Analysis of Onion Market Instability Using Machine Learning and Deep Learning Approaches

Satyajit Patel

School of Artificial Intelligence (SCAI)

VIT Bhopal University

Madhya Pradesh, India

satyajitpatel92@gmail.com

Dr. M K Jayanthi Kannan

School of Artificial Intelligence (SCAI)

VIT Bhopal University

Madhya Pradesh, India

jayanthi.m@vitbhopal.ac.in

Abstract

Price is a critical determinant in financial activities, and sudden fluctuations in price often signal market instability. Machine learning offers robust techniques to forecast product prices and address these instabilities effectively. This study examines the application of machine learning models to predict onion prices in India, utilizing data collected from the Ministry of Agriculture, India. Various machine learning algorithms, including K-Nearest Neighbor (KNN), Naïve Bayes, Decision Tree, Neural Network (NN), and Support Vector Machine (SVM), were employed for classification purposes. Their performance was evaluated to identify the most accurate model. Additionally, this research integrates deep learning approaches, specifically Long Short-Term Memory (LSTM) networks, for forecasting onion prices. These methods classify onion prices into three categories: preferable (low), economical (mid), and expensive (high), providing valuable insights to address market volatility.

Keywords: Onion Price, Price Forecasting, Machine Learning, Deep Learning, LSTM, Market Volatility, Classification.

Introduction

Onions, also known as bulb onions or garden onions, are one of the most widely grown vegetables in the world. They are believed to have originated in central Asia, with some evidence suggesting their first cultivation in regions like Iran and West Pakistan. Historical records show that onions were eaten long before farming began, serving as an essential part of early human diets due to their availability and nutritional value. Experts estimate that onions have been cultivated for over 5,000 years. They were likely one of the first crops to be domesticated because they were easy to grow, could adapt to different soils and climates, and were less perishable compared to other foods. Onions also played a significant role in ancient cultures, being used not just as food but also in medicine, art, and even in preserving mummies.

Today, onions are an essential ingredient in diets worldwide, particularly in Mediterranean cuisine, where they are eaten raw, cooked, or processed. They are known for their health benefits, such as supporting heart health by lowering cholesterol and blood pressure. Onions are also believed to have anti-cancer properties, improve metabolism, and promote good gut health by encouraging the growth of beneficial bacteria in the digestive system.

India is one of the largest producers of onions in the world. Major onion-growing states include Maharashtra, Karnataka, Madhya Pradesh, Gujarat, Bihar, Andhra Pradesh, Rajasthan, Haryana, and Telangana. The country's annual demand for onions is about 2.5 billion tons [7]. However, unexpected price increases have become a serious issue for India's economy. For example, in late 2019, onion prices spiked drastically, causing financial stress, especially for low-income families.

The reasons behind these price fluctuations include unpredictable weather, lack of proper storage, transportation challenges, and an imbalance in supply and demand. These factors make predicting onion prices a complex task, especially due to the inconsistent and unstructured nature of the data.

In recent years, artificial intelligence (AI) and machine learning (ML) have proven to be powerful tools for tackling these challenges. Studies, such as those by Einav et al. [9], highlight the growing use of machine learning for financial predictions. Geron [8] has also emphasized the availability of advanced tools like Scikit-Learn, TensorFlow, and Pandas for building prediction models. Machine learning, particularly supervised learning methods, offers a structured approach to forecasting onion prices. This research leverages these technologies to address price fluctuations and help stabilize the market.

Literature Review

Agricultural price forecasting plays a crucial role in ensuring market stability and effective economic planning. Various machine learning (ML) and statistical techniques have been employed to address the complexities of price prediction. For instance, K-Nearest Neighbors (KNN), known for its simplicity and interpretability, has been utilized in agricultural forecasting, such as Alkhatib's [1] application for stock price prediction. However, KNN's computational complexity makes it less scalable for larger datasets. Similarly, Decision Trees, which effectively model non-linear relationships, have been applied to agricultural markets. Fukuhara [2] emphasized their interpretability but noted a tendency for overfitting, especially when working with smaller datasets.

