Cryptocurrency Price Prediction Using News and Social Network Data

Berk Buzcu
Özyeğin University
Istanbul, Turkey
berk.buzcu@ozu.edu.tr

Mert Ozgun
Özyeğin University
Istanbul, Turkey
mert.ozgun@ozu.edu.tr

Doğa Yılmaz
Özyeğin University
Istanbul, Turkey
doga.yilmaz.11481@ozu.edu.tr

***Abstract*—Cryptocurrencies have emerged as digital assets in recent years. These advancements bring the need for trading and algorithmic operations on them. It is well established that these assets are highly emotional and require more than technical analysis to quantify a buy or sell signal. This study aims to generate actionable alternative data with the help of the average sentiment of people on Twitter and other social media platforms. In this context, we trained two Deep-Learning based sentiment analysis models and created a hybrid strategy that mixes alternative data and technical analysis.**

***Index Terms*—Cryptocurrencies, Fin-BERT, Sentiment Model, MACD**

I. Introduction

Crypto-currencies are digital assets utilized in trading for commodities, or profit through real currencies. The most recent volatility coming from the ability to quickly buy and sell this commodity has made room for investors to exploit and turn quick profits in a short period of time.

Bitcoin is the chief example of Cryptocurrencies. Each user has their own private online wallets which they use in buying and selling this commodity from online Bitcoin traders. These operations are made to be decentralized, meaning they are free from influence of third parties, granted by the heavy usage of cryptographic methods throughout the system.

Meanwhile the trends of day to day currencies are easier to predict and work on, cryptocurrencies prove to be influenced by other factors as well. This could be attributed to the lack of a regulatory power, since the crypto currencies are relatively self regulated by the economy participants, it is heavily affected by social trends as well as other economic factors. [4] In order to find such a relation, we explore the News and Social Network Data, mainly utilizing sentiment analysis methods.

Analyzing sentiment on the other hand is a cheap and effortless way to gain additional trading strategies. Sentiment analysis tools are easy to use and extremely powerful when it comes to picking winning trades.

II. Related Work

Cryptocurrency market analysis is a well known research topic, and learning methods weren't lacking in such. One instance would be the [1] attempt to utilize Decision tree and regression techniques on the values of cryptocurrencies. They have noticed due to the pattern of the rising prices in the cryptocurrencies, a complicated regression pattern fits the case better than a decision tree. But this research has only utilized the prices without accounting for the mentioned factors that was mentioned.

Another research done in this field is the [2] which also utilized Twitter data to make an analysis on tweet sentiments and utilized this analysis to find a link between the cryptocurrency values and the twitter trends. One key finding in this research is that they noticed a strong relation between twitter and finance markets and so on twitter's effects on the cryptocurrency markets. In this research, they utilized a simple Ensemble algorithm for classification, the random forest classifier for the sentiment analysis aspect of it and then correlated the sentiments of these tweets wholly to the price fluctuations of cryptocurrencies and found a 68 correlation. This is inline with our main intuition.

A. BERT

Bidirectional Encoder Representations from Transformers (BERT) is a state-of-the-art language representation model by Devlin et al. [5] The novelty of BERT is the ability of creating state-of-the-art models for a wide range of tasks without substantial task specific architecture. The resulting model can compete with state-of-the-art methods of various problems such as question answering or sentiment analysis. Within the scope of this project we used a fine tuned version of BERT for sentiment analysis. [5]

B. Fin-BERT

FinBERT is a state-of-the-art financial sentiment analysis model, which is a fine tuned version for the BERT to better suit the Financial Domain and sentinmenta analysis. The aim of financial sentiment analysis task is different from general sentiment analysis. The purpose behind it is guessing how the markets will react with the information presented in the data. [6]

C. LSTM

Another deep learning approach utilized in sentiment analysis is Long short-term memory (LSTM) which is a Recurrent Neural Network with the capacity to learn long term dependencies. Given LSTM's robustness when it comes to Text Classification per its capacity to understand the context of texts by inspecting the long term dependencies rather than the classic feed forward strategy in RNNs, we have decided to utilize it as a benchmark to make a baseline for our method. [7]

D. Back-testing

Back-testing allows traders and data researchers to test out the day-trading strategies they implemented thoroughly. The power of back-testing comes from testing a hypothesis on a previous date and calculating Return on Investment (ROI) and other metrics accordingly. Back-testing works as an inductive reasoning mechanism. Its theory suggests that a given strategy and pattern that worked on the past data will work on the future, vice versa.

