Commodity Price Prediction Application

Kaavish Project Proposal

By

Sidra Aamir (sa07316@st.habib.edu.pk)
Owais Aijaz (oa07610@st.habib.edu.pk)
Aiza Imran (ai07597@st.habib.edu.pk)
Syed Hamza (hs07141@st.habib.edu.pk)

October 1, 2024

In partial fulfillment of the requirement for

Bachelor of Science
Computer Science

Dhanani School of Science and Engineering

Habib University

Fall 2024

1 Abstract

Pakistan's agricultural sector is extremely significant to the economy of the country and employs a large segment of the population. However, it struggles with the price volatility of key commodities such as wheat, cotton, maize, and rice, which creates uncertainty for crop traders and other stakeholders. The prices of these commodities are affected by factors like weather, demand fluctuations, and global trade, and this fluctuation can lead to financial instability and market inefficiencies. To address this challenge, a robust solution is required which must be capable of forecasting commodity prices with precision.

Our proposed solution is a machine learning-based model designed to predict agricultural commodity prices using a combination of historical data and external variables. By integrating techniques such as Random Forest, Gradient Boosting, and deep learning models (LSTM and GRU), the system will analyze past trends and real-time factors like weather and market conditions. We aim to develop a user-friendly application to incorporate our model, that will provide insights for farmers, traders, and policymakers, enabling better decision-making and minimizing financial risks.

This project aims to enhance the efficiency of agricultural markets, support the economic stability of rural communities, and contribute to sustainable farming practices by offering a data-driven approach to navigating price fluctuations.

2 Problem definition

Pakistan is an agricultural country and this sector accounts for around 24% of the country's GDP, employs half of the labour force, and provides the majority of foreign exchange profits [1]. However, price volatility is a constant issue in this area. Crop traders and investors, in particular, are heavily impacted by price fluctuations due to complexities in predicting market trends. Crop pricing is influenced by several factors, including production levels, consumer demand, and climate conditions, making crop price forecasting a complex yet vital topic [2].

The problem this project aims to address is the unpredictable nature of commodity prices in Pakistan's agriculture sector, which leads to significant financial uncertainty for crop traders and other stakeholders. Commodity price forecasting is needed to consider all impactful trends to provide a reliable resource for investors and governments to reduce risks related to price volatility [3]. Accurate price prediction models could assist traders in deciding the best time to sell their produce to maximize benefit. Our project aims to utilize historical data on commodity price trends and employ machine learning techniques to develop a commodity price prediction model, aiding stakeholders in making informed decisions. This project has the potential to improve agricultural market efficiency, reduce economic losses, and establish sustainable agricultural practices by optimizing production strategies.

3 Social relevance

Given that approximately 64% of Pakistan's population resides in rural areas and relies on agriculture for their livelihoods [1], this project holds significant social relevance. By developing a reliable commodity price prediction model, we aim to empower crop traders and agricultural stakeholders with accurate market insights. This will ensure that the returns on their agricultural products are fair and commensurate with market conditions, helping them avoid financial losses and minimize crop wastage.

Accurate price prediction enhances national food security by improving supply chain efficiency and preventing the overproduction or underproduction of crops. This will lead to the development of a sustainable agricultural system and stimulate greater economic and social strength in Pakistan's farming communities.

4 Originality/Novelty

The development of an advanced Commodity Price Prediction Model offers significant value by addressing existing limitations in agricultural price forecasting systems. Current agricultural commodity markets, especially in developing countries such as Pakistan, suffer from inefficiencies due to volatility in prices, unpredictable weather patterns, and inadequate market data. Accurate price prediction is vital to mitigate the risks faced by stakeholders such as traders, investors, and policy makers.

4.1 Existing Solutions

Commodity price forecasting systems have been researched extensively, with models typically relying on statistical techniques such as ARIMA (Auto-Regressive Integrated Moving Average), GARCH (Generalized Autoregressive Conditional Heteroskedasticity) or simple linear regression approaches. However, these methods often fall short in addressing the complex, non-linear nature of agricultural commodity markets. Price fluctuations in agriculture are heavily influenced by a multitude of factors, including climatic conditions, export-import trends, global supply chains, government policies, and socio-economic indicators, which are challenging for traditional models to capture [4]. Research conducted by A. K. Mishra et al. (2018) indicates that traditional statistical models often under-perform when used to predict prices under such multi-factorial conditions, especially in volatile markets.

4.2 Proposed Innovation

The novelty of this project lies in the integration of advanced machine learning models and real-time external factor analysis into a price prediction system that is specifically designed for the complexities of agricultural commodities. Unlike existing systems, the proposed model will leverage both historical price data (2004-2024) and real-time external variables, including weather patterns, international trade dynamics, and global market trends, to improve prediction accuracy.

