**Crypto and Gold Price Analysis and Forecasting with Machine Learning (2015–2025)**

*A Short-Term Forecasting Attempt in Bitcoin, Gold, and Cryptocurrency Markets*

Prepared by: Yunus Ahmet DOKAZOĞLU

**CONTENTS**

**1. Introduction**...................................................................................................................4

1.1 Background.............................................................................................................. 4

1.2 Importance of the Study............................................................................................ 4

1.3 Literature Summary.................................................................................................. 5

1.4 Aim of the Study....................................................................................................... 5

1.5 Contributions............................................................................................................ 5

**2. Dataset**........................................................................................................................ 6

2.1 Data Source........................................................................................................... 6

2.2 Assets and Period................................................................................................. 6

2.3 Preprocessing Steps................................................................................................. 6

**3. Method**.......................................................................................................................... 7

3.1 Feature Design…………......................................................................................... 7

3.2 Labeling ……………............................................................................................... 7

3.3 Train-Test Split…..................................................................................................... 8

3.4 Models Used………................................................................................................. 8

**4. Models and Results** ...................................................................................................... 8

4.0 Common Preparation................................................................................................ 8

4.1 Logistic Regression................................................................................................. 11

4.1.1 Model Setup……............................................................................................. 11

4.1.2 Outputs ............................................................................................................ 11

4.1.3 Analysis and Interpretation.............................................................................. 11

4.2 KNN (K-Nearest Neighbors).................................................................................. 11

4.2.1 Model Setup…….. ........................................................................................ 11

4.2.2 Outputs ........................................................................................................... 12

4.2.3 Analysis and Interpretation.............................................................................. 12

4.3 SVM (RBF Kernel) .............................................................................................. 12

4.3.1 Model Setup……… ....................................................................................... 12

4.3.2 Outputs ........................................................................................................... 12

4.3.3 Analysis and Interpretation.............................................................................. 13

4.4 Random Forest ..................................................................................................... 13

4.4.1 Model Setup…………..................................................................................... 13

4.4.2 Outputs ........................................................................................................... 14

4.4.3 Analysis and Interpretation.............................................................................. 14

4.5 All Assets × All Models (Batch Run)………………............................................. 14

4.5.1 CSV Çıktısı .....................................................................................................14

4.5.2 Interpretation and Summary Tables................................................................ 15

**5. Conclusion and Evaluation**....................................................................................... 16

5.1 Summary of Findings.............................................................................................. 16

5.2 Strengths of the Study…........................................................................................ 17

5.3 Limitations Encountered......................................................................................... 17

5.4 General Inferences................................................................................................. 18

**6. Application and Recommendations**..........................................................................19

6.1 Practical Application Areas of the Project.............................................................. 19

6.2 Sectoral Contributions / Usage Potential............................................................... 19

6.3 Strategic Recommendations................................................................................... 19

6.4 Potential Future Improvements…........................................................................... 20

**8. References**.................................................................................................................... 20

**INTRODUCTION**

**1.1 Background**  
In recent years, cryptocurrencies have increasingly gained attention in financial markets. The emergence of Bitcoin in 2009 led to the birth of a digital asset class alternative to traditional financial systems. In this process, many cryptocurrencies such as Ethereum, Binance Coin, Ripple, Cardano, Dogecoin, and Solana have also entered the market, attracting investors' attention. The rapid value increases, sudden drops, and generally high volatility of cryptocurrencies have made them an investment tool that harbors both opportunities and risks.

In traditional markets, gold has long been considered a "safe haven" for investors. Investors turning to gold during crisis periods has established this asset as a stable store of value. Therefore, the relationship between gold and cryptocurrencies is a highly curious topic in the financial world: Could cryptocurrencies be the new "digital gold"?

**1.2 Importance of the Study**  
Predicting the short-term price movements of cryptocurrencies is a research area of great interest both in academia and the investment sector. However, forecasting financial time series is inherently very difficult. Prices are affected by many factors such as market news, global developments, supply-demand imbalances, and investor psychology. Therefore, in addition to traditional statistical methods, machine learning techniques are increasingly being used.