In order to perform back-testing on a given model, there needs to be a set time frame that the strategy has clear boundaries. The hypothesis should be quantitative rather than being qualitative. Testing the strategy with several time frames will ensure that the hypothesis is robust enough to be utilized in future data.

III. Models

In our project, we utilized a technical analysis based model and combined it with our sentiment analysis module to maximize returns.

A. Naive Sentiment Model

This model is based on the sentiment analysis from the previous data analysis and comment classification task. Sentiment classification data is binned to 1-hour intervals; the decision behind binning was made as 1 hour since it gives a reliable number of data points and stays relevant. Binned data is used to calculate average positive and negative comment counts. The decision of buy and sell is made by checking if the sentiment is inclined on some side or another. [9]

B. Moving Average Convergence Divergence (MACD) Model

MACD is a trend-following momentum strategy that utilizes two different moving averages to calculate whether an asset is oversold or overbought. The calculation results in a line that indicates a signal boundary. This line is used to trigger a buy or sell operation. This indicator helps analyze a security's price movements, whether it strengthens or weakens in a given time frame or not. The signal and MACD itself is used to crossover to combine both long term and short term consequences. Crossover strategies are best used in high volatility assets like cryptocurrencies to ensure that the model is error prune. [8]

$$MACD = EMA_{12} - EMA_{26}$$

C. Sentiment Aware Moving Average Convergence Divergence Model

This model is a hybrid that utilizes both the MACD indicator and our alternative data. It emulates a MACD indicator while the trades are regulated using the sentiment model. Line for MACD is calculated as normal but the decision to enter the market or sell is given by the crossover indicator and the sentiment for that given time period [12].

IV. Backtesting

Our project utilizes two similar technical analysis based models on both alternative data and market data. The first model is the Simple Moving Average (SMA) method. In this method, the basic assumption is that the market will respond to particular events with a certain lag in time. For our case, sentiment data perform best with 24 hours of shifted average. In SMA, since our main focus was on finding the best lag for the market response, we did not utilize a crossover strategy. The second model is called Moving Average Convergence Divergence (MACD). The details of this model are explained in Section III.

Our backtesting is done through a Python framework called Backtrader. [3] Backtrader allows traders and researchers to write seamless trading strategies without the need to deal with market dynamics. It abstracts out

Algorithm 1: Sentiment Aware MACD Indicator

Result: Market Action Signal
Compute MACD and MACD signal
if *next hour* **then**
  if *pending order* **then**
    return
  else
    if *not in market* **then**
      if *sentiment positive and MCross > 0* **then**
        send buy signal
      end
    else
      if *sentiment negative and MCross < 0* **then**
        send sell signal
      end
    end
  end
end

TABLE I
MODEL RESULTS ON FINANCIAL PHRASE BANK

Model	**Loss**	**Accuracy**	**F1 Score**
LSTM	0.0377	0.70	0.70
FinBERT	0.37	0.86	0.8

most of the infrastructure needed for testing a strategy. The only modification to Backtrader that we required was enabling Fractional Trading since the price of cryptocurrencies are high that we may want to trade fractions of a Bitcoin. Also, we choose Bitmex as the exchange since we gathered our data from their resources. The back trading algorithm is run with \$1000 initial balance. Also, we added Bitmex commission which is \$0.00075, to our trading strategy.

Backtesting for our strategies is done in 3 months, between February 2021 and May 2021. We conducted separate experiments for each month; the reason for this was that due to the market being in a bullish stage, our experiment did not give reliable results by running in a prolonged timeframe.

