OPEN ACCESS

Volume: 12

Special Issue: 1

Month: October

Year: 2024

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Citation:
Raghu, HN, and Aluregowda. "Unlocking Future Commodities Markets: Innovative Approaches to Price Prediction Using Supervised Machine Learning." *Shanlax International Journal of Arts, Science and Humanities*, vol. 12, no. S1, 2024, pp. 133–40

DOI:
https://doi.org/10.34293/sijash.v12iS1-Oct.8295

Unlocking Future Commodities Markets: Innovative Approaches to Price Prediction Using Supervised Machine Learning

Mr. H.N. Raghu
Assistant Professor, Department of Management
Government First Grade College, KR Pete, Mandya District, Karnataka, India

Dr. Aluregowda
Associate Professor and Head
Department of Management Studies and Research Centre
PES College of Engineering, Mandya, Karnataka, India

Abstract

This study discovers the application of managed machine learning methods for expecting commodity prices, addressing the limitations of outdated forecasting approaches. By incorporating varied data sources, containing historical prices, macroeconomic signs, and geopolitical aspects, the study employs a range of machine learning systems such as Gradient Boosting, Support Vector Machines, and Deep Learning replicas. The research validates that these advanced methods can considerably enhance predicting accuracy equated to conventional methods. Through robust model training procedures, including cross-validation and hyper parameter tuning, the study recognizes key features and advances the projecting power of commodity price replicas. The results emphasize the potential of machine learning to offer more reliable and vibrant estimates in the face of market unpredictability and external worries. This effort provides valued understandings for merchants, stakeholders, and legislators, educating decision-making and risk administration in complex and erratic product markets.

Keywords: Prediction, Supervised Machine Learning, Forecasting Accuracy, Feature Selection, Model Evaluation, Gradient Boosting.

Introduction

Commodity markets are essential to the global economy, providing the basis for the trading of crucial raw materials such as oil, natural gas, metals (e.g., gold, silver, copper), and agricultural products (e.g., wheat, corn, coffee).These marketplaces not only simplify trade but also allow manufacturers and consumers to hedge alongside price risks.(Schofield, (2021)) For instance, oil-exporting countries heavily rely on these markets for economic stability, while companies involved in production or agriculture depend on foreseeable commodity prices to accomplish operational costs. As cargos serve as raw inputs for numerous-businesses, price oscillations can impact supply chains, manufacture costs, and eventually consumer prices for finished goods (Dhandayuthapani & Sudha, (2018)).

Forecasting commodity prices precisely, however, is challenging due to a wide variety of factors that familiarize instability. Supply and demand inequalities are the primary drivers of price variations. For instance, a poor harvest due to weather circumstances can up agricultural product prices, while political uncertainty in major oil-producing nations can interrupt supply chains, impacting oil prices. Other outside factors like macroeconomic situations (inflation rates, currency fluctuations), geopolitical risks (trade sanctions, wars), and natural tragedies (hurricanes, droughts) create supplementary doubts in forecasting future values (Sehgal et al (2012)).

Furthermore, the rise of financial assumption has added another layer of difficulty to price forecast. Investors and risk-takers often trade commodities for profit relatively than for physical use, causing short-term price swings depend on market emotion. Worldwide merchandises markets are also influenced by technical advances (such as mining or agricultural techniques) and procedure changes (e.g., tariffs, subsidies), additional confusing price trends. (Seilan (2012))

Given these variables, orthodox economic outlines, such as those founded on supply-demand evenness or time-series analysis, habitually encounter problems in summarizing the complexities and variations inherent in commodity marketplaces. These outlines operate under the assumption of a relatively stable atmosphere and are generally ill-equipped to accommodate unexpected market evolutions or non-linear interdependencies amongst variables. For instance, apparently insignificant alteration in geopolitical dynamics could apply an exceedingly unequal effect on oil prices, a singularity that proves thought-provoking to forecast employing linear models. (Manokhin (2022)).