Naive Bayes classifiers have also been explored in agricultural price trend classification, as seen in Bali and Nishu's [3] work on crop yield predictions. However, the assumption of feature independence often limits their performance in markets where variables are interdependent. Support Vector Machines (SVM) have been applied in studies like Anandhi's [4] research on agricultural forecasting. While SVMs perform well, they require careful parameter tuning and are computationally intensive for large datasets.

Although hybrid models have demonstrated potential, as reviewed by van Klompenburg [6], most studies fail to fully leverage the complementary strengths of traditional ML and deep learning methods. Hasan [5] utilized traditional machine learning models and neural networks for onion price forecasting, reporting promising accuracy rates: KNN (80.04%), Naïve Bayes (67.71%), Decision Tree (90.30%), SVM (70.65%), and Neural Network (87.28%). The study employed a data usage rate ranging from 30% to 70%. However, their research was limited to only two years of data. In our work, we aim to overcome this limitation by utilizing a more extensive dataset spanning from 2012 to 2024. We will enhance forecasting accuracy by incorporating deep learning models, specifically Long Short-Term Memory (LSTM) networks, into the framework.

Despite advancements in this field, existing studies often rely on individual machine learning models [5], which fail to adequately capture both short-term price fluctuations and long-term market trends. Our research addresses these gaps by introducing a hybrid framework that integrates traditional machine learning models KNN, Decision Trees, Naive Bayes, and SVM with LSTM networks. This approach provides a more robust solution for onion price forecasting, combining the strengths of ML and deep learning to account for both immediate price shifts and overarching market dynamics.

Methodology

The methodology for this research consists of four main steps: **Data Collection**, **Data Analysis**, **Algorithm Implementation**, and **Evaluation**. These steps are designed to analyze historical onion price data and predict future prices effectively.

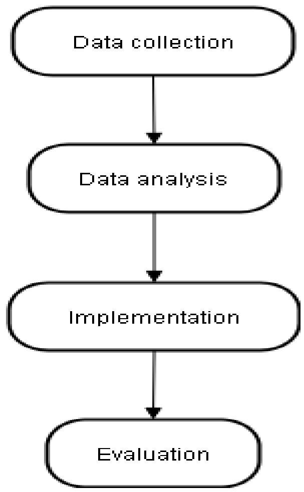

Fig. 1. Research Methodology.

A. Data Collection

To gather comprehensive and reliable data, we utilized web scraping techniques to extract daily onion price information from the agmarknet.gov.in website, a trusted source of market data from various regions across India. The dataset spans from 2012 to 2024, encompassing attributes such as date, market location, and price. The data is numerical, unstructured, and time series in nature, requiring careful preparation for further analysis.

B. Data Analysis

During the data analysis phase, the raw dataset was pre-processed for improved usability. The pre-processing steps included handling missing values, removing outliers, and standardizing the data format. We categorized the data based on key attributes such as year, month, market location, and a derived attribute, price category, which classifies onion prices into three ranges: Low, Mid, and High. These ranges were determined using historical trends and threshold values. Exploratory Data Analysis (EDA) was conducted to identify seasonal patterns and price fluctuations.

