Study case 2024

Analyzing Bitcoin Price Trends Using Crypto News (2021-2023)

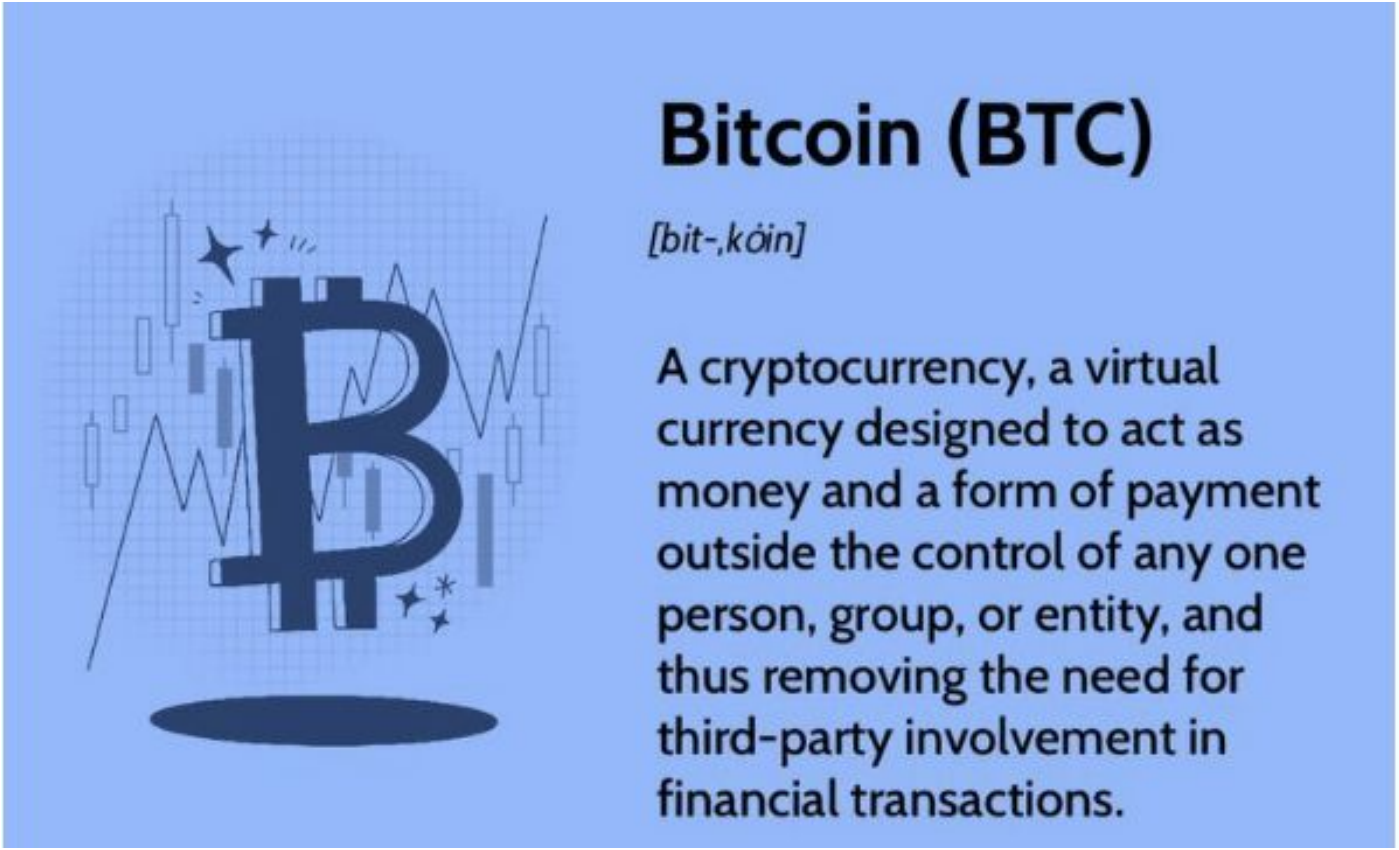
Exploring the correlation between sentiment, significance and market trends.