To overcome these restrictions, there has been arising interest in the application of supervised machine learning approaches to enhance the accuracy of commodity price projections. These models possess the competence to analyze widespread volumes of historical data, separate latent forms, and explain complex interrelationships between a multitude of aspects (Amin, 2020). Machine learning systems, when educated on complete datasets that incorporate price routes, macroeconomic signs, geopolitical incidences, and even investor emotion metrics, are capable of conveying more precise and adaptive predictions. This scientific progression holds the promise of supplementing market competence, thereby permitting traders, investors, and representatives to render more open-minded decisions within an increasingly volatile product environment. (Oktoviany et al (2021))

Managed machine learning is increasingly being engaged to address the difficulties associated with estimating commodity prices in dynamic and complicated market situations. In contrast to orthodox models that depend on recognized rules or linear associations, supervised algorithms arise insights from historic datasets, thereby increasing their capacity to recognize complex, non-linear configurations that frequently impact fluctuations in commodity prices (Dimitriadou et al., (2018)).

Relating to the estimate of commodity prices, overseen ML tasks by training models on extensive datasets that encompass historical price information alongside critical determinants, such as supply and demand dynamics, macroeconomic indicators, climatic conditions, and geopolitical occurrences. The model assimilates the interrelations among these variables and the resultant price movements, thereby enabling it to generate forecasts based on novel or previously unobserved data (An et al. (2019)).

A variety of controlled learning algorithms, such as Neural Networks, Decision Trees, Random Forests and Linear Regression aid as sophisticated practices for showing and predicting product prices. Regression analyses may plan the forthcoming price of petroleum by utilizing historical data in combination with macroeconomic metrics such as interest charges and inflation rates. More advanced practices, such as Neural Networks, hold the competence to uncover latent patterns and complicated interdependencies amongst various factors, thereby providing deep insights into oscillations in prices. (Manokhin, (2022)) (Panella et al (2012))

The ability of machine knowledge to analyze wide volumes of data from diverse bases comprising unconventional datasets such as social media sentiment and cable descriptions meaningfully enhances its predictive effectiveness. This multi-faceted method aids in improving prognostic incorrectness encouraged by sudden market interruptions or irregularities that orthodox models may fail to identify. Through the continuous integration of new data, managed machine learning models are able to adapt their answers to shifting market dynamics, so yielding forecasts that are more robust and consistent. (F.Y et al (2017))

In inference, machine learning effectively meets the difficulties connected with commodity price prediction by augmenting accuracy, regulating to fluctuating situations, and capitalizing on diverse data sources, thus providing a more dynamic and empirically-driven procedure in distinction to outdated forecasting methods.

Objectives

1. To gain a complete understanding of the structure and changing aspects of commodity markets
2. To discover and analyze advanced supervised machine learning methods for predicting commodity prices.
3. To assess how these innovations can improve future market efficiency, risk administration, and decision-making.