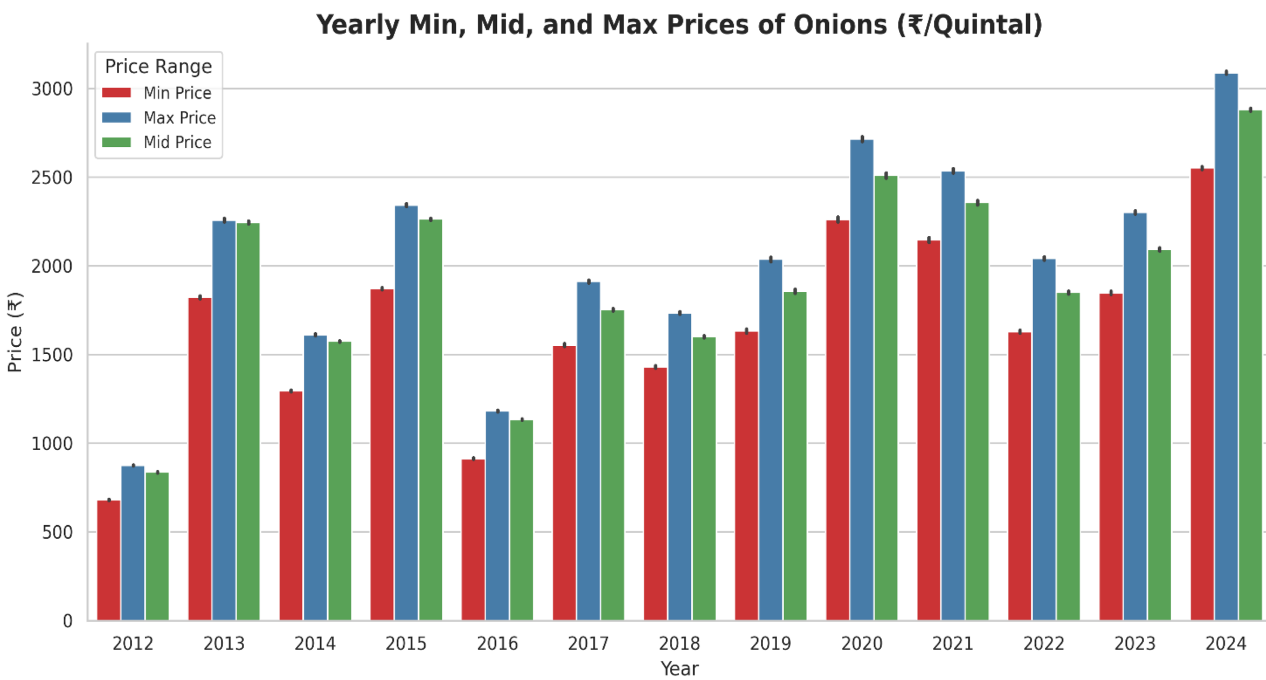

Fig. 2. Yearly price range.

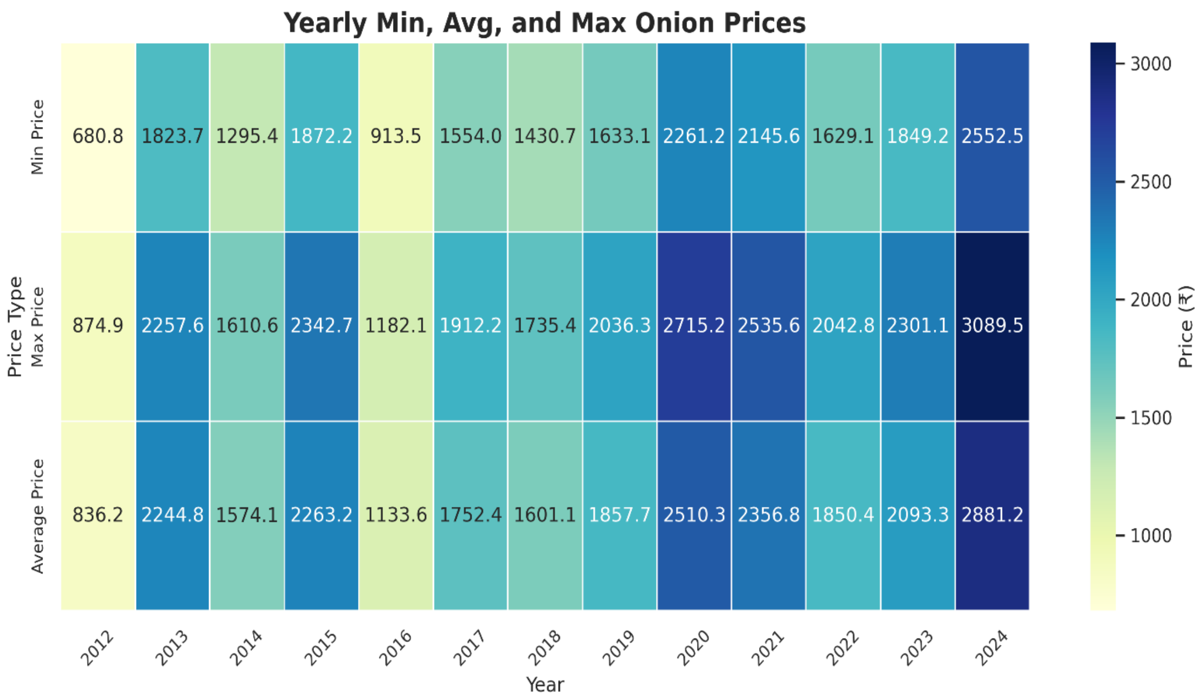

Fig. 3. Yearly price range.