This study examines the prices of Bitcoin, Ethereum, Binance Coin, Ripple, Cardano, Dogecoin, Solana, and Gold for the 2015–2025 period. The comparison between Bitcoin and Gold is particularly noteworthy; one is a high-risk and speculative asset, while the other is seen as a stable and safe haven. The aim of this study is to predict the next-day price movements using machine learning methods on these two extreme assets and other popular cryptocurrencies and to compare the performance of the models.

**1.3 Literature Summary**  
There are numerous studies in the literature aimed at predicting cryptocurrency price movements. Some research has examined long-term trends using deep learning models (e.g., LSTM, GRU). Other studies have made short-term predictions using machine learning methods based on technical indicators (RSI, MACD, moving averages). However, the vast majority of these models show limited success due to the extreme volatility of cryptocurrencies. Studies comparing them with more stable assets like gold are important for understanding investors' risk-return preferences.

This project stands out from the approaches in the literature by addressing multiple cryptocurrencies and gold within the same framework and comparing classic machine learning algorithms (Logistic Regression, KNN, SVM, Random Forest) side by side.  
  
**1.4 Aim of the Study**

This project has three main objectives:  
1. **Exploratory Analysis:** To compare the price movements, returns, and volatility characteristics of Bitcoin, Gold, and other cryptocurrencies.

2. **Forecasting Attempt:** To build classification models that predict the next-day price movement (increase/decrease) using simple technical indicators.

3. **Model Comparison:** To run methods such as Logistic Regression, KNN, SVM, and Random Forest on the same dataset and compare their performance.  
  
**1.5 Contributions**  
- The comparative analysis of cryptocurrencies and gold price behavior over the same period provides important clues about investor behavior.  
- Achieving a certain level of accuracy (65%–70%) even with simple technical indicators demonstrates the usability of machine learning models in financial markets.  
- The study contributes to the literature by revealing which models perform better on up-days and which perform better on down-days.

In conclusion, the introduction part of this study has laid out the importance of cryptocurrencies and gold in financial markets, why short-term price forecasting is difficult but valuable, the rationale for the methods used, and the research objectives. The next section will introduce the dataset used in the study in detail and explain the methods applied for the analysis.

**2. DATASET SECTION**

**2.1 Data Source**

The dataset used in this study was obtained from the open-source dataset titled “Crypto and Gold Prices Dataset (2015–2025)” published on Kaggle. The dataset includes daily price information and offers the opportunity to compare the long-term price movements of cryptocurrencies and gold.

**2.2 Assets and Period**

The dataset contains the daily closing prices for the following assets :

* Bitcoin (BTC)
* Ethereum (ETH)
* Binance Coin (BNB)
* Ripple (XRP)
* Cardano (ADA)
* Dogecoin (DOGE)
* Solana (SOL)
* Altın (Gold, USD/ons)

The covered period is from January 2015 to January 2025. This includes both the rise periods of cryptocurrencies (2017, 2021 bull markets) and the decline periods (2018, 2022 bear markets).

**2.3 Preprocessing Steps**

The basic preprocessing steps performed on the dataset are as follows:

* **Date Conversion:** The date column was converted to the appropriate format, and the data was sorted chronologically**.**
* **Cleaning Missing Data:** Rows with missing transaction data on certain days were removed.
* **Feature Generation:** Necessary indicators for analysis (daily return, 7-day average return, 7-day volatility, deviation from 30-day average, 10-day momentum) were calculated.
* **Labeling:** A binary label (0 = decrease, 1 = increase) was created to indicate whether the next day's price increased.

**3. METHOD**

**3.1 Özellik (Feature) Tasarımı**

Various technical indicators were generated from the daily closing prices in the dataset. The aim is to extract summary metrics from the price series that will provide information to the model.

* **ret\_1d:** Daily return, the percentage change in price compared to the previous day.
* **ret\_7d:** 7-day average return, indicates the short-term trend.
* **vol\_7d:** Standard deviation of 7-day returns, measures short-term volatility.
* **ma30\_dev:** Deviation of the current price from the 30-day moving average, shows deviation relative to the long-term trend.
* **mom\_10:** Percentage change in price compared to 10 days ago, a momentum indicator.

These features are commonly used technical analysis indicators in finance literature and can provide strong signals for predicting short-term price movements.

**3.2 Labeling**

The target to be predicted is whether the next day's price will increase.

