Legacy of the Market King:
The Freezer Gambit

By Helloworld 2.0 - Data_Crunch_045
University of Moratuwa

Introduction

For the **AgroChill** forecasting system, we have designed a robust pipeline that forecasts weekly fresh produce prices one month ahead for various regions and commodities in Agrovia. Our solution integrates historical weather and price data, scales and preprocesses the data, applies time-series modeling techniques, and continuously updates the predictions in real-time. This allows AgroChill to decide optimally when to sell fresh or freeze produce, ultimately maximizing profitability while minimizing waste.

System Architecture

The **AgroChill** system is built on a modular architecture, consisting of the following key components:

- **Data Ingestion**: Data from multiple sources, including price and weather datasets, is ingested into the system.

- **Data Processing Pipeline**: The data undergoes cleaning, merging, feature engineering, and reindexing to ensure consistency and completeness. The cleaned dataset is used as input for our forecasting models.

- **Forecasting Engine**: The core component of the system utilizes an **N-BEATS** model from the **Darts** library for time-series forecasting.

- **API Layer**: A set of REST APIs is exposed to interact with the model, allowing users to retrieve predictions, submit updated data, and access forecasting results.
    1. `/api/predict` **(POST)**: This API predicts the future crop prices for a given crop and region. The user provides the crop and region details, and the API returns the predicted prices for the next 4 weeks, including the respective dates.

    2. `/api/data/weather` **(POST)**: This API allows users to submit weather data, including temperature, rainfall, and humidity, for a specific region and date. The API uses this data to predict the crop yield impact score and stores both the weather data and the prediction for future use.

    3. `/api/data/prices` **(POST)**: This API enables users to submit crop price data for a specific region and date. The user sends the crop, region, date, and price, and the API stores this information in the dataset and returns a success message confirming the storage.

    4. `/api/insights` **(POST)**: This endpoint analyzes price predictions for a specific crop and region, calculates depreciation rates for frozen produce, and provides

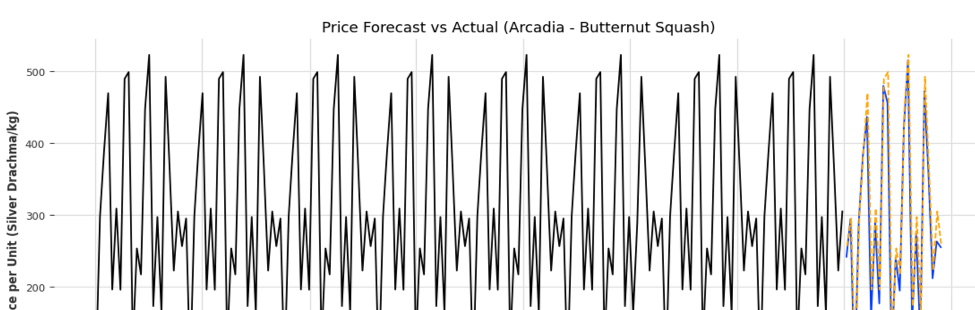
actionable insights on when to sell (fresh vs. frozen) to maximize profits.

- **Containerization**: The entire system is packaged in a **Docker** container, ensuring portability and scalability. This guarantees easy deployment across environments.

Forecasting Methodology

Our system predicts future commodity prices using the N-BEATS deep learning model from the Darts library, combined with weather impact features derived through a separate regression pipeline. Below is a breakdown of the methodology:

1. **Handling Missing Weeks**:
    - Weekly gaps are filled by reindexing each `(Region, Commodity)` group with a full date range and forward-filling missing values, ensuring consistent weekly data for modeling.
2. **Data Aggregation and Cleaning**:
    - Price and weather datasets are merged on both region and date. After removing duplicate entries, we aggregate features such as price, temperature, humidity, and rainfall to weekly means. This harmonized dataset enables us to align pricing data with corresponding weather conditions.
3. **Feature Scaling**
    - The weekly time series data is scaled using a Min-Max scaling approach. This helps to normalize values like price and crop yield impact scores, improving model stability and convergence during training.
4. **Time Series Conversion**
    - The cleaned and scaled data is converted into Darts `TimeSeries` objects. One series captures the target variable (price), while another represents the past covariates, including the crop yield impact score, which enriches the model with external weather-related context.
5. **Model Training**
    - We train a deep learning model based on the N-BEATS architecture. The model learns temporal patterns in historical prices while considering past covariates related to weather conditions. The model is trained across all available series for various crop-region combinations.
6. **Rolling Forecast**
    - The trained model generates a rolling forecast for the next 4 weeks, extending from the most recent data point in each series. The predicted prices are then rescaled back to their original values for interpretation. These forecasts are presented alongside historical prices to evaluate performance visually.