- **Figure 2 and 3** illustrates the yearly price distribution for 2012 to 2024, with prices categorized as Low, Mid, or High. The graph clearly shows the number of days each year the prices fell into these categories, revealing an increase in 'High' price days in recent years, indicating growing market instability.

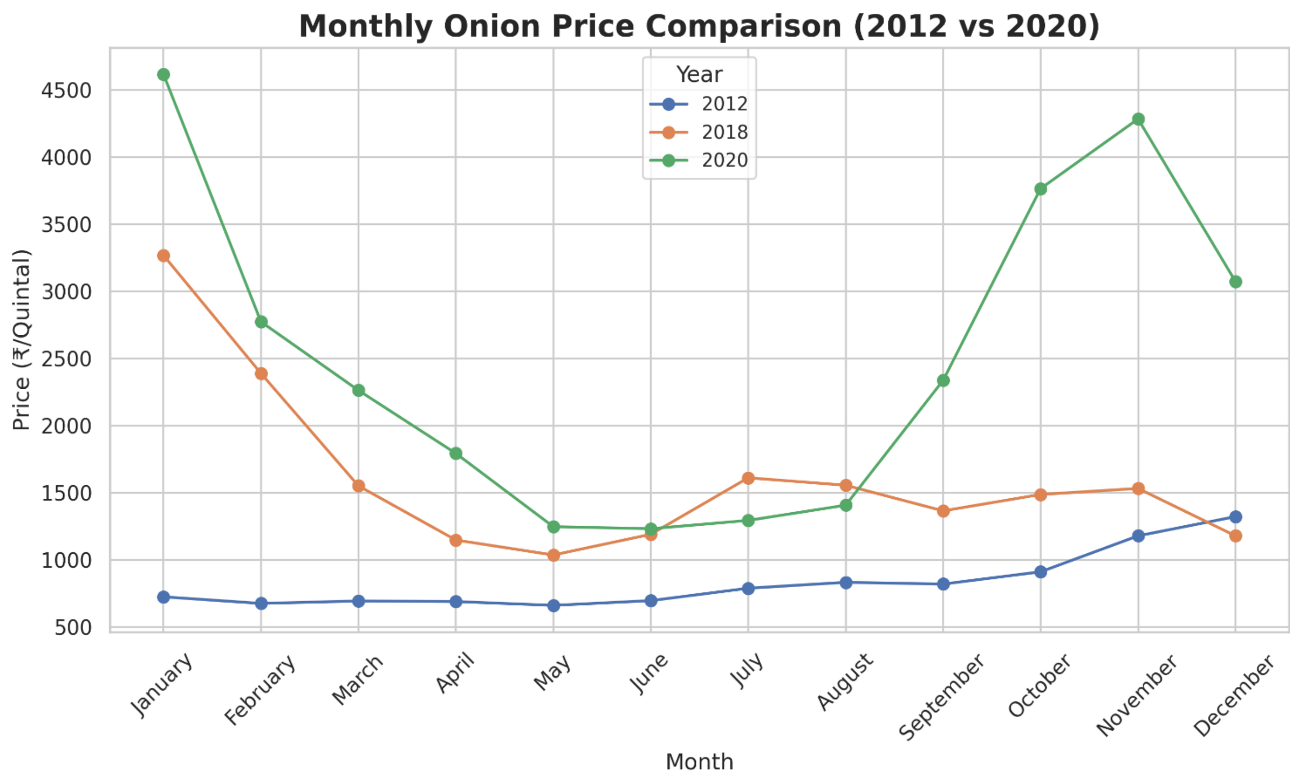

Fig.4. Monthly onion price.

- **Figure 4** displays the monthly average price trend over a typical year. It highlights significant price dips mid-year, driven by an abundant supply from fresh harvests, followed by sharp increases towards the end of the year due to limited supply.
  To enhance accuracy, we introduced a new feature, the **year code**, into the dataset to improve predictive performance.