Francescut R., 12 December 2024 - CodeOp

What is Bitcoin???

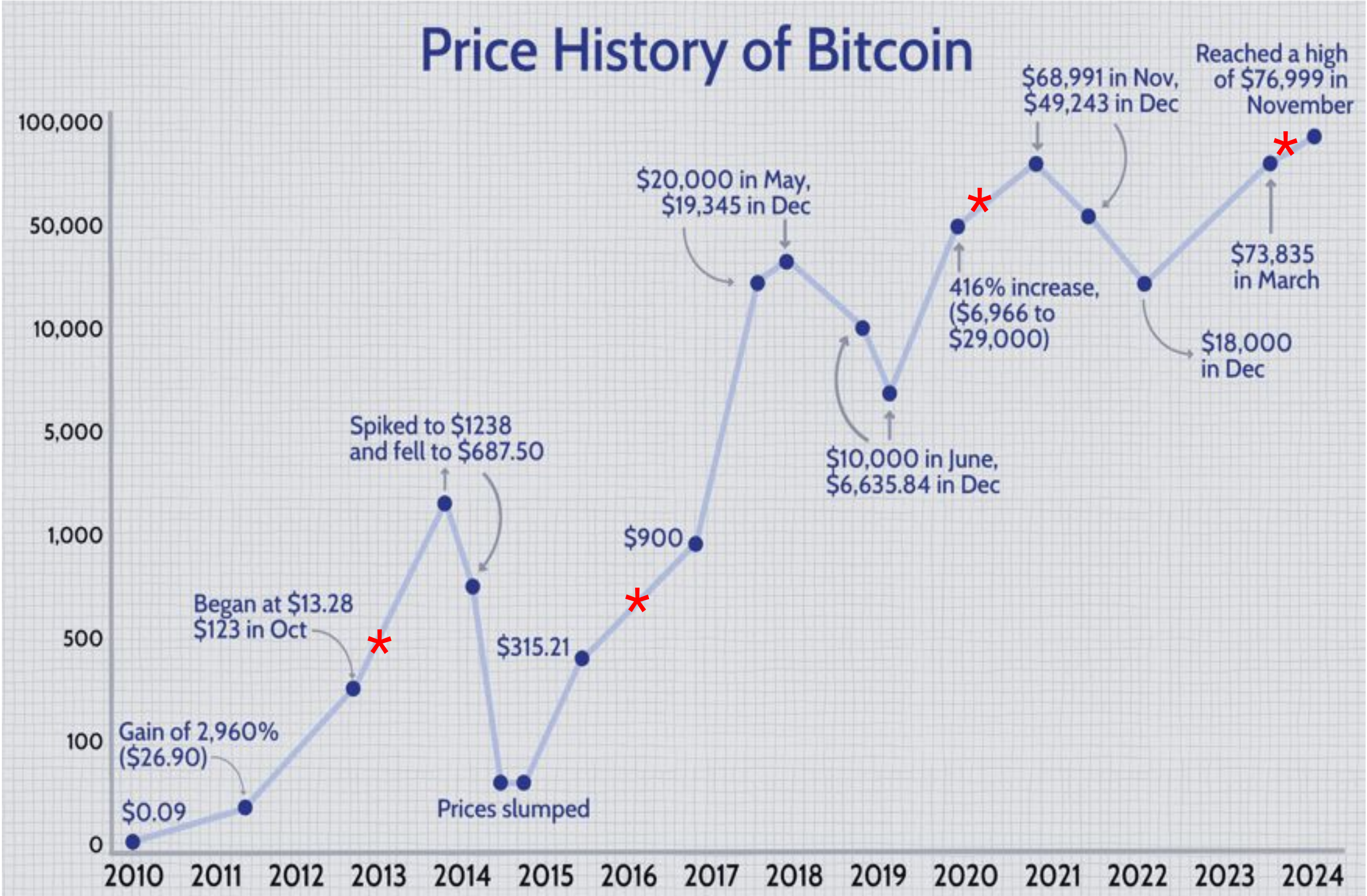
Miners use computational power to validate and record transactions on the Bitcoin blockchain.