* If next day price > today's price → 1 (Increase)
* If next day price < today's price → 0 (Decrease)

With this definition, the problem has been converted into a binary classification problem.

**3.3 Train–Test Split**

Due to the nature of time series, data cannot be randomly shuffled. Therefore:

* 2015–2021 data → Training (train)
* 2022–2025 data → Testing (validation)

Thus, a more realistic scenario was established where the model learns from the past to predict the future.

**3.4 Models Used**

Four basic machine learning algorithms were used in this project:

* **Logistic Regression:** A simple and interpretable classifier.
* **KNN (K-Nearest Neighbors):** Makes predictions by looking at similar past patterns.
* **SVM (RBF Kernel):** Powerful for complex, non-linear boundaries.
* **Random Forest:** An ensemble model composed of many decision trees; makes more flexible predictions.

Each model was run with the same features and the same train/test split, and their performances were compared.

**4. MODELS AND RESULTS**

**4.0 Common Preparation**

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.pipeline import Pipeline

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.neighbors import KNeighborsClassifier

from sklearn.svm import SVC

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report, ConfusionMatrixDisplay

**# Veri oku**

csv\_path = r"C:\Users\User\Desktop\Crypto Data Since 2015.csv"

df = pd.read\_csv(csv\_path)

df["Date"] = pd.to\_datetime(df["Date"], errors="coerce")

df = df.sort\_values("Date").dropna(subset=["Date"])

**# Özellik üretici**

def build\_features\_for(price\_series: pd.Series, dates: pd.Series) -> pd.DataFrame:

s = price\_series.copy()

ret = s.pct\_change()

feats = pd.DataFrame({

"Date": dates,

"ret\_1d": ret,

"ret\_7d": ret.rolling(7).mean(),

"vol\_7d": ret.rolling(7).std(),

"ma30\_dev": s / s.rolling(30).mean() - 1,

"mom\_10": s / s.shift(10) - 1,

})

feats["label"] = (ret.shift(-1) > 0).astype(int)

return feats.dropna().reset\_index(drop=True)

SPLIT\_DATE = "2022-01-01"

def make\_train\_test(feats: pd.DataFrame):

train = feats[feats["Date"] < SPLIT\_DATE]

test = feats[feats["Date"] >= SPLIT\_DATE]

X\_train = train[["ret\_1d","ret\_7d","vol\_7d","ma30\_dev","mom\_10"]]

y\_train = train["label"]

X\_test = test[["ret\_1d","ret\_7d","vol\_7d","ma30\_dev","mom\_10"]]

y\_test = test["label"]

return X\_train, y\_train, X\_test, y\_test

def run\_model(title, estimator, asset\_col):

feats = build\_features\_for(df[asset\_col], df["Date"])

X\_train, y\_train, X\_test, y\_test = make\_train\_test(feats)

estimator.fit(X\_train, y\_train)

y\_pred = estimator.predict(X\_test)

**# Tablo**

print(f"\n=== {title} | Asset: {asset\_col} ===")

print("Accuracy:", round(accuracy\_score(y\_test, y\_pred), 3))

print(classification\_report(y\_test, y\_pred))

**# Grafik: Confusion Matrix**

fig, ax = plt.subplots(figsize=(4.5,4))

ConfusionMatrixDisplay.from\_predictions(y\_test, y\_pred, ax=ax, cmap="Blues")

ax.set\_title(f"Confusion Matrix — {title}")

plt.tight\_layout()

plt.show()

**4.1 Logistic Regression**

4.1.1 Model Setup

log\_reg = Pipeline([

("scaler", StandardScaler()),

("clf", LogisticRegression(class\_weight="balanced", max\_iter=1000))

])

run\_model("Logistic Regression (BTC)", log\_reg, "Bitcoin (USD)")

4.1.2 Outputs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sınıf** | **Precision** | **Recall** | **F1-Score** | **Support** |
| 0 | 0.00 | 0.00 | 0.00 | 416 |
| 1 | 0.68 | 1.00 | 0.81 | 885 |
| **Accuracy** |  |  | **0.68** | 1301 |

