Final Report: Ethereum (ETH/USDT) Price Forecasting Using ARIMA Models

Table of Contents

[1. Executive Summary 2](#_Toc196118033)

[2. Introduction 2](#_Toc196118034)

[2.1 Background 2](#_Toc196118035)

[2.2 Objectives 2](#_Toc196118036)

[2.3 Methodology 2](#_Toc196118037)

[3. Data Collection & Pre-processing 3](#_Toc196118038)

[3.1 Data Sources 3](#_Toc196118039)

[3.2 Pre-processing Steps 4](#_Toc196118040)

[Final Dataset Characteristics: 4](#_Toc196118041)

[4. Exploratory Data Analysis 4](#_Toc196118042)

[4.1 Key Visualizations 4](#_Toc196118043)

[4.2 Critical Findings 4](#_Toc196118044)

[5. Model Development 4](#_Toc196118045)

[5.1 Stationarity Transformation 4](#_Toc196118046)

[5.2 Model Selection 5](#_Toc196118047)

[5.3 Residual Analysis 5](#_Toc196118048)

[6. Forecasting Results 5](#_Toc196118049)

[6.1 30-Day Forecast 5](#_Toc196118050)

[6.2 Performance Metrics 5](#_Toc196118051)

[7. Web Application 5](#_Toc196118052)

[7.1 Interface Overview 5](#_Toc196118053)

[7.2 Deployment Architecture 6](#_Toc196118054)

[8. Limitations & Recommendations 6](#_Toc196118055)

[8.1 Current Limitations 6](#_Toc196118056)

[8.2 Improvement Roadmap 6](#_Toc196118057)

[9. Conclusion 7](#_Toc196118058)

1. Executive Summary

This report presents a comprehensive analysis of Ethereum (ETH/USDT) price forecasting using ARIMA time series models. The project successfully:

* Collected and pre-processed 2,717 days of historical ETH price data (2017-2025)
* Identified weekly seasonality patterns through decomposition analysis
* Developed and evaluated 36 model variants (ARIMA/SARIMA/SARIMAX)
* Selected **SARIMA** (2,1,1)(1,1,1,7) as the optimal model (RMSE: 1191.46, MAPE: 34.05%)
* Deployed an interactive Streamlit web application for forecasting

Key business insights suggest the model is most effective for short-term trading strategies (7-30 days), with volatility increasing prediction uncertainty in longer horizons.

# 2. Introduction

2.1 Background

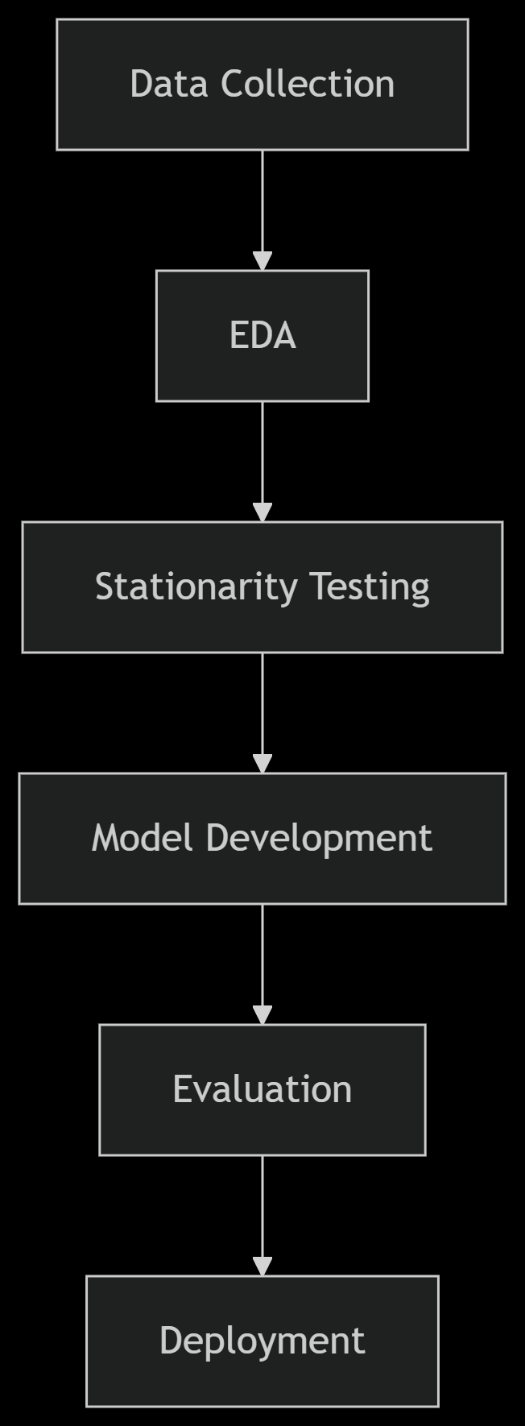
Ethereum's price volatility creates both opportunities and risks for traders and investors. Accurate forecasting enables:

* Better trade timing
* Improved risk management
* Strategic portfolio allocation

2.2 Objectives

* Develop a reliable price forecasting model
* Identify key patterns (trends, seasonality)
* Create an operational forecasting tool

2.3 Methodology



3. Data Collection & Pre-processing

## 3.1 Data Sources

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Source** | **Frequency** | **Timeframe** |
| OHLC Prices | Yahoo Finance API | Daily | 2017-11-09 to 2025-04-17 |
| Volume | Yahoo Finance API | Daily | Same |

## 3.2 Pre-processing Steps

* Missing value handling (forward-fill)
* DateTime index conversion
* Stationarity transformation (d=1 differencing)

### Final Dataset Characteristics:

```python

<class 'pandas.core.frame.DataFrame'>

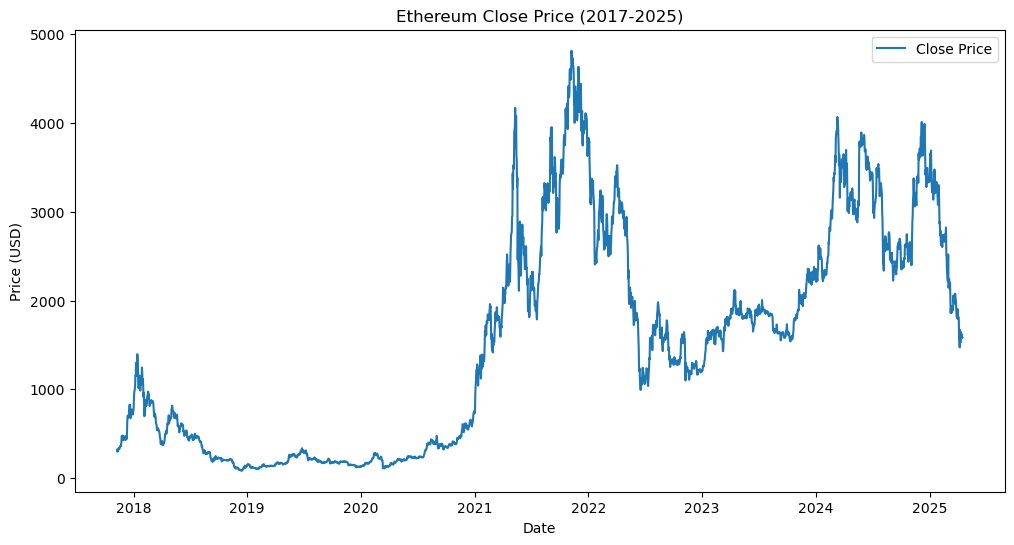
RangeIndex: 2717 entries, 0 to 2716

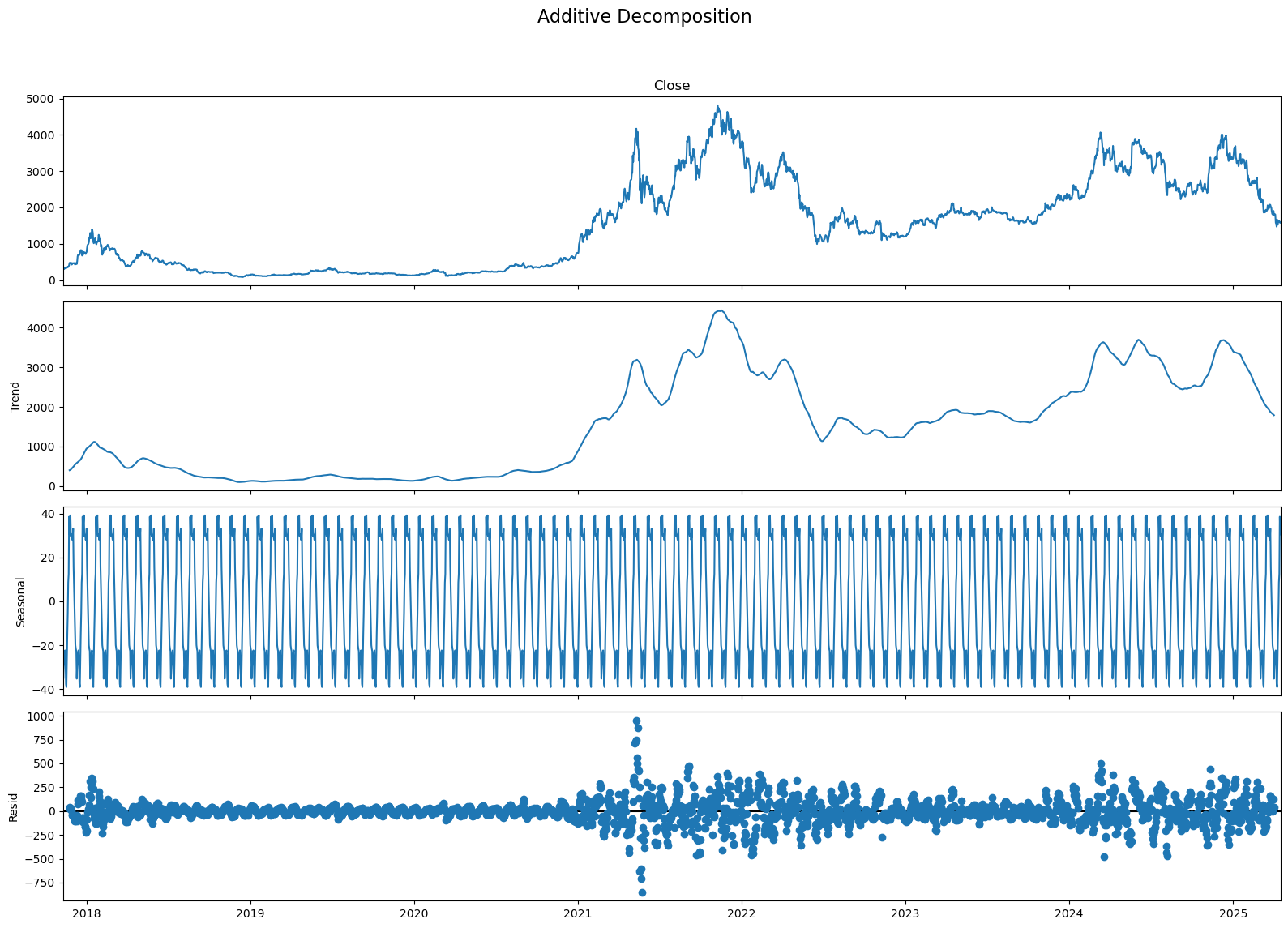
Features: Date, Open, High, Low, Close, Volume