As a reward for their work, they receive a certain amount of newly created Bitcoin.



The "halving" is a pre-programmed event that reduces the mining reward by half approximately every four years.

- 3 January 2009
- 28 Novembr 2012
- 9 July 2016
- 11 May 2020
- 19 April 2024



At the moment the rewards stand at 3,125BTC for successful miners.

However, my problem statement:

Bitcoin prices are highly volatile and influenced by multiple factors, including news and market sentiment.

Objective

Investigate how crypto news sentiment and significance influence Bitcoin price trends.

Focus on predicting and understanding significant price changes using a combined dataset of crypto news and Bitcoin price data.

Dataset Scope

Crypto News:
Over 3 years (2021-2023), with sentiment, user-defined importance, and engagement metrics.

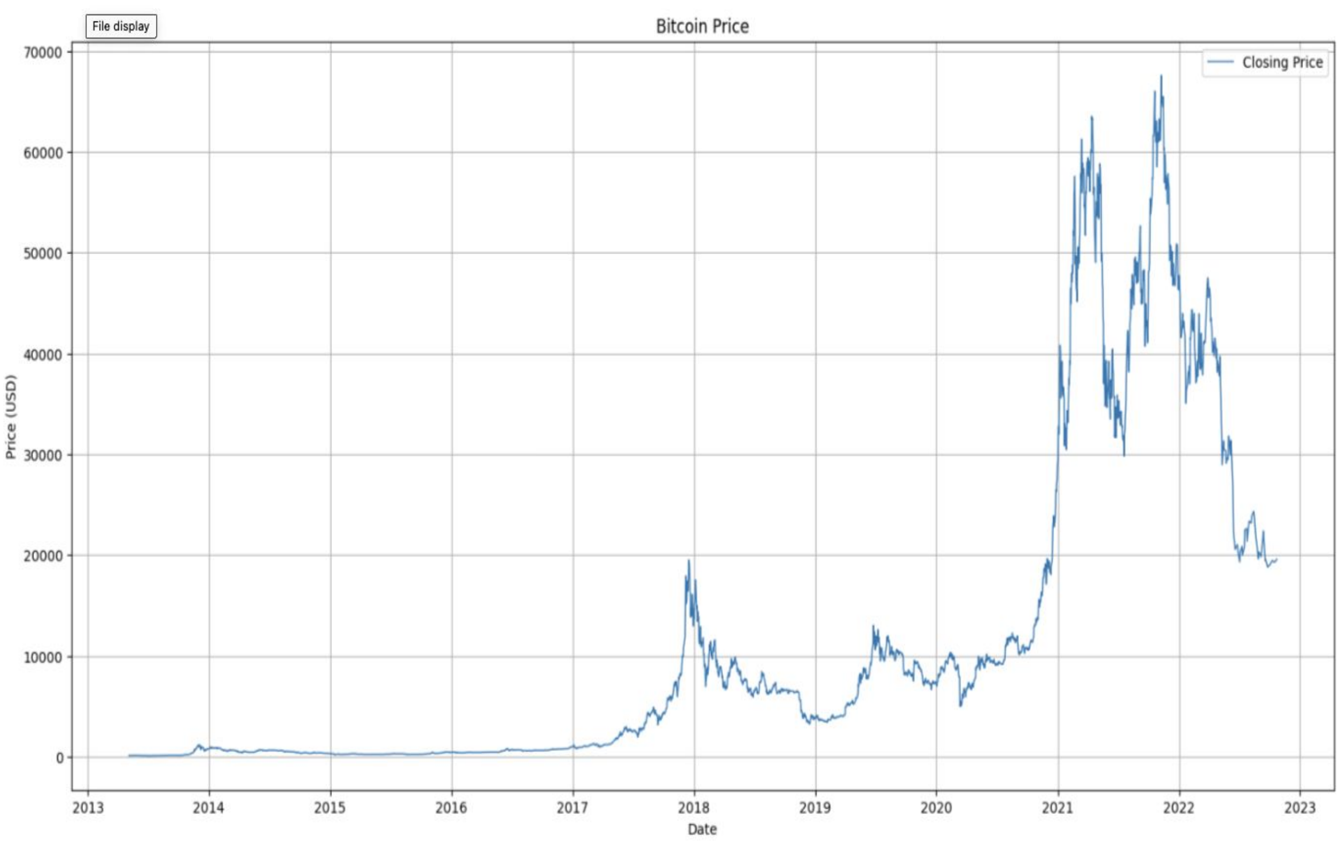
Bitcoin Prices:
3-day moving averages, closing prices, volatility and volume metrics.