V. BENCHMARK

Due to LSTM's powerful capacity in Text Classification, we utilized it in Sentiment Analysis, which is a sub-problem of Text Classification. In the chart below, we detail our LSTM layers.

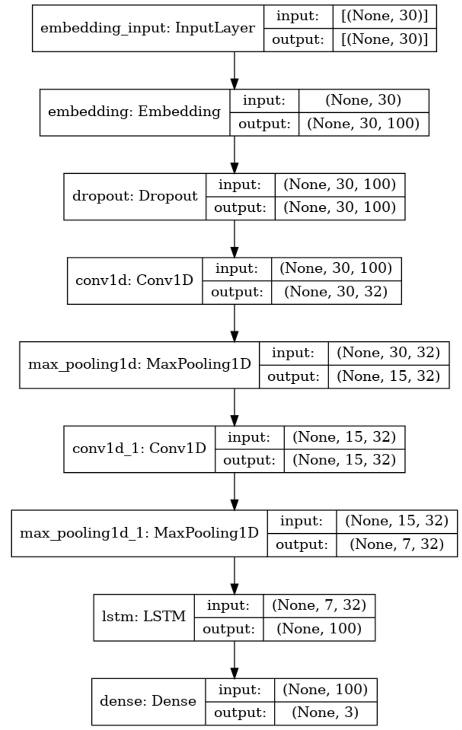

Fig. 1. LSTM Architecture

The table above displays the difference in various metrics between our models. LSTM meanwhile performing considerably well, ends up in a worse position against FinBERT on text classification tasks. One of the key factors in this difference is that Fin-BERT is fine tuned to better work in finance domain whereas our LSTM is more generally applied.

Since sentiment analysis is our main task, and FinBERT has performed better in this task, we have decided to utilize FinBERT as our main prediction module for the rest of the tasks.

VI. RESULTS

As described in the previous sections, we have tested Naive Sentiment, MACD and Sentiment Aware MACD models using Bitcoin Tweets data-set. Data-set contained bitcoin related tweets from February 2021 to April 2021. The return of investment for each month and for the whole data-set can be found in the Appendix section.

As it can be seen from the data shown in the table 2, by just simply buying and holding BTC we can achieve 28.25% ROI. However we can further improve our ROI with our proposed model to 34.02%. ROI for other methods are also available in the Figure 2.

	February 2021	March 2021	April 2021	February-April 2021
Sentiment Aware MACD	30.85% ROI	2.86% ROI	4.81% ROI	34.02% ROI
MACD	18.98% ROI	5.92% ROI	4.11% ROI	30.97% ROI
Naive Sentiment	18.07% ROI	-0.46% ROI	8.55% ROI	27.22% ROI
Buy And Hold	15.27% ROI	19.43% ROI	-15.18% ROI	28.25% ROI

Fig. 2. ROI for all Methods for each month in data-set

Because of the volatile nature of cryptocurrencies, the main challange here is to beat the market, [10] which also means beating the buy and hold strategy. We keep this as a benchmark and try to improve upon this trading strategy.

Three of our experiments are conducted in a monthly basis and one in a three month period. As it can be seen from Figure 2, our proposed model outperforms 3 of 4 experiments conducted. Sentiment Aware MACD is at it's peak by more than 30% ROI in February 2021. In March, we can see that the buy and hold strategy outperforming all of the models, which might be caused by several factors like big whales entering/exiting the market and other factors. Overall, we can see that our proposed mixed model succeeds at beating the market by 6%, which is a remarkable improvement at scale.

VII. Conclusion

Beating the market in a volatile asset such as Bitcoin is a complex challenge to tackle. Since the distribution and the scale of such assets are much higher than traditional stocks, there are many ways to predict the market direction by investigating the alternative data [11]. In this project, we implemented a Deep Learning based sentiment analysis tool to detect and monitor the market's sentiment. Even though our initial model does not work well, when we mix our method with a well-established technical analysis indicator like MACD, we see significant improvements in our results.

