**Cross Validation Image**  
  
  
  
The implementation of a manual cross-validation technique tailored for time series data analysis of Bitcoin prices, using a rolling window approach rather than random splitting due to the sequential nature of the dataset. Specifically, the script establishes a series of temporal folds, each encompassing 363 days, against a substantial training set size of 3000 days, ensuring that a complete integer number of folds is utilized without exceeding the available data.

The image shows five distinct cross-validation folds, each divided into training and test segments—the training segment is illustrated with a blue line, while the test segment is depicted in red. This clear segmentation is indicative of the model's training phase on historical data (blue) before being subjected to the prediction phase (red), providing a pragmatic framework to evaluate the model's predictive accuracy on data it has not previously encountered.

A close examination of the visualized data across the five folds reveals notable price volatility within the Bitcoin market, characterized by drastic fluctuations and several prominent peaks, each of which is captured within different folds. This variability is crucial for assessing the model's robustness, as it ensures that the model's predictive capabilities are tested against a diverse array of market conditions and behaviors, ranging from steady inclines to abrupt declines. The ability of the model to adapt to these conditions can significantly inform its reliability and effectiveness in real-world applications.

**Random Forest image**  
  
The evaluation of a Random Forest model applied to Bitcoin price prediction is presented, illustrating the model's varying degree of success across different temporal segments. The graph in the study delineates a comparison of actual and predicted Bitcoin prices, with the model achieving an average Mean Absolute Percentage Error (MAPE) of 26.57, indicating a moderate prediction accuracy, and an average Mean Squared Prediction Error (MSPE) of 12.88, reflecting the variance of the predictions from the actual values. Notably, the model's Akaike Information Criterion (AIC) averages at 34613.21, serving as a gauge for the model's relative quality. The predictive trajectory closely mirrors the actual price movements, though it struggles with the market's pronounced peaks and troughs, underscoring the inherent volatility and unpredictability of cryptocurrency markets.

A detailed cross-validation, represented in the accompanying table, reveals that the model's performance fluctuates across folds; Fold 3 exhibits the highest MAPE, indicating lower predictive accuracy, whereas Fold 5 boasts the lowest MAPE and AIC, suggesting a superior model fit and predictive precision for that segment. These variations underscore the challenges in forecasting financial time series data, where external factors often induce significant predictive discrepancies.

**XG Boost Image**

the XG Boost model applied to Bitcoin price predictions indicate that the model has a close following of the actual price movements, denoted by the red line, with its predictions. The model records an average Mean Absolute Percentage Error (MAPE) of 25.99, reflecting a moderate level of accuracy in its predictions. The Mean Squared Prediction Error (MSPE) stands at 12.68, which points to the model's ability to predict with a reasonable degree of precision, and the average Akaike Information Criterion (AIC) is 34660.88, suggesting the model's goodness of fit to the data.

The table provides a breakdown of performance metrics across five different folds, with Fold 3 showing a notably higher MAPE, indicating that the predictions for this particular fold were less precise relative to the actual values. In contrast, Fold 5 demonstrates the lowest MAPE and a relatively lower AIC, indicating a more accurate set of predictions and a better fit model for the data within this fold.

The graph reflects the model's challenge in capturing the extreme peaks and troughs of the Bitcoin price, a common difficulty in the volatile cryptocurrency market. These findings emphasize the complexities of financial time series forecasting and the need for robust modeling techniques to handle the unpredictable nature of such data.

**Lagged Information**

In the construction of the time series predictive model for Bitcoin pricing, the introduction of lagged features constitutes a critical enhancement to the analytical framework. Specifically, a lagged feature was generated by shifting the price data by one day, thus creating a new column Lag\_1d in the dataset. This methodological step is imperative for encapsulating the temporal autocorrelation characteristic of financial time series, where previous values bear predictive power over future values.

Subsequent to the introduction of Lag\_1d, the dataset was purged of rows containing NA values, which resulted from the lagging process. These NA entries, indicative of initial time points lacking historical data, were removed to preserve the data's integrity. This cleansing step ensures that the model's inputs are devoid of potential biases or errors that could compromise the validity of subsequent predictive insights.