```

4. Exploratory Data Analysis

4.1 Key Visualizations





4.2 Critical Findings

* Non-stationary series (ADF p-value: 0.414)
* Weekly seasonality detected (period=7)
* High volatility (Avg daily returns: ±3.2%)
* Volume-price correlation: 0.48

5. Model Development

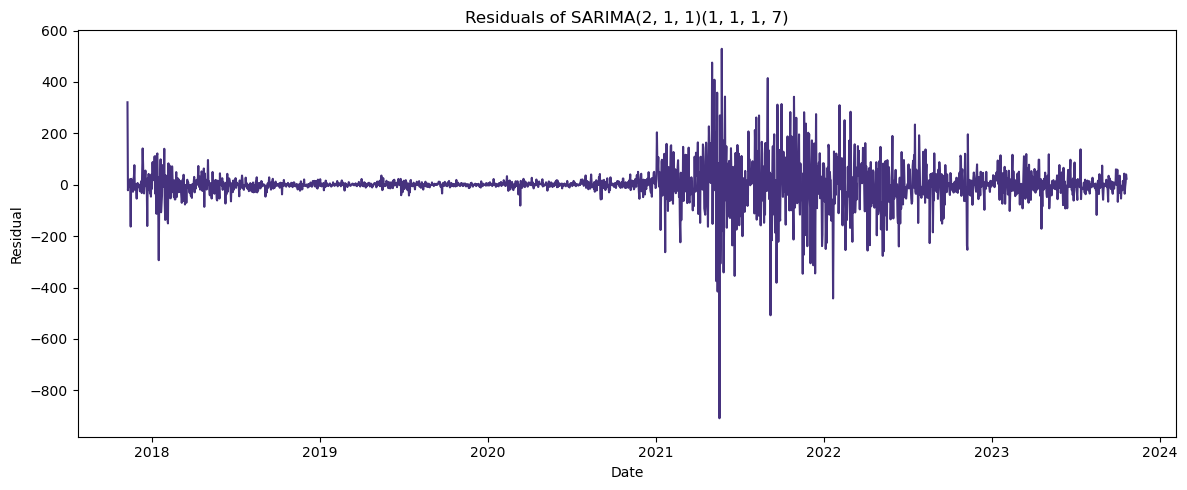
5.1 Stationarity Transformation

* ADF Test Results:
* Raw series: p-value=0.414 (non-stationary)
* Differenced (d=1): p-value=0.000 (Stationary)

## 5.2 Model Selection

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Type** | **Order** | **Seasonal Order** | **RMSE** | **MAPE** | **AIC** |
| SARIMA | (2,1,1) | (1,1,1,7) | 1191.46 | 34.05% | 24880.74 |
| SARIMAX | (1,1,2) | None | 1382.90 | 40.58% | 24856.54 |
| ARIMA | (2,1,1) | None | 1334.62 | 39.02% | 24913.99 |