The volatile nature of Bitcoin creates pressure on MACD to perform many buys and sell operations because of the constantly changing market directions. Our sentiment model acts as a regulatory layer for MACD only to buy or sell operations if these two entities agree on a bull or bear market. This way, the strategy buys and sells less than 90% frequency than a regular MACD does. This shows that combining qualitative and quantitative indicators might be a master strategy that can beat up the market.

References

[1] K. Rathan, S. V. Sai and T. S. Manikanta, "Crypto-Currency price prediction using Decision Tree and Regression techniques," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 190-194, doi: 10.1109/ICOEI.2019.8862585.

[2] A. Jain, S. Tripathi, H. D. Dwivedi and P. Saxena, "Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis," 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018, pp. 1-7, doi: 10.1109/IC3.2018.8530659.

[3] Backtrader.com. n.d. Welcome - Backtrader. [online] Available at: ¡https://www.backtrader.com¿ [Accessed 6 June 2021].

[4] Bitcoin: A Peer-to-Peer Electronic Cash System

[5] Jacob Devlin and Ming-Wei Chang and Kenton Lee and Kristina Toutanova BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

[6] Dogu Araci, FinBERT: Financial Sentiment Analysis with Pretrained Language Models

[7] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[8] Hung, Nguyen. (2016). Various moving average convergence divergence trading strategies: A comparison. Investment Management and Financial Innovations. 13. 363-369. 10.21511/imfi.13(2-2).2016.11.

[9] Cartea, Á., Jaimungal, S., & Sánchez-Betancourt, L. (2021). Deep Reinforcement Learning for Algorithmic Trading. SSRN Electronic Journal. doi: 10.2139/ssrn.3812473

[10] Jiang, X. (2020). Bitcoin Price Prediction Based on Deep Learning Methods. Journal Of Mathematical Finance, 10(01), 132-139. doi: 10.4236/jmf.2020.101009

[11] Cartea, Á., Jaimungal, S., & Sánchez-Betancourt, L. (2021). Deep Reinforcement Learning for Algorithmic Trading. SSRN Electronic Journal. doi: 10.2139/ssrn.3812473

[12] Beaudan, P., & He, S. (2019). Applying Machine Learning to Trading Strategies: Using Logistic Regression to Build Momentum-Based Trading Strategies. SSRN Electronic Journal. doi: 10.2139/ssrn.3325656

[13] Backtrader for Backtesting (Python) - A Complete Guide - AlgoTrading101 Blog. (2021). Retrieved 6 June 2021, from https://algotrading101.com/learn/backtrader-for-backtesting/