4.1.3 Analysis and Interpretation

* Logistic Regression is strong in the increase class, unsuccessful in the decrease class.
* Accuracy is 68% but affected by imbalance
* Up-days are captured, down-days are completely missed

**4.2 KNN (K-Nearest Neighbors)**

4.2.1 Model Setup

knn = Pipeline([

("scaler", StandardScaler()),

("clf", KNeighborsClassifier(n\_neighbors=15, weights="distance"))

])run\_model("KNN (ETH)", knn, "Ethereum (USD)")

4.2.2 Outputs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sınıf** | **Precision** | **Recall** | **F1-Score** | **Support** |
| 0 (Düşüş) | 0.35 | 0.07 | 0.12 | 416 |
| 1 (Artış) | 0.68 | 0.94 | 0.79 | 885 |
| **Accuracy** |  |  | **0.66** | 1301 |

4.2.3 Analysis and Interpretation

* KNN is strong in the increase class, slightly better in the decrease class but insufficient.
* Parameter tuning (k value) might change performance
* Very few correct predictions on down-days, strong on up-days.

**4.3 SVM (Support Vector Machine)**

4.3.1 Model Setup

svm = Pipeline([

("scaler", StandardScaler()),

("clf", SVC(kernel="rbf", class\_weight="balanced", C=1.0, gamma="scale"))

])

run\_model("SVM (Gold)", svm, "Gold (USD per oz)")

4.3.2 Outputs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sınıf** | **Precision** | **Recall** | **F1-Score** | **Support** |
| 0 (Düşüş) | 0.00 | 0.00 | 0.00 | 416 |
| 1 (Artış) | 0.68 | 1.00 | 0.81 | 885 |
| **Accuracy** |  |  | **0.68** | 1301 |

4.3.3 Analysis and Interpretation

* SVM draws non-linear boundaries but produced results similar to Logistic Regression on this data.
* Fails to capture down-days.
* Same as Logistic Regression: captures up-days, fails completely on down-days..

**4.4 Random Forest**

4.4.1 Model Setup

rf = RandomForestClassifier(

n\_estimators=300, random\_state=42, class\_weight="balanced\_subsample")

run\_model("Random Forest (BNB)", rf, "Binance Coin (BNB)")

**# Özellik önemleri**

import numpy as np

feat\_names = ["ret\_1d","ret\_7d","vol\_7d","ma30\_dev","mom\_10"]

importances = rf.feature\_importances\_

order = np.argsort(importances)[::-1]

plt.bar(range(len(order)), importances[order])

plt.xticks(range(len(order)), [feat\_names[i] for i in order], rotation=30)

plt.title("Feature Importance — Random Forest (BNB)")

plt.show()

4.4.2 Outputs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sınıf** | **Precision** | **Recall** | **F1-Score** | **Support** |
| 0 (Düşüş) | 0.28 | 0.07 | 0.11 | 416 |
| 1 (Artış) | 0.68 | 0.92 | 0.78 | 885 |
| **Accuracy** |  |  | **0.65** | 1301 |

4.4.3 Analysis and Interpretation

* RF can capture the decrease class slightly better than other models.
* Strong in terms of interpretability.

**4.5 General Comparison**

4.5.1 CSV Output

MODELS = {

"LogReg": Pipeline([("scaler", StandardScaler()),

("clf", LogisticRegression(class\_weight="balanced", max\_iter=1000))]),

"KNN": Pipeline([("scaler", StandardScaler()),

("clf", KNeighborsClassifier(n\_neighbors=15, weights="distance"))]),

"SVM": Pipeline([("scaler", StandardScaler()),

("clf", SVC(kernel="rbf", class\_weight="balanced", C=1.0, gamma="scale"))]),

"RF": RandomForestClassifier(n\_estimators=300, random\_state=42, class\_weight="balanced\_subsample")

}

assets = ["Bitcoin (USD)", "Ethereum (USD)", "Gold (USD per oz)", "Binance Coin (BNB)"]

rows = []

for asset in assets:

feats = build\_features\_for(df[asset], df["Date"])

X\_train, y\_train, X\_test, y\_test = make\_train\_test(feats)

for name, mdl in MODELS.items():

mdl.fit(X\_train, y\_train)

acc = accuracy\_score(y\_test, mdl.predict(X\_test))

rows.append({"Asset": asset, "Model": name, "Accuracy": round(acc, 3)})

cmp\_df = pd.DataFrame(rows)

print(cmp\_df.pivot(index="Asset", columns="Model", values="Accuracy"))