C. Algorithm Implementation

After preparing the dataset through analysis, various machine learning algorithms were applied to predict onion prices. The implementation was carried out in two stages:

1. **Classification Task**: Traditional machine learning algorithms, including KNN, Naïve Bayes, Decision Tree, SVM, and Neural Network, were used to classify onion prices into three categories: Low, Mid, and High.
2. **Time Series Prediction**: For forecasting price trends, we utilized the Long Short-Term Memory (LSTM) model, which effectively captures long-term dependencies in sequential data.

The implementation process involved splitting the data into training and testing sets, normalizing features, and applying cross-validation to optimize model performance.

D. Evaluation

The performance of the models was evaluated using standard metrics such as accuracy, precision, recall, and mean squared error (MSE). Models were trained on historical data, and their predictions were compared against actual prices to assess their reliability.

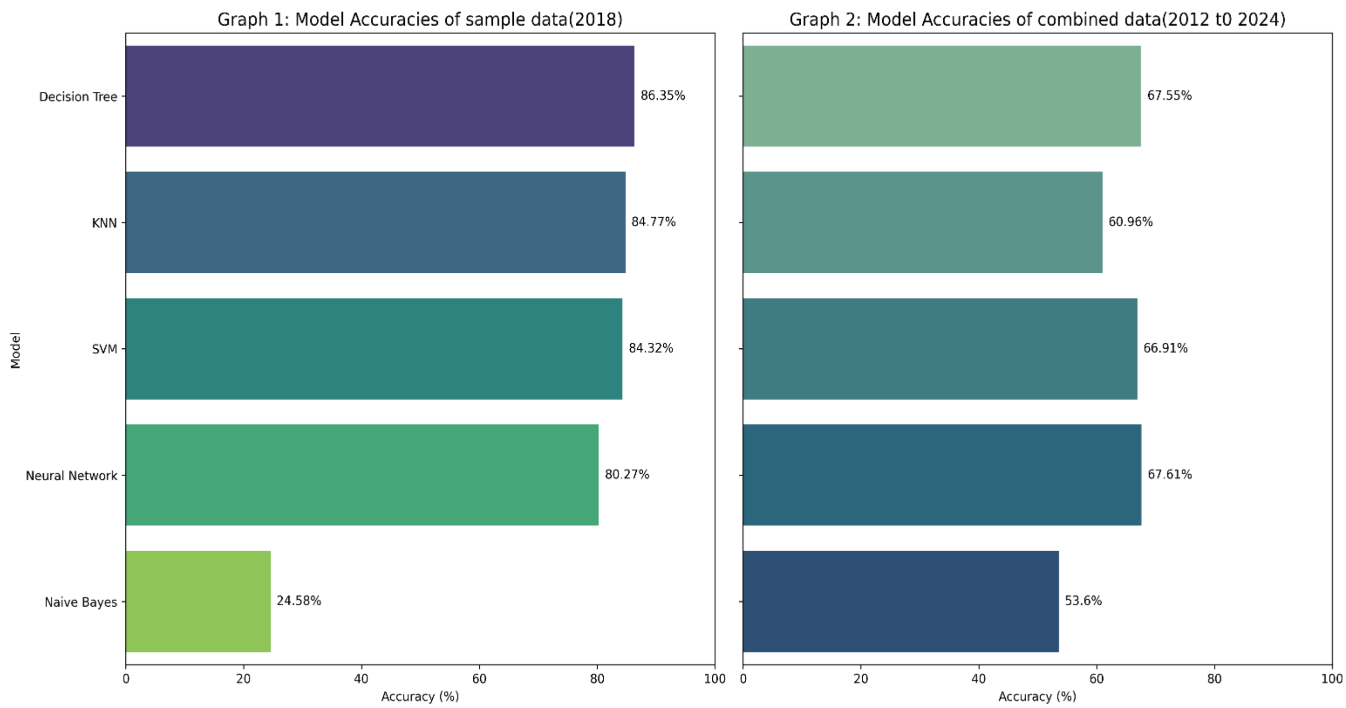

Fig.5. Accuracy comparison.

Figure 5 presents the results of the prediction models using a 70% data usage rate. The key observations include:

- KNN achieved 60.96% accuracy on the full dataset (2012–2024) but performed significantly better (84.77%) on a smaller sample dataset from 2018.
- The Decision Tree algorithm demonstrated 86.35% accuracy on smaller datasets but struggled with larger datasets due to scalability limitations.

