Forecasting Dogecoin Price with Time Series Data

H. Majeed, S. Bukhari, S. Gohar

November 9, 2025

1. Introduction

Forecasting cryptocurrency has been an interesting topic for many time-series analysts due to high fluctuations and higher reward metrics. This report examines the daily closing price of Dogecoin, exploring the trends, seasonality, and residual patterns. The dataset used is from OpenML[1], it contains 1532 values of daily opening price, high price, low price, closing price, and volume for Dogecoin cryptocurrency from November 2017 to January 2022. The data distribution is presented in Table 1 with the price fluctuations being mere decimals.

	Open	High	Low	Close	Volume
count	1532	1532	1532	1532	1532
mean	0.053032	0.056574	0.049593	0.053151	1049953580
std	0.106257	0.114770	0.097944	0.106358	3809881008
min	0.001046	0.001210	0.001002	0.001038	1431720
max	0.687801	0.737567	0.608168	0.684777	69410680685

Table 1. Distribution of Dogecoin features

2. Data Visualization and Exploratory Analysis

2.1 Time-Series Visualization

From the time-series graph presented in Figure 1. the closing price remained steady up until 2021, where it experienced a sharp rise reaching its peak value between Feb-Mar 2021. This was followed by a sharp decline until mid 2021, where the price showed gradual fluctuations. This shows an upward trend marked by significant speculative bubbles, which is typical of cryptocurrencies. It also can be seen that reaching a higher price point also makes the price more volatile as can be seen from the peak size.

[1] https://www.openml.org/search?type=data&status=active&id=43472&sort=runs

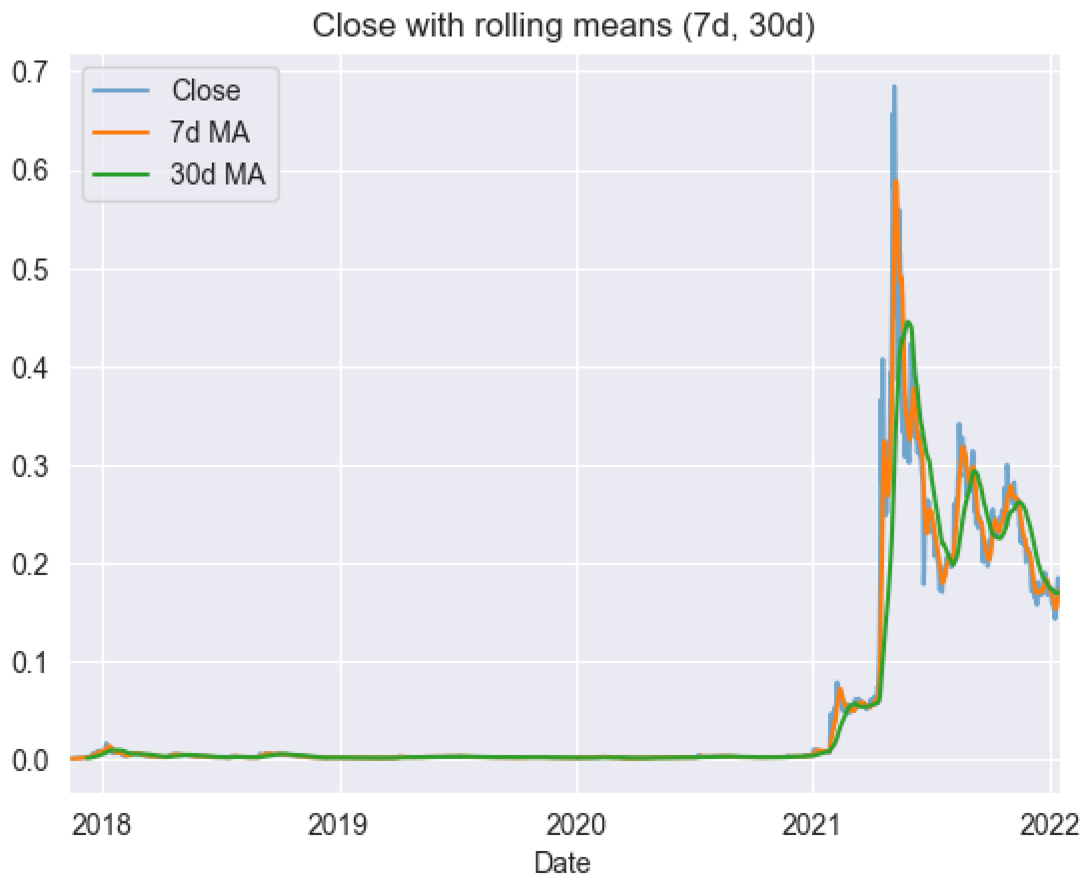

Figure 1. Time-series graph of closing price for Dogecoin 2018-2022

2.2 Time-Series Decomposition Analysis

To isolate the components which influence the price, we applied Seasonal-Trend-Decomposition using LOESS (STL) method which decomposes the data into season, trend, and residue. The trend shows a non-linear growth especially during the covid-19 period. This trend is not just a straight line which hints towards the use of non-linear machine learning models to learn the relationship of closing price.