Problem Statement

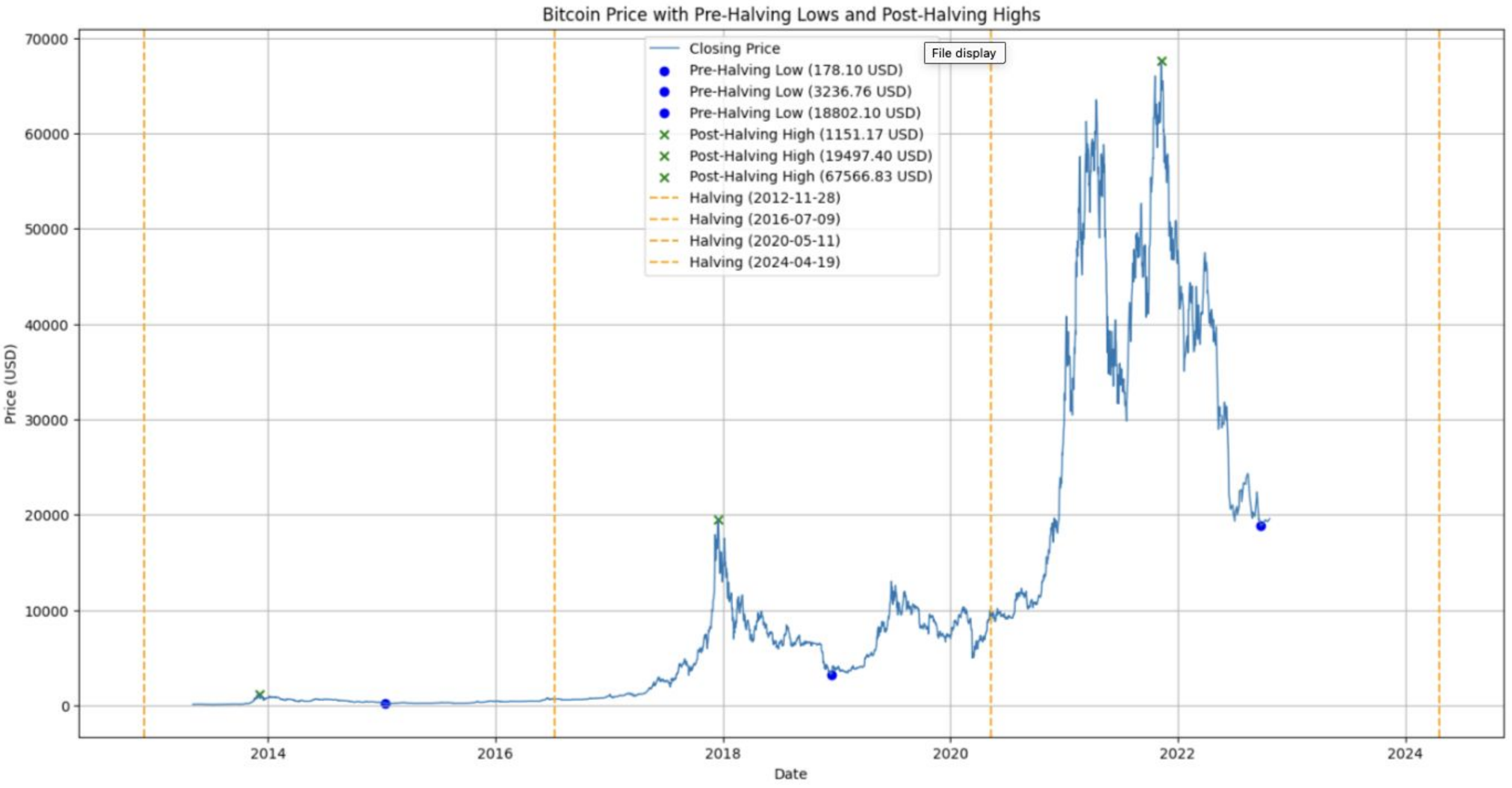
Can we identify patterns in crypto news that predict or correlate with significant Bitcoin price changes?