Furthermore, deep learning models such as Long Short-Term Memory (LSTM), GRU (Gated Recurrent Units), and Ensemble methods could be employed to capture long-range dependencies in time-series data and generate more reliable forecasts. Additionally, a user-friendly application will ensure accessibility for all stakeholders, traders and investors, allowing them to make data-driven decisions in real-time. By allowing customization, users can tailor the forecasts to their specific regional contexts or commodities of interest, making the system highly adaptable and personalized. A study by Zhang et al. (2020) demonstrated the power of deep learning models in improving forecast accuracy over traditional models, particularly in commodity price predictions when integrated with external economic indicators [5]

5 Scope and Deliverables

The project will span a year and involve four team members, each contributing to distinct phases of the project. Given the duration and group size, the project scope is designed to be both challenging and feasible, with specific milestones and clearly defined tasks for each phase.

5.1 Justification of Scope

Data Collection & Pre-processing

With four group members and a year-long timeline, we can dedicate ample time to gathering, cleaning, and organizing data from various sources. This includes consolidating datasets, filling in missing information, standardizing formats, and addressing inconsistencies. Each member will manage a specific subset (crop prices, weather, economic data). Our deliverable will be a cleaned, unified dataset covering historical prices, weather (temperature, humidity, air quality), and economic factors (tariffs, market trends, seasonality).

Modeling & Prediction

Identifying key factors affecting crop prices is essential for accurate predictions. The scope includes experimenting with multiple feature selection and statistical techniques (e.g., Principal Component Analysis, correlation analysis). This phase will be research-intensive and can be managed effectively with four members working collaboratively on analysis and testing. Additionally, we will test various machine learning models (such as Linear Regression, Decision Trees, and Random Forests) to achieve high accuracy, the extended project timeline allows for comparing model performances and fine-tuning hyperparameters to achieve the best accuracy, with a goal of 90-95%.

System Development

To show our predictions to the audience, we will create a user-friendly, cross-platform forecasting system either as a website or mobile application. It will have a dashboard that displays crop price predictions and visualizations of the factors influencing price changes. A year-long project gives the team ample time to design a user-friendly interface, develop the necessary backend, and integrate machine learning predictions into a real-time dashboard.

Testing & Validation

Given the importance of model accuracy and system performance, dedicating time to testing is critical. Testing both the machine learning models and the application functionality ensures reliability so we will conduct thorough testing.

5.2 Foreseeable Deliverables

1. **Cleaned and Structured Dataset** A consolidated and cleaned dataset, including historical price data, weather data (temperature, humidity, air quality, etc.), and economic factors (tariffs, market trends, seasonality).
2. **Feature Selection Report**We will deliver a report identifying key features that have the greatest impact on crop price prediction.

3. **Machine Learning Models** Implemented and tested machine learning models (e.g., Linear Regression, Random Forests) with fine-tuned hyperparameters to achieve 90-95% accuracy in crop price prediction.
4. **Forecasting System (Website/Mobile App)** A fully functional cross-platform website or mobile application displaying crop price predictions with an interactive dashboard.
5. **Comprehensive Documentation** Full project documentation, including a user manual, technical specifications, and a final project report.
6. **Final Presentation & Demonstration** A presentation showcasing the system's functionality, key insights from the feature selection process, model accuracy, and a live demonstration of the forecasting system.

6 Feasibility

6.1 Datasets

- **Historical Commodity Prices**: This dataset will include historical pricing data for various agricultural commodities (cotton, sugar, wheat, maize, rice, edible oils, spices, etc.) from 2004 to 2024. This data is essential for training machine learning models to predict future prices accurately.
- **External Factors**: Additional datasets will encompass weather patterns, economic indicators, global market trends, and government policies. These factors are crucial for enhancing the model's predictive accuracy by providing context to the historical price data.

Access: Collected historical data will be supplemented with external datasets obtained through APIs or web scraping techniques.

6.2 Compute Resources

- **Cloud Computing Services**: Platforms like AWS or Google Cloud will be necessary for scalable storage and processing of large datasets. This infrastructure will support real-time data integration and model training.

Access: Cloud services will be selected based on budget and computational power requirements for model training and deployment.

6.3 Hardware

- **Development Machines**: Team members will need personal computers with adequate processing power (preferably with GPUs) to handle data processing and model training tasks effectively.
- **Testing Devices**: Various smartphones and tablets will be required for testing the mobile application across different platforms and screen sizes.

Access: Team members will utilize their own devices, supplemented by necessary hardware acquired through institutional resources.

6.4 Software

- **Machine Learning Libraries**: Python libraries such as Pandas, NumPy, Scikit-learn, TensorFlow/Keras will be essential for data manipulation, machine learning model development, and neural network implementation.
- **Data Visualization Tools**: Software like PowerBI or Tableau will be required for creating interactive visualizations that present both historical data and future predictions.
- **Frontend Development**: React.js will be used to develop a user-friendly interface that works seamlessly on both mobile and web platforms.
- **NLP Tools**: Natural Language Processing libraries (e.g., NLTK or SpaCy) will be necessary for analyzing unstructured data sources to enhance prediction capabilities.

Access: Most software libraries are open-source or available through institutional licenses, making them accessible for the project team.