Model Evaluation

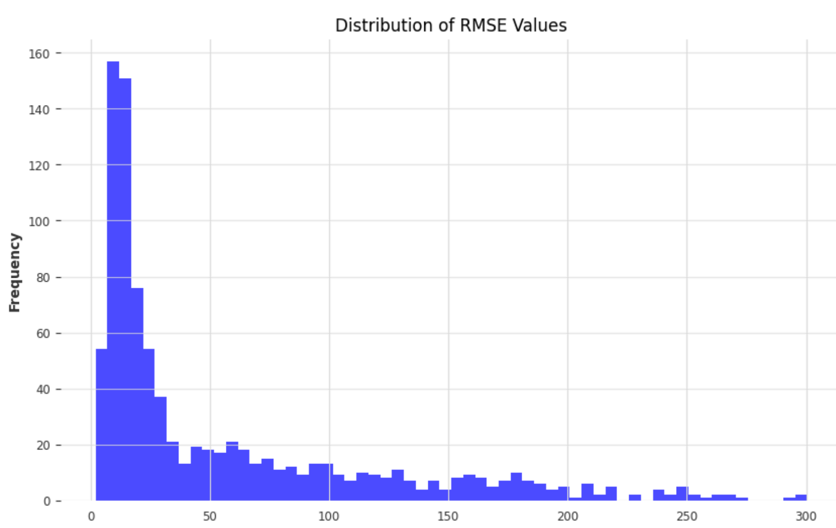
In the model evaluation, we used the **Root Mean Squared Error (RMSE)** to assess the performance of our prediction model. RMSE measures the square root of the average of the squared differences between predicted and actual values. Essentially, it quantifies how far off our model's predictions are from the actual outcomes, with lower values indicating better model performance. Here's how we evaluated the model:

1. **Calculation of RMSE**: For each prediction, we calculated the squared difference between the predicted and actual prices, averaged these squared differences across all data points, and then took the square root of the result. This provided us with the RMSE value for each prediction.

2. **Summary Statistics**: The summary statistics provided for RMSE are:

    - **Count**: 925 samples were used in the evaluation.
    - **Mean**: The average RMSE across all predictions is 55.98, meaning that, on average, our predictions were off by approximately 56 units from the actual prices.
    - **Standard Deviation**: The RMSE values vary with a standard deviation of 62.95, indicating some degree of variability in the prediction errors.
    - **Min and Max**: The smallest RMSE value was 2.09, and the largest was 300.31, showing that while most predictions were close, some had substantial errors.
    - **25th, 50th, 75th Percentiles**: The RMSE values at these percentiles (12.62, 24.27, and 80.82, respectively) provide insight into the distribution of errors. The median RMSE is 24.27, suggesting that half of the predictions were within this range of error.

The RMSE statistics indicate a spread in the model's prediction accuracy, with many predictions being relatively close to actual values (the median RMSE is around 24.27), but a significant portion of predictions deviating more (with RMSE as high as 300.31)



Business Insights & Recommendations

Based on the forecasting results, we derive actionable insights that help AgroChill optimize its operations:

1. **Market Trends & Price Prediction**:
   - We identify periods when prices are expected to rise or fall, enabling AgroChill to decide whether to store or sell produce at optimal times.
2. **Storage & Freezing Strategy**:
   - When prices are predicted to fall in the coming weeks, the system recommends freezing produce to avoid losses, helping AgroChill manage its storage effectively.
3. **Weather Impact on Pricing**:
   - Our system reveals the impact of weather conditions (e.g., temperature, rainfall) on crop prices, allowing AgroChill to prepare for fluctuations due to environmental factors.

Deployment Strategy

Our solution is packaged into a Docker container, ensuring it is portable, scalable, and easy to deploy across different environments. The following components are included in the deployment:

- **Model**: The time-series prediction model is embedded within the Docker container. This ensures that the model is always up-to-date and can seamlessly integrate with the latest data, enabling accurate predictions at any time.
- **APIs**: Exposed APIs allow external systems to interact with the application. These APIs enable users to retrieve forecasted prices, submit new weather data, and upload price data.
- **Cloud Deployment (Optional)**: For improved scalability, availability, and public access, the solution can be deployed on a cloud platform such as AWS or GCP. This will provide benefits like automatic scaling, high availability, and easy management, ensuring the solution can handle increasing loads effectively.

Challenges & Future Improvements

- **Data Completeness**: Missing weather data was a challenge, but we handled this with imputation techniques. In the future, we can use more sophisticated models for imputation.

- **Model Improvement**: We can add future covariates like month, year, and day information, as well as seasonal trends, to provide the model with richer context and improve the accuracy of crop price predictions.

Conclusion

The **AgroChill** forecasting system provides a highly effective solution for predicting agricultural commodity prices. Through the integration of weather data, time-series forecasting models, and real-time data processing, the system enables AgroChill to make better decisions on when to sell or store produce, ensuring maximum profitability and reducing waste.