What are the most frequent keywords in news articles linked to price shifts? changes?

Graph of the trend of the Bitcoin close by Year



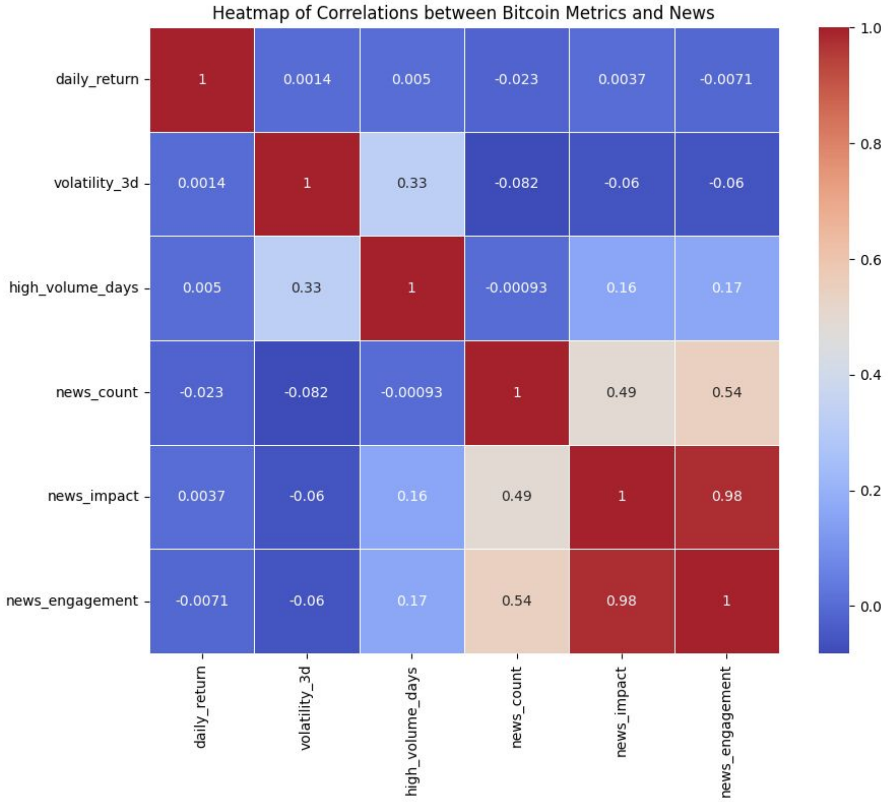
Graph of the trend of the Bitcoin close by Year VS. and the Halvings with their respective minimum and maximum



```python
halving_dates = [
    pd.Timestamp('2012-11-28'),
    pd.Timestamp('2016-07-09'),
    pd.Timestamp('2020-05-11'),
    pd.Timestamp('2024-04-19')  # Estimated next halving
]

pre_halving_lows = {}
for halving_date in halving_dates:
    start_date = halving_date - pd.DateOffset(years=2)
    valid_data = bitcoin_df[start_date:halving_date]
    if not valid_data.empty:
        pre_halving_low = valid_data['close'].idxmin()
        pre_halving_lows[halving_date] = pre_halving_low
post_halving_highs = {}
for halving_date in halving_dates:
    end_date = halving_date + pd.DateOffset(years=2)
    valid_data = bitcoin_df[halving_date:end_date]
    if not valid_data.empty:
        post_halving_high = valid_data['close'].idxmax()
        post_halving_highs[halving_date] = post_halving_high
```

```python
plt.figure(figsize=(18, 9))
plt.plot(bitcoin_df.index, bitcoin_df['close'], label='Closing
Price', linewidth=1)

for halving, low_date in pre_halving_lows.items():
    price = bitcoin_df['close'].loc[low_date]
    plt.scatter(low_date, price, color='blue', marker='o',
label=f'Pre-Halving Low ({price:.2f} USD)')

for halving, high_date in post_halving_highs.items():
    price = bitcoin_df['close'].loc[high_date]
    plt.scatter(high_date, price, color='green', marker='x',
label=f'Post-Halving High ({price:.2f} USD)')

for halving in halving_dates:
    plt.axvline(halving, color='orange', linestyle='--',
label=f'Halving ({halving.date()})')

plt.title('Bitcoin Price with Pre-Halving Lows and Post-Halving
Highs')
plt.xlabel('Date')
plt.ylabel('Price (USD)')
plt.grid(True)
plt.legend()
```

