Smart Cumulative Stock Forecasting System

**Abstract**

Managing stock is the most critical thing for companies. Companies need to be very careful in monitoring their stock tracking. If there is a mistake in setting up the inventory tracking, it can lead to a challenging situation for the companies to rectify. In this study, the performance of Lime and Shap methods in classifying stock forecasting systems has been evaluated. Lime provides an interpretable method compatible with regression models to explain feature weights, while Shap offers a comprehensive interpretability framework to assess the contributions of features to model predictions.

Keywords: Cumulative Stock,, LIME, Shap, Machine Learning, Explainable Artificial Intelligence(XAI)

**Introduction**

Efficient stock management is vital for companies across industries to ensure smooth operations and meet customer demands. Accurate tracking and forecasting play a pivotal role in achieving this goal. The integration of intelligent systems, leveraging technologies like artificial intelligence and machine learning, has become increasingly popular for enhancing stock management practices. By intelligent systems, companies can improve their supply chain, production processes, and overall profitability.

In this study, we focus on evaluating the effectiveness of two prominent methods, Lime and Shap, in enhancing the interpretability and reliability of stock forecasting models. Lime offers interpretability for classification and regression models, allowing stakeholders to gain insights into feature weights and their impact on predictions. On the other hand provides, Shap a comprehensive interpretability framework, enabling the assessment of feature contributions to model predictions.

Through a systematic and critical analysis, our study aims to provide valuable insights that can significantly contribute to the advancement of intelligent stock management systems. By understanding the performance of Lime and Shap, companies can make better-informed decisions and minimize stock-related challenges. As a result, our research aims to assist businesses in optimizing their stock management strategies.

**Objective**

The purpose of this report is to examine the objectives and effectiveness of explainable artificial intelligence methods, namely LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations), used in the process of developing data-driven classification models for monitoring stock. This approach, which aims to optimize the ideal stock level that needs to be maintained for companies.

Explainable artificial intelligence addresses a common issue with machine learning models today: the difficulty of explaining and understanding their decisions. Understanding and explaining the model's predictions is critical for both applied engineering and end-users when developing data-driven classification models for stock prediction systems. At this point, explainability techniques such as LIME and SHAP make complex models more understandable and transparent, facilitating the process of making reliable and accurate decisions.

The main goal of this report is to perform a data-driven analysis using LIME and SHAP methods to understand and explain the predictions of the regression models developed on the stock dataset. While LIME focuses on individual data instances to understand why a specific prediction was made, SHAP provides a Shapley value-based approach to explain the contribution of each feature to the prediction. By using these two methods together, we aim to determine which features and instances have a significant impact on the predictions of the classification and regression models for stock values and enhance the reliability of these models

**The Innovations We Bring**

* Using LIME with Regression and Classification Models
* Using SHAP with Regression and Classification Models

**Previous Studies**

* From Local Explanations to Global Understanding with Explainable AI for Trees" - Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan H. Baumer, and Su-In Lee (2020)
* Explainable AI for Practitioners Designing and Implementing Explainable ML Solutions Michael Munn & David Pitman Foreword by Ankur Taly

**Dataset**

The data set used in this study is an time series and titled “Daily Flour Amount Tracking” This dataset contains real-world data collected to detect the occupancy status of amount of flour an company has. It is presented in tabular form and includes various features obtained from sensors that measure different time intervals and are used to determine the occupancy status.

**LIME and SHAP Explanatory Analyses**

The Lime and Shap methods were used to understand the internal workings of the regression models and determine which features have a more significant impact on the predictions. Lime provides a local explanation of the model by explaining the predictions of selected instances. On the other hand, Shap presents a global explanation by quantifying the contribution of each feature in the dataset to the model's predictions.

Through Lime and Shap analyses, we identified the important features in the predictions of each classification model and their respective contributions. These explainability analyses offer a valuable perspective to comprehend which features are more influential in stock managing prediction.

Evaluation Metrics

The performance of the classification models was evaluated using common metrics such as accuracy and F1 score. These metrics were used to assess the accuracy and reliability of the models.

**Results**

At the beginning of the study, the stock dataset's features and structure were examined, and data preprocessing steps were completed, including feature selection. Subsequently, both traditional machine learning (ML) methods based regression models were developed. These models were trained to capture important features and patterns used to predict the performance of stock forecasting systems.

Next, the use of LIME and SHAP methods for explaining and understanding the classification models' predictions was explored. LIME created interpretable and simple models by focusing on data instances to understand why specific predictions were made. On the other hand, SHAP provided explanations by quantifying the contribution of each feature to the predictions. By using these two methods together, the process of how the stock regression model generated predictions was presented in a more understandable and transparent manner.