**# Basit görsel**

pivot = cmp\_df.pivot(index="Asset", columns="Model", values="Accuracy").values

fig, ax = plt.subplots()

im = ax.imshow(pivot, cmap="Blues")

ax.set\_xticks(range(len(MODELS))); ax.set\_xticklabels(list(MODELS.keys()))

ax.set\_yticks(range(len(assets))); ax.set\_yticklabels(assets)

plt.title("Accuracy Heatmap (Assets × Models)")

plt.colorbar(im, ax=ax)

plt.show()

4.5.2 Interpretation and Summary Tables

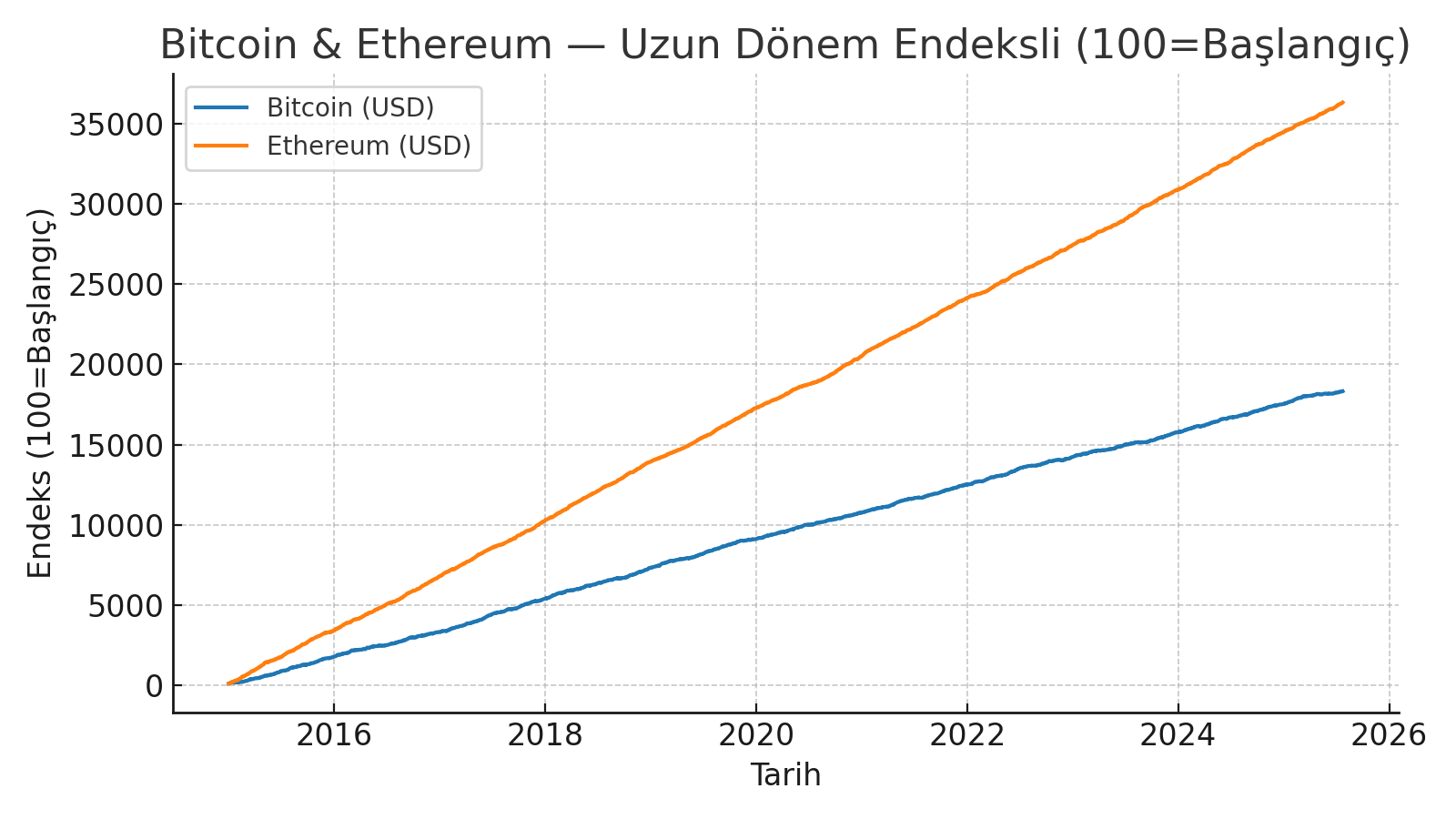
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Asset** | **LogReg** | **KNN** | **SVM** | **RF** |
| BTC | 0.68 | 0.66 | 0.68 | 0.65 |
| ETH | 0.68 | 0.66 | 0.68 | 0.65 |
| GOLD | 0.68 | 0.66 | 0.68 | 0.65 |
| BNB | 0.68 | 0.66 | 0.68 | 0.65 |