Heatmap of Correlation between Bitcoin Metrics and News Sentiment



Key Insights from the Heatmap:

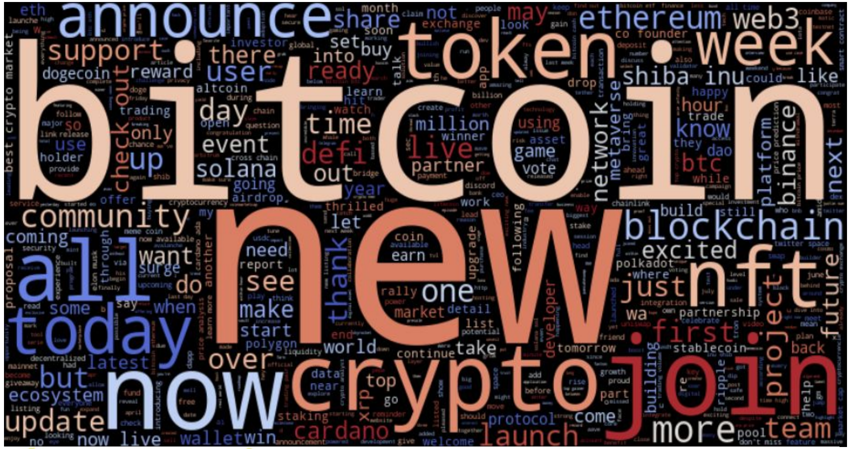
- **Daily Return**: Shows **minimal correlation** with all news metrics, indicating news doesn't significantly impact short-term price changes.
- **Volatility (3-day)**: Positively correlates (**0.33**) with **high volume days**, suggesting volatility increases with trading activity, as expected.
- **High Volume Days**: Slight correlation with **news impact (0.16)** and **news engagement (0.17)**, showing some relationship between news and trading volume.

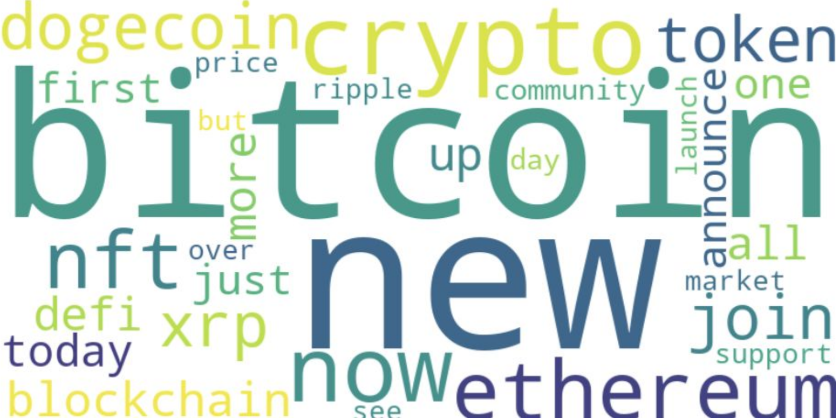
News Metrics:

- **News Count**, **Impact**, and **Engagement** are strongly correlated with each other (**0.49 to 0.98**).
- However, they have **limited influence** on Bitcoin's volatility and returns.