In conclusion, this study demonstrates that a data-driven approach for predicting systems, combined with explainable artificial intelligence techniques, enables the development of more effective and interpretable models. This explanatory analysis conducted on the daily stock dataset provides an important foundation for future research in the field and contributes to making prediction systems more efficient and reliable.

**Suggestions**

**Further development can be done regarding adapting the SHAP library to neural network models.**

**SHAP & LIME Statistics**

**SHAP**

**Fig 1 Shap Feature İmportances**

Shap feature an importance graph is a visualization that shows how each feature affects a particular prediction made by a machine learning model. It aids in comprehending the significance and influence of particular features on the output of the model.

Day numbers have a high effect on the algorithm, as expected, and algorithm shape values are becoming more powerful day by day. **(Fig 1)**

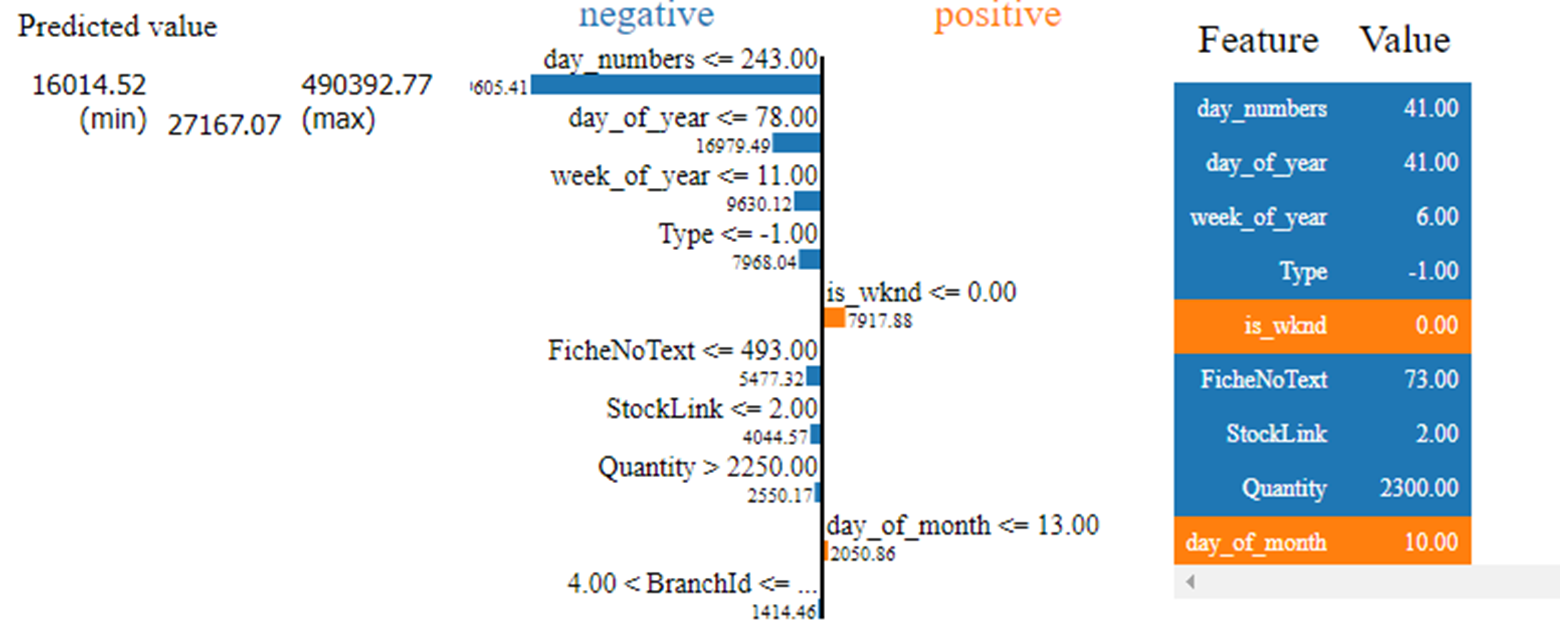
**Fig 2 Shap Values Graph**

This graph illustrates the cumulative impact of daily numbers some values have double colors because our data set has multiple types in a day. **(Fig 2)**

**Fig 3 Shap Waterfall Graph**

This shape graph shows the difference between f(x) and E[f(x)] values for a fancy column, where f(x) represents our test data set's split before values. The shap value of the original data set is E[f(x)]. **(Fig 3)**

**Lime**

**Fig(4 Lime Predicted Values Effects)**

In this lime output, a column is selected for modeling, and this column is calculated by the lime algorithm. They have negative or positive values, depending on how they affect the predicted value. Is it increasing or decreasing. **(Fig 4)**