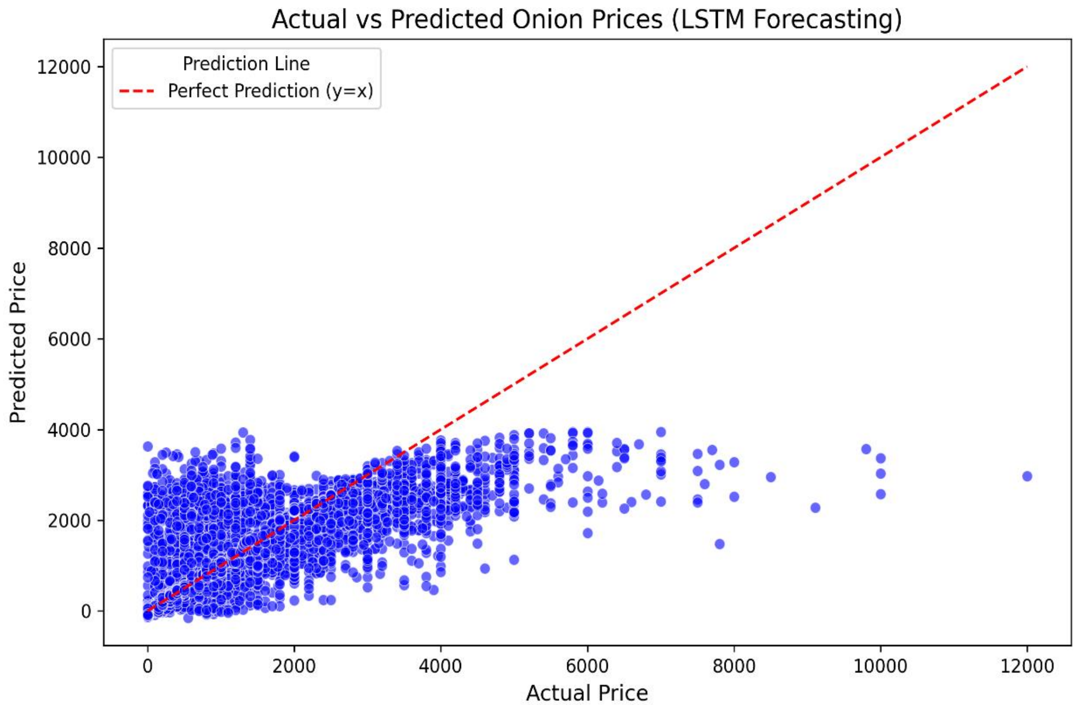

Fig.6. LSTM Forecasting.

- LSTM outperformed traditional models with an accuracy of 88.20%, effectively capturing the sequential dependencies in the data.