Conclusion: News metrics correlate with each other but have little effect on Bitcoin's daily return and volatility

Word Cloud of News Titles/Significant





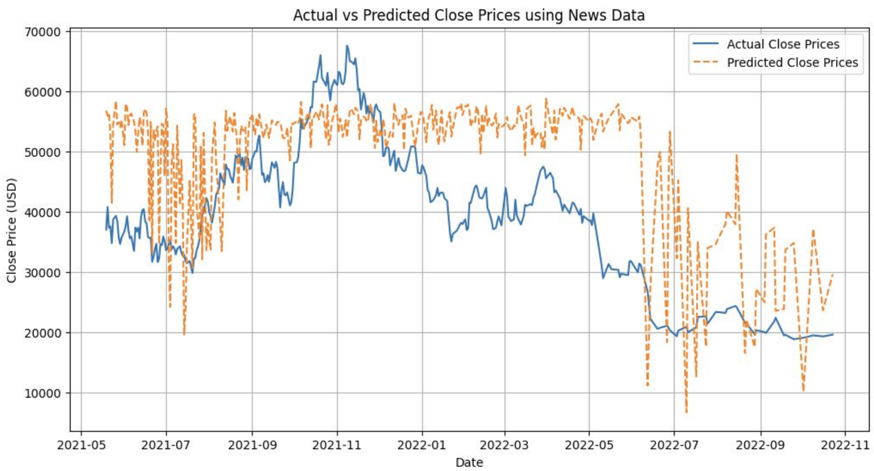
Considerations:
the presence of **time-sensitive terms** like “now” “first” “announce” or “today” in the significant price change dataset appears to be similar to their frequency in the general news dataset.

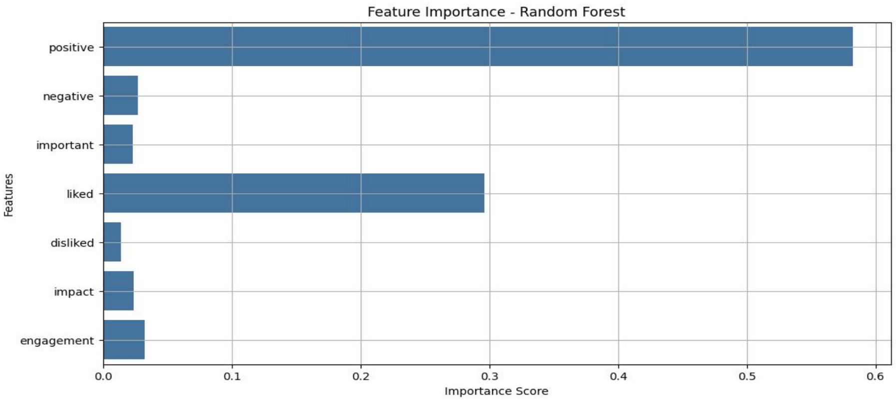
This suggests that these words are **not significantly more present** during price changes and may not reliably indicate market movements on their own.

It would be interesting to develop further analyses based on the **contextual relationships between words** in the news dataset. For example, exploring how certain keywords appear together in specific contexts or headlines could reveal more nuanced patterns.

Additionally, analyzing the **frequency and sentiment shifts** of these words over time might provide deeper insights into their impact on market movements.

Machine Learning and Prediction





Model Performance Metrics:

```
Mean Absolute Error (MAE): 11275.479805196559
Mean Squared Error (MSE): 172990882.0286103
Root Mean Squared Error (RMSE): 13152.599820134812
R^2 Score: -0.38070089107332006
```

The high error values and negative R^2 indicate that the model performs poorly and is less accurate than a simple mean prediction.

Feature Importance

Key Features:

- “positive” (Importance: ~0.58) – The most considered model’s feature.
- “liked” (Importance: ~0.30) – The second most relevant feature

Minor Impact:

- “negative,” “important,” “impact,” and “engagement” have minimal influence.

Poor Model Performance:

The Random Forest model struggles to predict Bitcoin prices accurately based solely on news data.

Limited Predictive Power:

While positive sentiment and engagement are the most important features, they are insufficient to explain price movements.