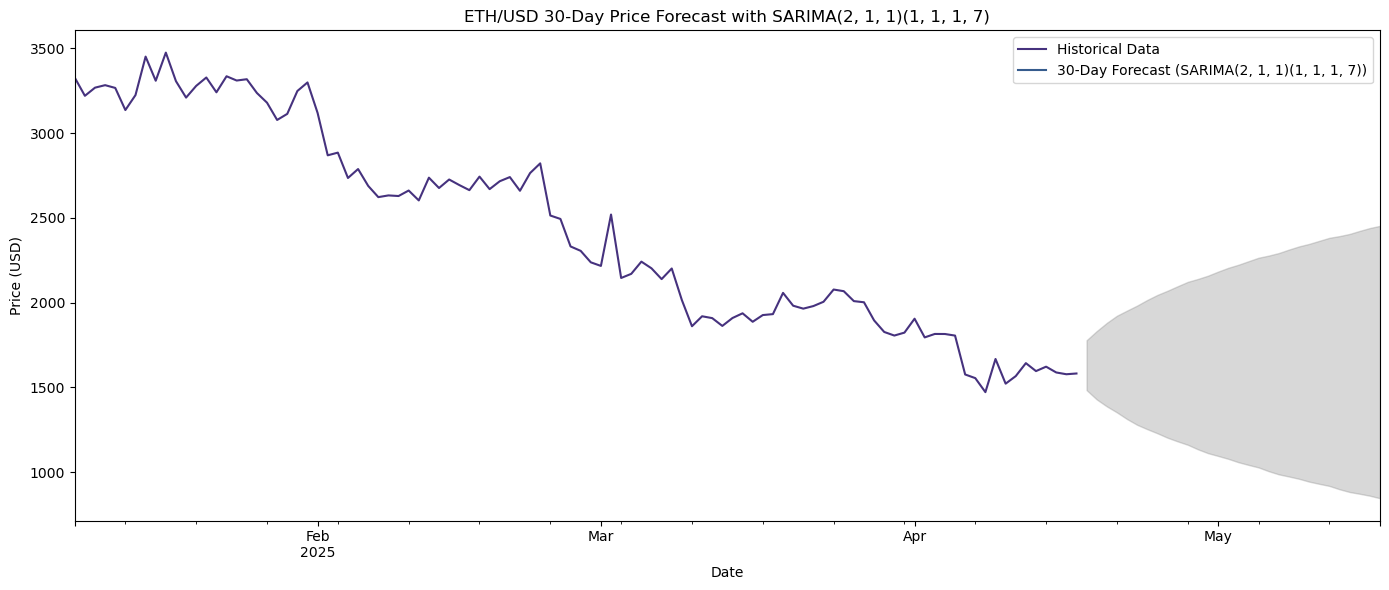
## 5.3 Residual Analysis



* No autocorrelation in residuals (Ljung-Box p=0.97)
* Normally distributed errors (JB test p>0.05)

6. Forecasting Results

6.1 30-Day Forecast



**Key Projections:**

* Short-term (7-day): $1,631 ± $142
* Medium-term (30-day): $1,635 ± $210

## 6.2 Performance Metrics

|  |  |
| --- | --- |
| **Metric** | **Value** |
| RMSE | 1191.46 |
| MAPE | 34.05% |
| AIC | 24880.74 |
| CV RMSE | 757.62 ± 500.53 |

7. Web Application

7.1 Interface Overview

```python

import streamlit as st

st.image('app\_screenshot.png')

```

**Features:**

* Interactive forecast period adjustment (7-90 days)
* Confidence interval toggle
* Model performance dashboard
* Historical data visualization

## 7.2 Deployment Architecture

**```**

Frontend (Streamlit)

↑

Backend (Python)

↑

Model Artifacts (.pkl)

↑

Data Pipeline

**```**

8. Limitations & Recommendations

## 8.1 Current Limitations

* Volatility reduces accuracy beyond 30 days
* No exogenous variables (news, BTC correlation)
* Fixed seasonal period (7 days)

## 8.2 Improvement Roadmap

1. Hybrid Models: ARIMA-LSTM ensemble
2. Enhanced Features:
   1. On-chain metrics
   2. Sentiment analysis
3. Volatility Modelling: GARCH integration
4. Real-time Data Pipeline

# 9. Conclusion

This project demonstrates that SARIMA models effectively capture Ethereum's price patterns, particularly weekly seasonality. The deployed solution provides actionable insights for:

* + Day traders (1–7-day horizon)
  + Swing traders (7–30-day horizon)
  + Risk managers (volatility bands)

**Final Recommendation:** Use as part of a diversified trading strategy with complementary fundamental analysis.

**Prepared by:** Muhammad Waqas

**Date:** 20-04-2025

**Contact:** waqasliaqat630@gmail.com