The seasonal component, which was calculated based on a 7-day period, shows consistent low magnitude cyclic pattern, this shows that there are external deterministic factors that which introduces predictability into the price movements.

Since the residuals are not uniformly distributed and contain high magnitude spikes during the high price time-period, this indicates that simple time-series model like ARIMA may struggle and deep learning models like RNN, LSTM, Transformer that handles non-linearity and are well suited for capturing dependencies will be necessary for accurate forecasting.

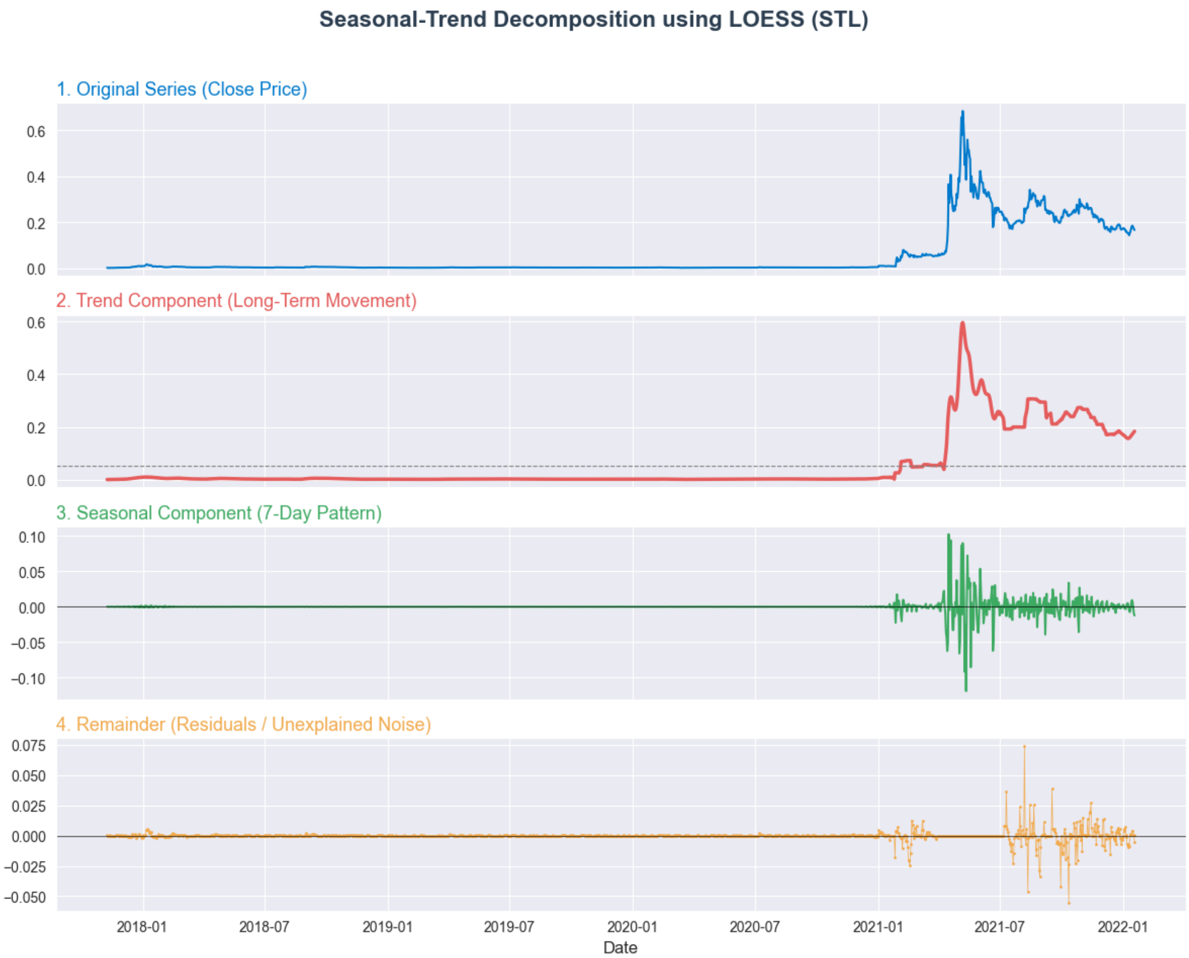

Figure 2. STL components visualization

2.3 Autocorrelation Analysis

We used the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to understand how the current price is correlated with its own past prices. The ACF bars remain outside the blue shaded confidence interval for many lags and shrinks very slowly. The slow, gradual decrease of the ACF is an indicator of a time series with a strong trend.