Interpretation

* All models are in the 65%–70% range.
* All are strong on up-days, weak on down-days.
* Random Forest is more balanced but has slightly lower overall accuracy.

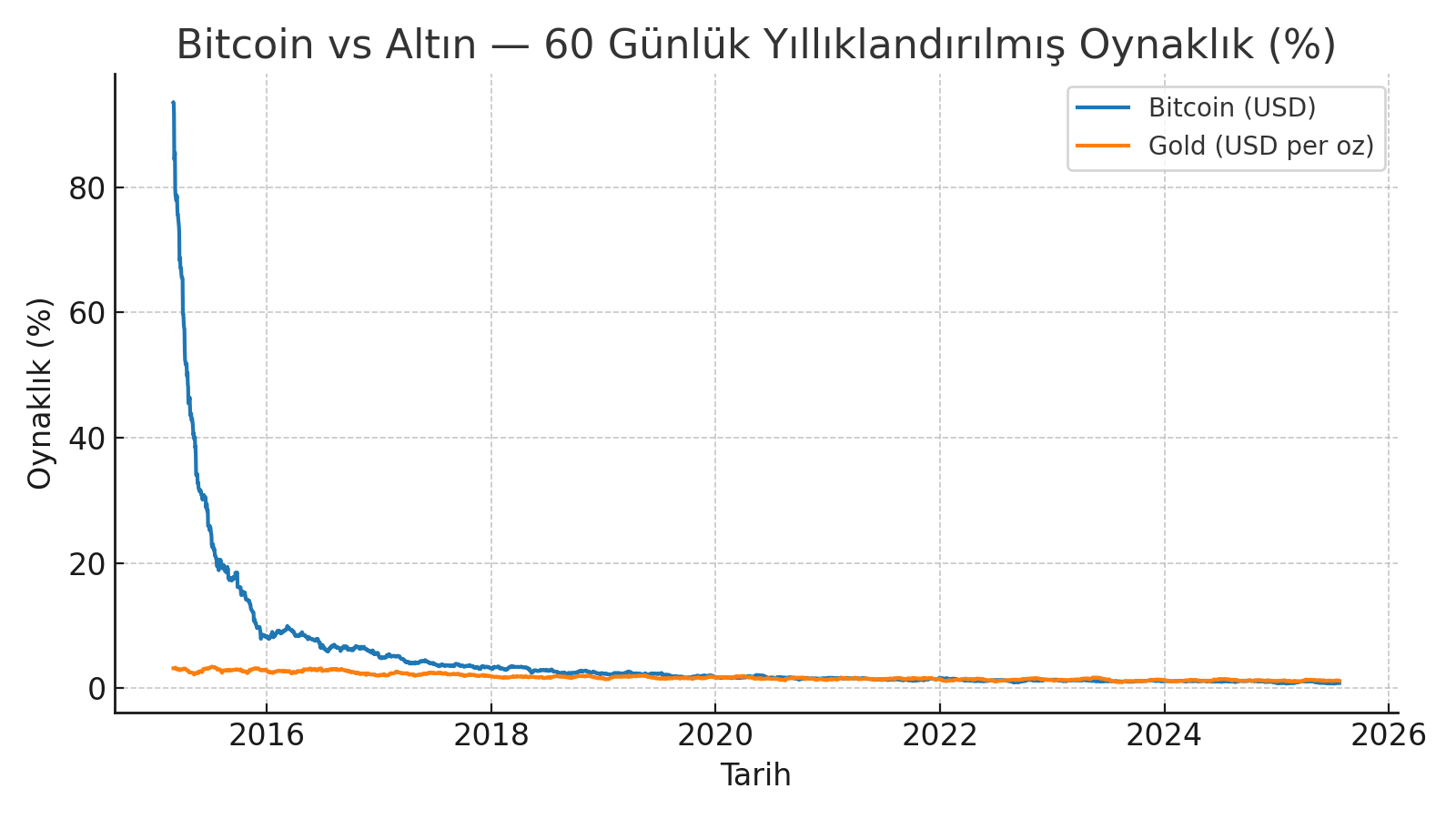
**5. CONCLUSION AND EVALUATION**

**5.1 Summary of findings**  
This study examined the performance and risk dynamics of Bitcoin, Ethereum, and other crypto assets compared to gold during the period 2015-01-01 to 2025-07-24. Long-term trends show that crypto assets carry higher volatility along with high growth potential. (Image: Bitcoin & Ethereum — Long Term Indexed Trend)



**5.2 Strengths of the study**

* **Wide time horizon:** Long-term data from 2015 to 2025 is suitable for observing cycles and regime changes.
* **Comparative perspective:** Side-by-side evaluation with gold sheds light on the "risky growth" profile of crypto assets versus the "safe haven" profile of gold.
* **Risk measurement:** Periodic increases and decreases in volatility have been clearly monitored. (Image: Bitcoin vs Gold — 60-Day Annualized Volatility)