Recommendation:

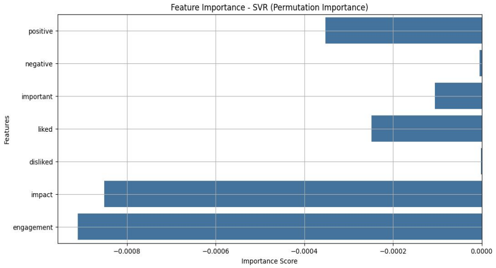
Incorporate additional data sources, such as market indicators and macroeconomic factors, or explore advanced models like LSTMfor time-series forecasting.

Conclusions

In summary, news data alone does not provide enough predictive power for accurate Bitcoin price predictions.

Support Vector Regression (SVR):

Linear Regression Results Summary



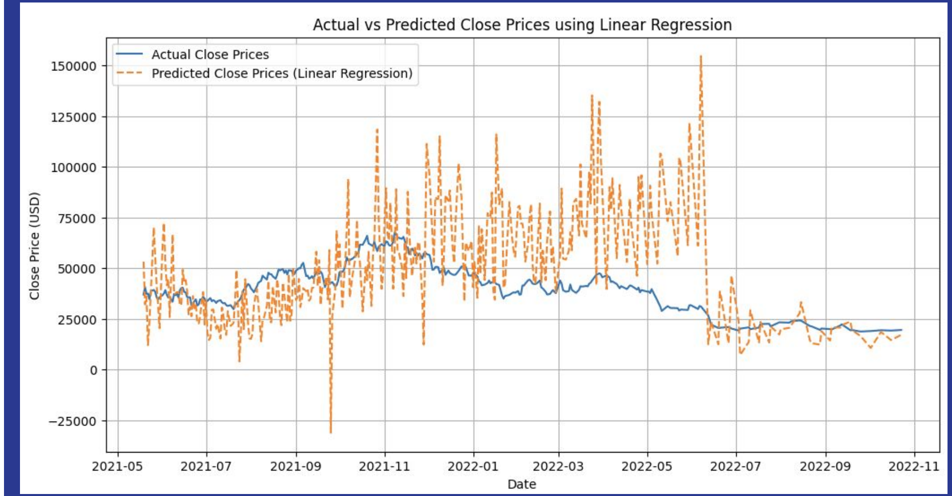
High Errors:
- **MAE**: 32,753.95
- **RMSE**: 34,608.68

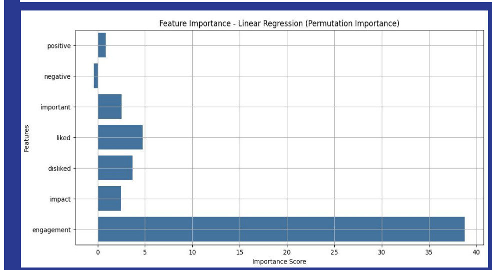
Poor Fit:
• **R^2 Score**: -8.56 (model performs worse than the mean prediction)

Graph:
The Orange dashed line is nearly flat and failing to follow the actual trend.

Conclusion:
The predicted values do not capture any upward or downward movement, suggesting the SVR model is **broken or misconfigured**.

Linear Regression:





Linear Regression Results Summary

High Errors:
- **MAE**: 19,437.05
- **RMSE**: 27,435.74

Poor Fit:
• **R^2 Score**: -5.01 (model performs worse than the mean prediction)

Graph:
Predicted prices show **high volatility** and **large deviations** from actual prices.

Conclusion:
Linear Regression **fails to capture** Bitcoin price trends, indicating it is **unsuitable** for this dataset.

Interesting metrics to investigate:

It appears that there is a recurring pattern in the timing of Bitcoin's price minimums and maximums surrounding each halving event. Specifically, historical data indicates that.

This cyclical behavior suggests that the market typically experiences a trough within this timeframe post-halving, followed by a potential move toward new highs.

However, predicting daily price surges or declines based solely on a **sector-specific dataset of crypto news** proves challenging.

The complexity arises due to the multitude of **variables** influencing economic conditions and market trends.

Factors such as **global economic policies, investor sentiment, regulatory developments,** and **macroeconomic indicators** all interplay, making precise short-term predictions inherently unreliable.