Review of Literature

Outdated stock market forecast methods use neural networks to uncover data patterns. Recent progressions employ machine learning replicas like Long Short-Term Memory and Convolutional Neural Networks for improved precision. This project integrates LSTM and CNN, applying a sliding window approach and evaluating performance with root mean square error. (Amar Gupta (2024)). This research introduces a complete crop price prediction approach using varied machine learning methods to account for ecological factors like rainfall and heat. By employing algorithms such as linear regression, decision trees, and XG-Boost, the study assesses and associates methods to enhance predicting correctness. The results provide actionable visions for agricultural investors, enlightening decision-making and risk organization in the face of market difficulties. (Minu Inba, (2024)). This study discovers market forecast using machine learning processes such as regression, time series models, and support vector machines. By concentrating on data preprocessing, feature engineering, and model assessment with metrics like MSE, RMSE, and MAPE, the study intentions to enhance prediction correctness. The outcomes provide valued understandings into the efficiency and confines of these methods in predicting stock prices among market ambiguity. (Prachi Pathak (2024)). This study examines numerous machine learning systems, including Generalized Neural Network, Support Vector Regression, Random Forest, and Gradient Boosting Machine, for forecasting the wholesale price of Brinjal in Odisha, India. Results reveal that GRNN consistently overtakes other methods, including traditional ARIMA models, in forecasting accuracy across all markets, with RF acting comparably in four markets. (Ranjit Kumar Paul (2022)). This study discovers the application of Automated Machine Learning for predicting used creation equipment prices, addressing the encounters of spatial and progressive variations. By incorporating AutoML with dominion knowledge and employing a novel method appraisal score, the research validates that AutoML can decrease dependence on ML specialists and rationalize the predicting process for small and medium-sized enterprises (Horst Stuhler (2023)). This study suggests a predicting model for forecasting future gold prices using Linear Regression. It talks the encounters modeled by gold's role in monetary steadiness and its attraction as an investment amongst market unpredictability. Whereas many trainings have attentive on forecasts, this research highpoint the presentation of machine learning to product forecasting, explicitly gold, to provide visions into price trends influenced by economic variables. (Swati Kumari, (2020)). This learning presents a hybrid approach for price prediction, merging various machine learning and deep learning models, including Long Short-Term Memory networks, to forecast nifty 50 index values. Outcomes-validate that the LSTM-based univariate model, using one-week prior data, attains the highest accurateness in foretelling prices. (Sidra Mehtab (2020))

Theoretical Background

Managed machine learning is a dominant approach that involves training a model on a labeled dataset, where each input data features are associated with a known outcome called label. This training course permits the model to learn the association between the features and the labels, permitting it to make precise forecasts on new, unseen data. Basically, the model simplifies outlines from the training data to estimate outcomes for future cases.

Here are several key algorithms used in supervised learning, each with its own strengths. Linear Regression is one of the simplest approaches, demonstrating the relationship between a dependent factor and one or more independent factor under the belief of a linear relationship. This technique is widely used for forecasting continuous results. Decision Trees provides a more intuitive method by splitting data into subsections based on different standards, creating a tree-like structure that helps in making choices or forecasts. These trees handle both definite and numerical data efficiently. Neural Networks, on the other hand, consist of interconnected layers of nodes that mimic the human brain's processing competences. They are predominantly adept at catching complex patterns and associations within the data, making them appropriate for tasks demanding deep learning and complex feature connections. Composed, these systems provide a robust toolkit for addressing various predictive modeling tests.

Data-Driven Approaches to Commodity Price Prediction

Types of Data for Price Prediction

Commodity price forecast trusts on various types of data to produce accurate predictions. Historical price data forms the basis of analytical models, as it offers understandings into past price trends and patterns. Observing this data helps in identifying recurrent tendencies, seasonality, and long-term movements. External pointers, such as inflation,GDP development rates and interest rates, also play a dynamic role in forecasting product values. These displays influence the overall economic atmosphere, affecting supply and demand dynamics and accordingly, commodity prices. Weather data is mainly important for agricultural merchandises, as climatic conditions directly impact crop yields and quality. Variables such as heat, rainfall, and moisture can meaningfully affect production levels and, consequently, market prices. Geopolitical factors, including political steadiness, trade policies, and global relations, further impact product prices by inducing supply chains, trade flows, and market feelings. Incorporating these diverse data types helps create an inclusive model that captures the complicated nature of commodity price actions.

Innovative Data Preprocessing Techniques

Actual data preprocessing is critical for creating robust commodity price forecast models. Feature engineering includes creating new features or altering existing ones to increase the projecting power of the model. This process may include developing features from raw data, such as calculating moving averages, decomposing time series data into trend and seasonal components. Standardization is another essential preprocessing step, which ensures that different features contribute equally to the model by scaling data to a standard range. This process stops features with larger varieties from ruling the learning process and helps advance model junction and presentation. Management of missing data is also crucial as partial data can lead to biased or inaccurate forecasts. Methods such as imputation, where missing values are estimated based on other available information, or using processes that can handle missing values directly, are normally employed. Data amplification involves generating supplementary data from existing data to increase model robustness and simplification. This method can include adding noise, creating synthetic data points to expand the dataset and improve model presentation.