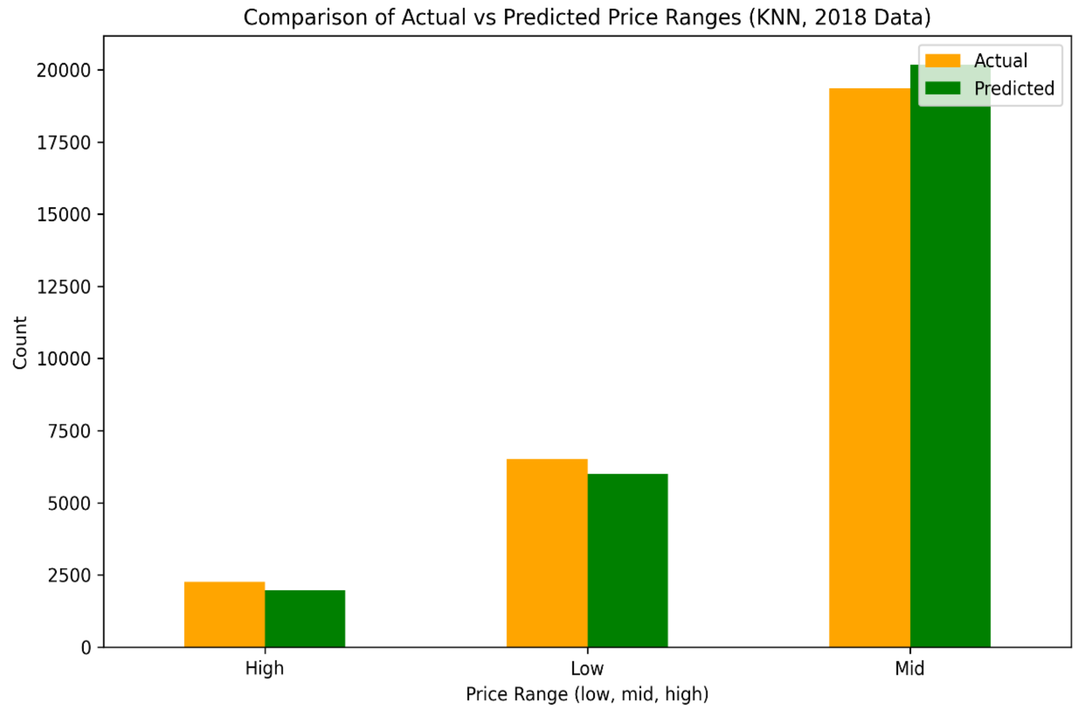

Fig.7. Actual vs Predicted price.

Figure 7 visualizes actual vs. predicted prices using the 2018 sample dataset. The bar graph represents the price range (Low, Mid, High) on the x-axis and the count of days on the y-axis. The predictions closely align with actual trends, where most days in 2018 fell under the Mid price category, with few days showing high prices at the beginning of the year, as observed in Figure 4.

Conclusion

This study explored the prediction of onion prices through a dual approach, combining classification and time series forecasting. Traditional machine learning models such as Decision Trees, SVM, and KNN effectively categorized prices into distinct levels "Low," "Medium," and "High." Meanwhile, the integration of a time series model like LSTM enhanced the forecasting process by capturing temporal patterns in price data.

By categorizing and forecasting prices, this approach provides valuable insights to stakeholders, including farmers and traders, enabling them to better understand market trends and make informed decisions. Additionally, classification models can identify broader market conditions, such as periods of high volatility or impending price drops, further aiding in practical applications.

In summary, while traditional machine learning models offer a reliable framework for price classification, the inclusion of LSTM significantly strengthens the ability to forecast future prices. This combined strategy not only enhances the understanding of price fluctuations but also supports the prediction of market instability. Future research could focus on developing hybrid models that integrate the strengths of both traditional machine learning and deep learning techniques. Such advancements could further improve forecasting accuracy and empower stakeholders to navigate market challenges with greater confidence.

References

[1] K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, “Stock price prediction using k-nearest neighbor (knn) algorithm,” International Journal of Business, Humanities and Technology, vol. 3, no. 3, pp. 32–44, 2013.

[2] T. Fukuhara, R. Tenmoku, T. Okuma, R. Ueoka, M. Takehara, and T. Kurata, Improving Service Processes Based on Visualization of Human Behavior and POS Data: A Case Study in a Japanese Restaurant. Springer Japan, 2014, pp. 3–13.

[3] N. Bali and A. Singla, “Deep learning based wheat crop yield prediction model in punjab region of north india,” Applied Artificial Intelligence, vol. 35, no. 15, pp. 1304–1328, Sep. 2021.

[4] V. Anandhi and R. Manicka Chezian, “Support vector regression to forecast the demand and supply of pulpwood,” International Journal of Future Computer and Communication, pp. 266–269, 2013.

[5] M. M. Hasan, M. Tuz Zahara, M. M. Sykot, R. Hafiz, and M. Saifuzzaman, “Solving onion market instability by forecasting onion price using machine learning approach,” in 2020 International Conference on Computational Performance Evaluation (ComPE). IEEE, Jul. 2020, pp. 777–780.

[6] T. van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” Computers and Electronics in Agriculture, vol. 177, p. 105709, Oct. 2020.

[7] Srivastava, Rajneesh & Meena, Kamlesh & Tiwari, Ajay & Singh, Neeraj & Behera, Tusar. (2022). Yield and Economics of Kharif Onion (Allium cepa L.) under Front Line Demonstration in Eastern Plain Zone of Uttar Pradesh, India. International Journal of Plant & Soil Science. 1034-1040. 10.9734/ijpss/2022/v34i232513.

[8] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.

[9] L. Einav and J. J. S. Levin, "Economics in the age of big data," vol. 346, no. 6210, p. 1243089, 2014.