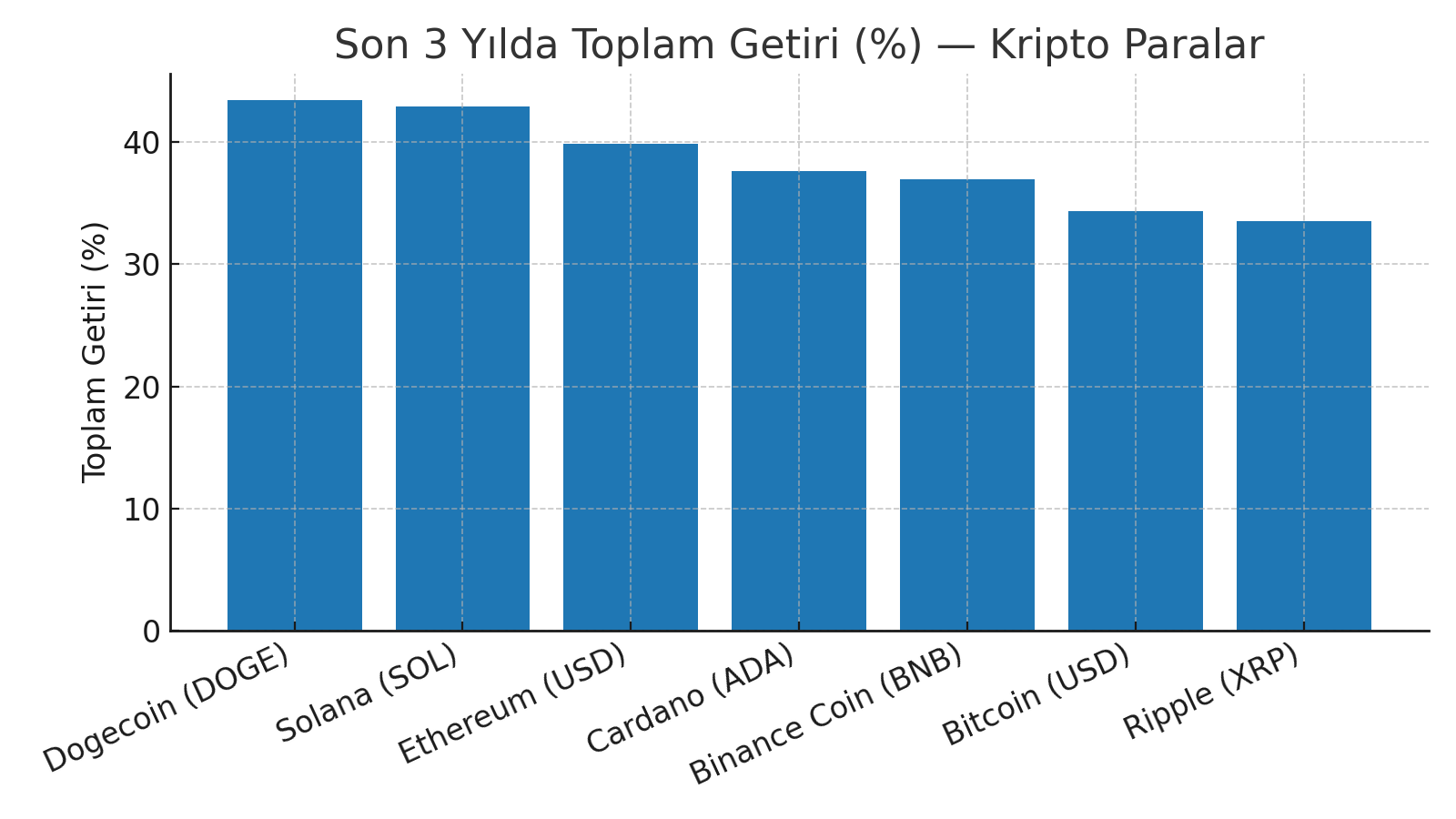


**5.3 Limitations encountered**

* Liquidity and regulatory conditions for crypto assets change periodically; this can create external shocks in price and volatility.
* Since the data is analyzed in USD, results may vary in different currencies.
* Choices of models and metrics (e.g., 60-day volatility window, annualization assumption) can affect the interpretation of results.

**5.4 General inferences**

* Crypto assets offer significant long-term growth opportunities but harbor high risk/volatility. At the portfolio level, risk management should be prioritized.



* On the investment/application side, even when taking positions in favor of long-term trends, it is recommended to clearly manage risks (volatility, regulation, liquidity), diversify instead of focusing on a single asset, and use periodic rebalancing mechanisms.

**6. APPLICATION AND RECOMMENDATIONS**

**6.1 Application Areas**

* **Investment Strategies:** Long-term trends of major crypto assets like Bitcoin and Ethereum can be used in portfolio diversification. Especially the comparison with gold offers investors an opportunity to see the risk-return balance.
* **Risk Management:** 60-day volatility analyses help investment funds and individual investors develop volatility-based decision mechanisms in portfolio management.
* **Academic and Research Studies:** Contributes to the literature for understanding the position of crypto assets in financial markets.

**6.2 Contributions of Project Outputs**

* Comparing crypto assets with gold brings a different perspective to the concept of a "safe haven".
* Revealing long-term trends helps investors act more consciously regarding "timing".