APPENDIX A
SENTIMENT AWARE MACD MODEL FIGURES

Fig. 3. Compound Model February 2021 Buy-Sell Graph

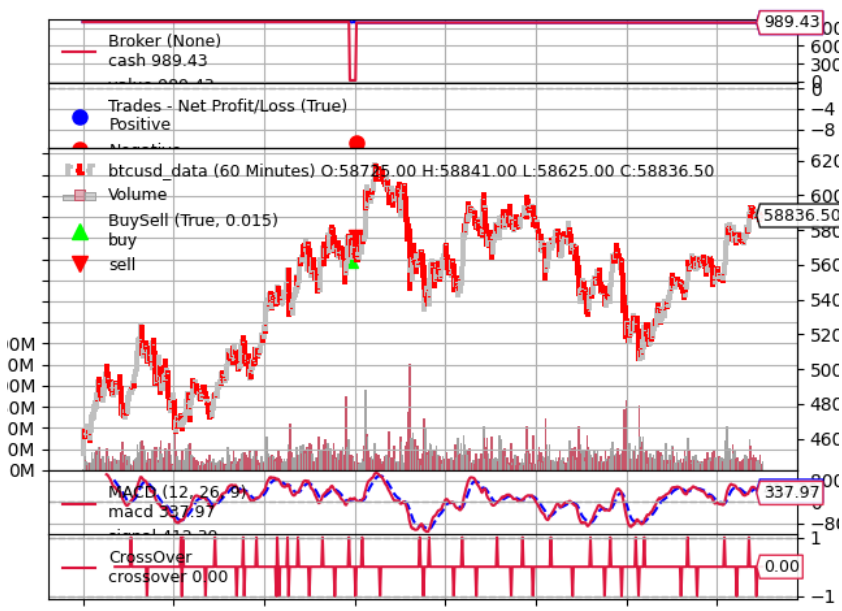

Fig. 4. Compound Model March 2021 Buy-Sell Graph

Fig. 5. Compound Model April 2021 Buy-Sell Graph

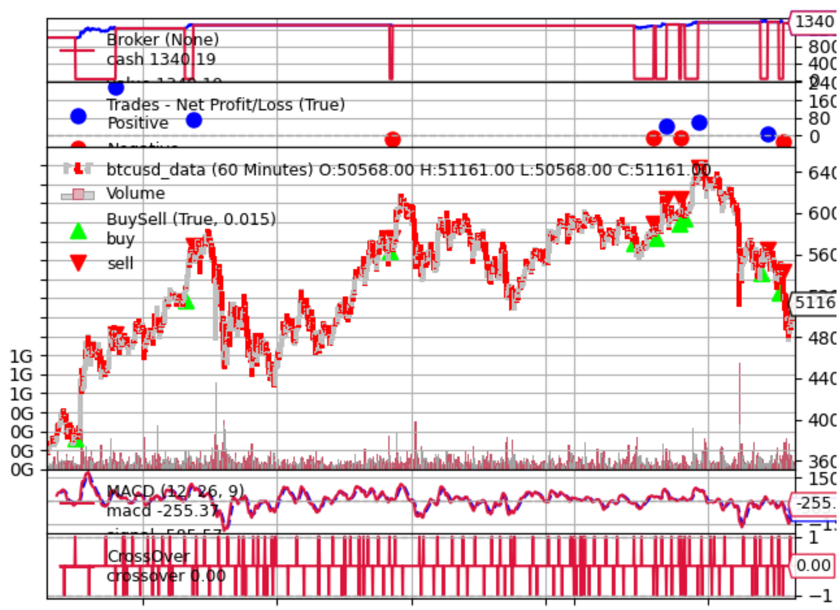

Fig. 6. Compound Model February-April 2021 Buy-Sell Graph

APPENDIX B
MACD MODEL FIGURES

Fig. 7. MACD Model February 2021 Buy-Sell Graph

Fig. 8. MACD Model March 2021 Buy-Sell Graph