Feature Selection and Importance

Feature selection is a pivotal step in data-driven approaches to commodity price prediction, as it involves identifying and retaining the most influential features that significantly impact price outcomes. Theoretical exploration of feature selection includes several methods and principles to determine feature importance. Statistical techniques, such as correlation analysis, assess the relationship between features and the target variable, helping identify features with the strongest associations. Machine learning algorithms like Decision Trees and Random Forests provide insights into feature importance by evaluating how each feature contributes to reducing prediction error. Regularization methods such as Lasso (L1 regularization) and Ridge (L2 regularization) can also be used to shrink less important features, effectively performing feature selection as part of the model training process. Wrapper methods, which involve selecting features based on model performance, and filter methods, which evaluate features independently of any model, offer additional strategies for identifying key predictors. Combining these approaches allows for a thorough exploration of features, ensuring that only the most relevant variables are included in the predictive model.

Overall, data-driven approaches to commodity price prediction encompass a broad range of data types and preprocessing techniques. By integrating historical, macroeconomic, weather, and geopolitical data, and employing innovative preprocessing methods, accurate and reliable models can be developed. Feature selection and importance analysis further refine these models, focusing on the most significant predictors to enhance forecasting accuracy and support informed decision-making in commodity markets.

Supervised Machine Learning Models for Price Prediction

Traditional Algorithms

Linear Regression is an essential statistical method used for modeling the association between a dependent factor and one or additional independent variables. In the background of commodity price forecast, Linear Regression accepts a linear connection amongst the price of the product and its impacting factors, such as supply, demand, and external variables. The theoretical foundations involve approximating the parameters of a linear equation that reduces the sum of squared variances between the observed and projected values. This ease makes Linear Regression easy to interpret but may limit its skill to capture multifaceted, non-linear relations in product price data.

Logistic Regression Spreads Linear Regression by forecasting categorical results, often used when the goal is to categorize whether a commodity price will decrease within a definite range or exceed a verge. Although more normally applied to classification problems, Logistic Regression can be improved for price prediction by transforming continuous outcomes into categorical ones. The theoretical basis lies in modeling the probability of a particular outcome using a logistic function, which ensures that predicted values fall between 0 and 1.

Decision Trees denote a more flexible method by splitting data into subsections depending on feature values to form a tree-like assembly of choices. Each node signifies a feature test, and each division represents a result, leading to forecasts at the leaf nodes. In product price forecast, Decision Trees can handle both arithmetical and categorical statistics, making them helpful for demonstrating complex connections between features. The theoretical groundwork involves recursively segregating the data to minimalize forecast error, using metrics like entropy.

Random Forests build upon Decision Trees by creating an ensemble of multiple trees, each trained on a random subset of the data. The final prediction is made by combining the outputs of all separate trees, often through popular voting for classification or averaging for regression. This method recovers model sturdiness and accuracy by decreasing the variance and modifying over fitting. The theoretic basis lies in the standard of bagging (bootstrap aggregating), where each tree studies from a dissimilar random sample of the data.

Innovative Algorithms

Gradient Boosting is an ensemble method that builds models successively, with each new model pointing to correct the errors made by the preceding models. Gradient Boosting includes training a series of weak learners, naturally Decision Trees, where individual learner focuses on the remaining mistakes of its predecessors. The theoretical basis is based on gradient descent, where the algorithm reduces the loss function by iteratively apprising model parameters. This method often leads to superior presentation by catching complex patterns and connections in the data, making it highly active for commodity price forecast.