6.5 Challenges and Bottlenecks

- **Data Quality and Availability**: Ensuring the availability and quality of datasets can be challenging, particularly for external factors that may not be consistently reported.

  Solution: Establish partnerships with NGOs and governmental agencies to access reliable datasets and ensure comprehensive coverage of external factors that affect agricultural commodity prices.

- **Model Complexity**: The multifactorial nature of agricultural commodity prices, influenced by various external factors, may complicate model training and reduce prediction accuracy.

  Solution: Employ advanced machine learning techniques such as ensemble methods or deep learning models (LSTM, GRU) to better capture complex patterns in time-series data, improving model performance.

- **Training Data Anomalies**: The training data from specific periods, such as the COVID-19 pandemic, may contain anomalies that could skew predictions and reduce model reliability.

  Solution: Carefully analyze historical data to identify anomalies and implement robust preprocessing techniques like outlier detection and handling to mitigate the effects of these anomalies on model training.

- **Unpredictable Events**: Just as the COVID-19 pandemic was unpredictable, future unforeseen events could disrupt market conditions, rendering existing models less effective.

  Solution: Design the model to incorporate real-time updates and adaptive learning techniques, allowing it to adjust to sudden changes in market conditions and remain effective despite unpredictable events.

- **User Adoption**: Ensuring that stakeholders, such as traders, adopt the application may require extensive user education and support.

  Solution: Develop comprehensive user manuals, conduct workshops, and provide ongoing support to demonstrate the application's benefits, ensuring smooth adoption and usage by all stakeholders.

7 Tech stack

1. Frontend

- **React.js** – For building the web user interface.
- **React Native** – For mobile app development, enabling cross-platform compatibility.
- **Material-UI** or **Tailwind CSS** – For responsive design and modern UI components.

2. Backend

- **Node.js** (with Express.js) – For building the server-side API.
- **MongoDB** – NoSQL database for storing application data.
- **Express.js** – For routing and handling HTTP requests.
- **Power BI** or **Tableau -** For creating interactive visualizations which can be embedded in the React app.

3. Machine Learning

- **Python** – Core language for building machine learning models.
- **TensorFlow** or **PyTorch** – For deep learning and neural network models.
- **Scikit-learn** – For traditional machine learning algorithms and pre-processing.
- **Pandas** and **NumPy** – For data manipulation, pre-processing, and data visualization.

4. NLP Tools

- **NLTK** and **SpaCy** – Apart from the core language Python, Natural Language Processing libraries will be necessary for analyzing unstructured data sources.

5. Deployment

- **Google Cloud** or **AWS Lambda** – For deploying the backend and machine learning models.

References

[1] P. B. of Statistics, "Agriculture statistics: Pakistan bureau of statistics," https://www.pbs.gov.pk/content/agriculture-statistics#:~:text=%E2%80%8BAgriculture%20constitutes%20the%20largest,source%20of%20foreign%20exchange%20earnings, accessed: Sep. 30, 2024.

[2] S. K. Purohit, S. Panigrahi, P. K. Sethy, and S. K. Behera, "Time series forecasting of price of agricultural products using hybrid methods," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 1388–1406, Sep. 2021.

[3] A. Rasheed, M. S. Younis, F. Ahmad, J. Qadir, and M. Kashif, "District wise price forecasting of wheat in pakistan using deep learning," https://arxiv.org/abs/2103.04781, accessed: Sep. 30, 2024.

[4] C. Sun, M. Pei, B. Cao, S. Chang, and H. Si, "A study on agricultural commodity price prediction model based on secondary decomposition and long short-term memory network," *Agriculture*, vol. 14, no. 1, p. 60, 2024, submission received: 23 November 2023 / Revised: 25 December 2023 / Accepted: 27 December 2023 / Published: 28 December 2023. [Online]. Available: https://doi.org/10.3390/agriculture14010060

[5] D. Zhang, S. Chen, L. Ling, and Q. Xia, "Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons," *IEEE Access*, vol. PP, no. 99, pp. 1–1, February 2020. [Online]. Available: https://www.researchgate.net/publication/339034087_Forecasting_Agricultural_Commodity_Prices_Using_Model_Selection_Framework_With_Time_Series_Features_and_Forecast_Horizons

Undertaking of Kaavish advisement as an External Supervisor

I hereby affirm that I have read the project details as described on the preceding pages and agree to undertake advisement of this Kaavish project as an External Supervisor. I understand that this role entails the following.

Meeting Meeting the project team regularly, at least once every two weeks, for the entire duration of the Kaavish. The meetings may be held remotely if required.

Advisement Providing supervision and advice to the team in order to ensure steady progress of the project toward its goals.

Liaison Liaising with the Internal Supervisor as required, e.g. to provide feedback or engage in grading.

Other Any other task, depending on availability and suitability, relevant to the Kaavish as communicated by the Internal Supervisor or Kaavish Working Group.

Name: ______________________________

Email: ______________________________

Phone: ______________________________

Designation: ______________________________

Affiliation: ______________________________

Signature: ______________________________