Fig. 9. MACD Model April 2021 Buy-Sell Graph

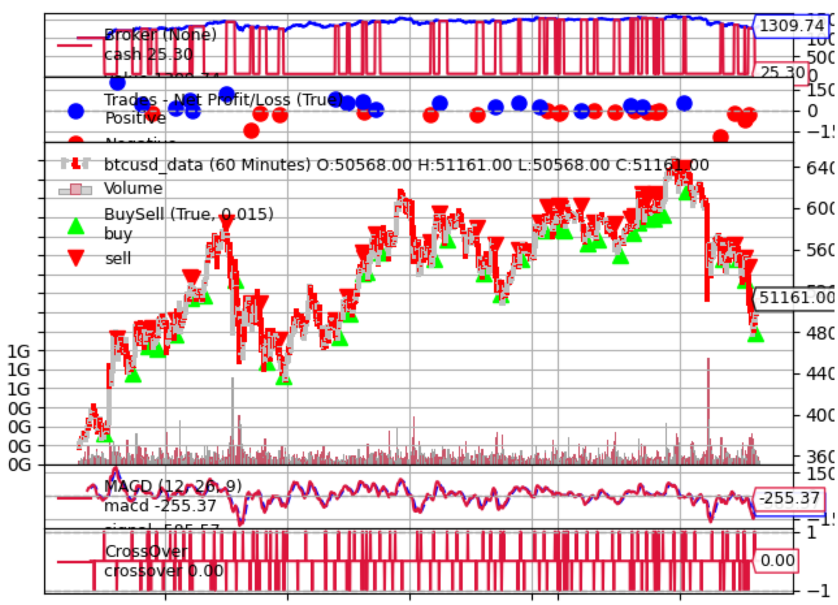

Fig. 10. MACD Model February-April 2021 Buy-Sell Graph

APPENDIX C
NAIVE SENTIMENT MODEL FIGURES

Fig. 11. Naive Sentiment Model February 2021 Buy-Sell Graph

Fig. 12. Naive Sentiment Model March 2021 Buy-Sell Graph

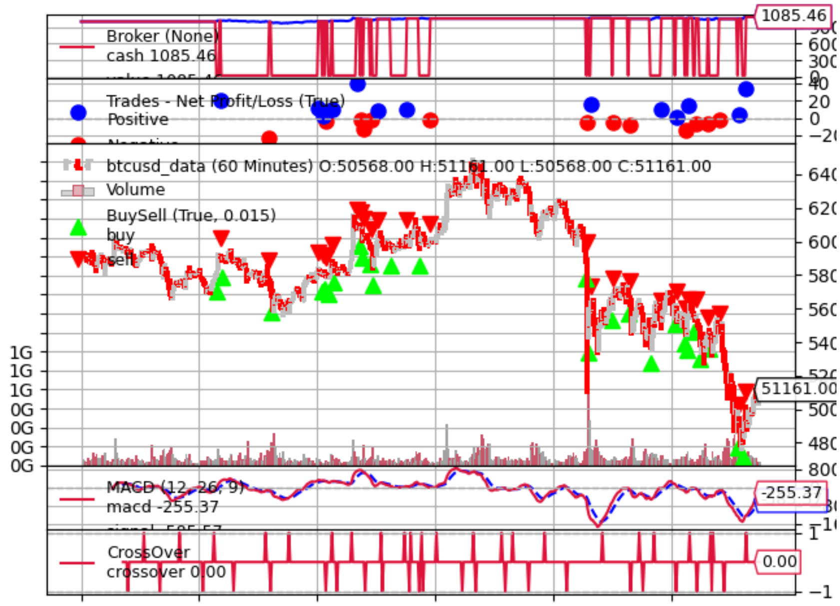

Fig. 13. Naive Sentiment Model April 2021 Buy-Sell Graph

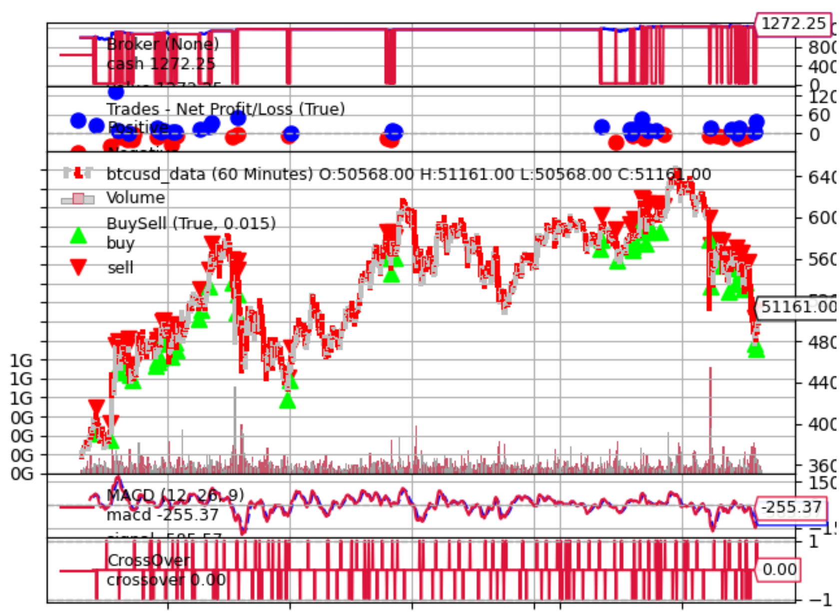

Fig. 14. Naive Sentiment Model February-April 2021 Buy-Sell Graph