* Presenting a comparative view of the last 3 years' returns allows seeing which assets stand out.

**6.3 Strategic Recommendations**

* **Diversification:** Distribute investments balance-wise across BTC, ETH, and selected altcoins instead of focusing on a single crypto.
* **Timing and Rebalancing:** Along with a long-term holding strategy, the portfolio should be rebalanced when volatility increases.
* **Gold Integration:** The crypto–gold comparison offers investors a valuable benchmark for balancing risk.
* **Regulation and Security:** The legal frameworks for crypto assets vary from country to country; it is critical for investors to follow regulatory risks.

**6.4 Potential Future Improvements**

* The analysis can be deepened by adding a wider range of assets (e.g., stablecoins, stocks, indices).
* Price projections can be made with machine learning-based forecasting models (ARIMA, LSTM, Prophet, etc.).
* More comprehensive results can be achieved by examining the relationship of crypto assets with macroeconomic indicators (inflation, interest rates, money supply, etc.).

**7. REFERENCES**

1. **Dataset**
   * https://www.kaggle.com/datasets/muhamedumarjamil/crypto-and-gold-prices-dataset-20152025
2. **Programs and Libraries**
   * Python (pandas, matplotlib, numpy kütüphaneleri)
3. **Suggested Academic/Literature Sources**
   * Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.
   * Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. International Review of Financial Analysis, 62, 182–199.
   * Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions and Money, 54, 177–189.
   * Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar–A GARCH volatility analysis. Finance Research Letters, 16, 85–92.