The PACF plot shows a large, significant spike at Lag 1, and then rapidly drops remaining statistically insignificant for subsequent lags. The strong correlation at Lag 1 indicates that the price at time t-1 has a direct and highly significant impact on the price at time t.

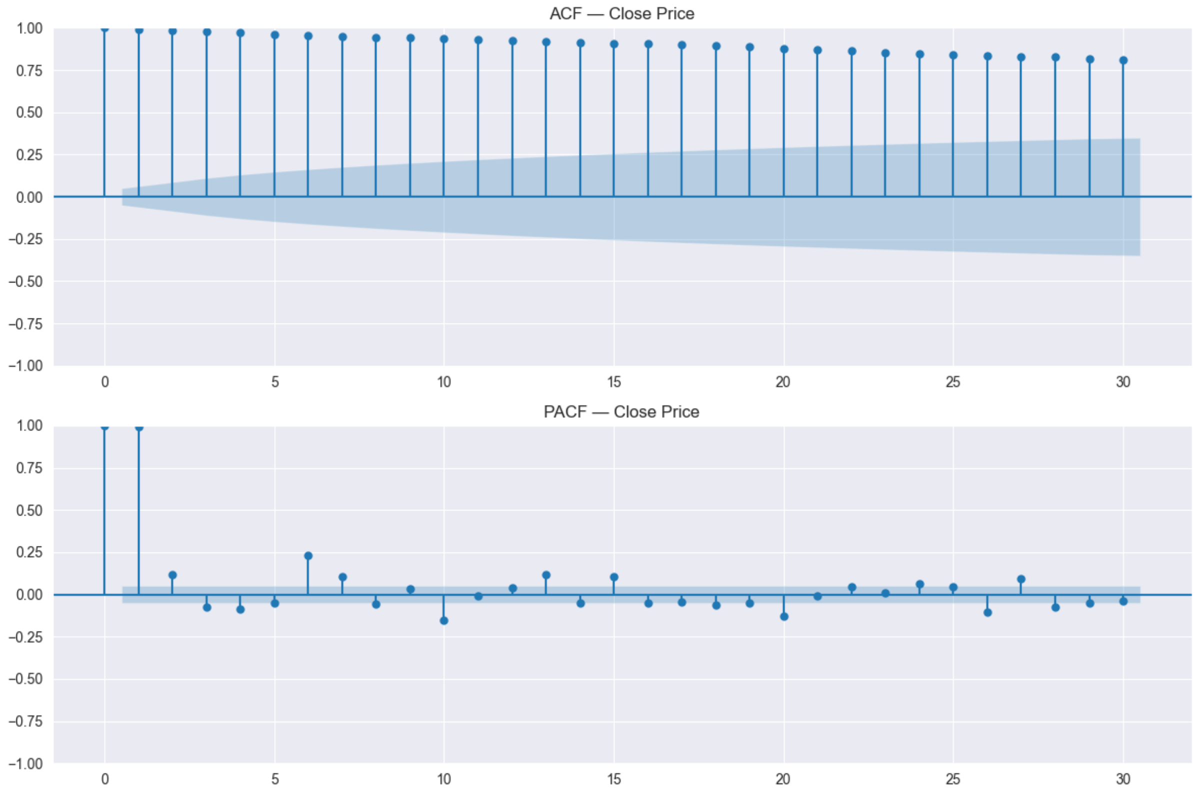

Figure 3. Autocorrelation plots of closing price

3. Plan for Data Partitioning

In time-series forecasting, partitioning while making sure chronology is preserved, and preventing data leakage where future information is inadvertently used to train a model is of highest important to ensure proper model training. The dataset will be split strictly by date. The oldest data will form the training set, and the newest data will be held out for testing. The earliest 70% of the data, will be used to fit the model's parameters. The next 15% of the data, would be used for tuning hyperparameters. The final 15% of the data (the most recent period), which will be used to provide an unbiased assessment of the model's forecasting performance on unseen data.

Given the highly volatile and dynamic nature of cryptocurrency, an even more robust evaluation would involve a rolling or expanding window cross-validation technique for hyperparameter tuning. This mimics a real-world trading scenario by continuously expanding the training set and re-evaluating the model monthly or quarterly on the most recent period. For this initial submission a fixed chronological split of 70/15/15 will provide a necessary baseline.