Support Vector Machines are dominant classifiers that work by finding the hyperplane that best divides data points of dissimilar classes in a maximum-dimensional space. For regression tasks, SVMs can be adjusted to Support Vector Regression, which aims to fit the data within a specified margin of acceptance. The theoretical basis includes solving a convex optimization problematic to maximize the margin amongst data points and the decision border. SVMs are particularly active in handling non-linear relations through kernel purposes, which can be beneficial for predicting commodity prices with complex patterns.

Deep Learning models, including Neural Networks, embody the cutting edge of machine learning methods. These models consist of multiple layers of interrelated nodes, where each layer learns ranked depictions of the data. Long Short-Term Memory systems, a type of Recurrent Neural Network, are predominantly suited for time series data due to their capability to capture temporal needs and long-term relations. The theoretical basis of Deep Learning lies in learning multifaceted, non-linear mappings from input data to forecasts through back propagation and gradient descent. Deep Learning models often outperform outdated algorithms by discovery complicated patterns in huge datasets.

Model Training and Evaluation

Model Training Practices include several critical stages to guarantee that a model completes well on hidden data. Cross-Validation is a method used to evaluate model performance by dividing the data into training and testing sets. The model is skilled on one subset and tested on alternative, and this process is frequent multiple times to get an average performance measure. This benefits in measuring the model's generalizability and evading over fitting.

Hyper parameter Tuning states to the process of enhancing the limits that control the learning process of an exemplary, such as the number of trees in a Random Forest or the learning rate in Gradient Boosting. Methods such as grid search is usually used to discover diverse hyper parameter mixtures and identify the optimum settings for model performance.

Evaluation Metrics afford quantitative actions of how well a model forecasts commodity prices. Mean Absolute Error calculates the average absolute difference between foretold and actual values, providing an upfront measure of prediction accurateness. Root Mean Squared Error highlights superior errors by squaring the modifications, which is useful for understanding the model's mistake circulation. Precision calculates the amount of correct forecasts, particularly applicable for cataloging tasks. F1 Score balances precision and recall, posing a combined measure of a model's ability to correctly identify optimistic circumstances. Each metric offers perceptions into diverse aspects of model recital, and selecting the suitable metric depends on the specific aims and necessities of the forecast task.

In swift, outdated and advanced machine learning algorithms each offer unique welfares for commodity price forecast. While traditional approaches like Linear Regression and Decision Trees provide foundational methods, advanced techniques such as Gradient Boosting, SVMs, and Deep Learning models offer improved predictive competences. Active model training and assessment developments ensure that these replicas are robust and precise, contributing to more reliable commodity price forecasts.

Challenges

Machine learning models used for product price forecast encounter numerous key contests. Over fitting happens when a model does extraordinarily well on training data but ailing on new data due to its unnecessary difficulty. Contrariwise, under fitting occurs when a model is too unsophisticated to capture the fundamental patterns, resulting in insufficient recital. Data excellence issues, such as noise, incomplete, or prejudice, further worsen these problems by misrepresenting the precision of estimates and leading to erratic predictions. Furthermore, machine learning models habitually struggle with market instability and unpredicted external issues, such as natural disasters or political disasters, which can meaningfully impact values in ways not replicated in past data. These circumstances introduce difficulties and worries that can limit the efficiency of predictive mock ups, making it difficult to attain robust and dependable forecasts in dynamic and erratic surroundings.

Implications

The work validates that supervised machine learning meaningfully improves commodity price forecast-precision by catching complex patterns and adjusting to altering market circumstances. It highlights the benefits of incorporating diverse data sources and using innovative procedures over outdated models, which often fight with volatility and non-linear associations. These insights provide valuable tools for dealers, investors, and policymakers, enlightening decision-making and risk administration in the face of dynamic and erratic commodity marketplaces.

Findings and Suggestions

Managed machine learning gives substantial improvements in commodity price forecast by processing huge datasets and adjusting to complex market circumstances. While traditional models provide introductory approaches, advanced algorithms improve accuracy and strength. Speaking encounters like over fitting, data excellence, and external surprises is crucial for enlightening projecting presentation. Forthcoming research should emphasis on participating diverse data sources, decontaminating preprocessing methods, and exploring progressive algorithms to improve projecting reliability in volatile markets.

Conclusion

In conclusion, this study highpoint the transformative influence of supervised machine learning on product price estimate. By leveraging different data sources and employing advanced systems, such as Gradient Boosting and Deep Learning, the study validates significant progresses in forecasting precision compared to traditional approaches. The incorporation of historical prices, external indicators, and external influences, coupled with robust model training and estimation methods, provides a more complete and adaptive method. This improves the ability of dealers, investors, and officials to navigate complex and unpredictable commodity markets efficiently, offering an extensive progression in predictive analytics for merchandises.

References

1. Amin, M N. (2020, December 20). Predicting Price of Daily Commodities using Machine Learning. https://doi.org/10.1109/3ict51146.2020.9312012.
2. An, J., Mikhaylov, A., & Moiseev, N. (2019, September 1). Oil Price Predictors: Machine Learning Approach. EconJournals, 9(5), 1-6. https://doi.org/10.32479/ijeep.7597.
3. Dhandayuthapani, S P., & Sudha, H. (2018, June 30). A Study on Performance of Commodities Derivatives Trading in India. Rekha Patel, Volume-2(Issue-4), 502-507. https://doi.org/10.31142/ijtsrd10926.
4. Dimitriadou, A., Gogas, P., Papadimitriou, T., & Plakandaras, V. (2018, October 13). Oil Market Efficiency under a Machine Learning Perspective.
5. Flarence, A. R., Bethu, S., Sowmya, V., Anusha, K., & Babu, B. S. (2018). Importance of supervised learning in prediction analysis. Periodicals of Engineering and Natural Sciences (PEN), 6(1), 201. https://doi.org/10.21533/pen.v6i1.288.
6. F.Y, O., J.E.T, A., Awodele, O., O, H J., Olakanmi, O D., & Akinjobi, J. (2017, June 25). Supervised Machine Learning Algorithms: Classification and Comparison. Seventh Sense Research Group, 48(3), 128-138. https://doi.org/10.14445/22312803/ijctt-v48p126.
7. Manokhin, V. (2022, April 18). Application of machine learning techniques to the analysis and forecasting of commodity prices. European Organization for Nuclear Research. https://doi.org/10.5281/zenodo.6467108.
8. Mehtab, S., Sen, J., Dutta, A., & Department of Data Science and Artificial Intelligence, Praxis Business School. (2020). Stock price prediction using machine learning and LSTM-Based deep learning models. In Department of Data Science and Artificial Intelligence, Praxis Business School [Journal-article]. https://arxiv.org/ftp/arxiv/papers/2009/2009.10819.pdf.
9. Oktoviany, P., Knobloch, R., & Korn, R. (2021, November 18). A machine learning-based price state prediction model for agricultural commodities using external factors. Springer Science+Business Media, 44(2), 1063-1085. https://doi.org/10.1007/s10203-021-00354-7.
10. Panella, M., Barcellona, F., & D'Ecclesia, R L. (2012, December 31). Forecasting Energy Commodity Prices Using Neural Networks. Hindawi Publishing Corporation, 2012, 1-26. https://doi.org/10.1155/2012/289810.
11. Schofield, N C. (2021, May 7). Fundamentals of Commodities and Derivatives. , 1-17. https://doi.org/10.1002/9781119349242.ch1.
12. Sehgal, S., Rajput, N., & Dua, R K. (2012, August 10). Price Discovery in Indian Agricultural Commodity Markets., 2(2), 34-34. https://doi.org/10.5296/ijafr.v2i2.2224.
13. Seilan, A. (2012, January 1). Performance Analysis of Indian Commodity Market.