

Author	Year	Title	Aims	Method	Sample	Conclusion	Comments
Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova	2019	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	We'll be using this model for our research paper	Train	5000	BERT introduces new architecture deep <i>bidirectional</i> to deal with NLP tasks.	I think BERT has an interesting potential with their own deep <i>bidirectional</i> architecture to tackle standard language models which are unidirectional that limits the architecture.
TonTon Hsien-De Huang, Po-Wei Hong, Ying-Tse Lee, Yi-Lun Wang, Chi-Leong Lok, and Hung-Yu Kao	2018	Hunting the underground inside story of the ethereum Social-network Opinion and Comment	Looking for a comparison of whether using LSTM is faster than using BERT	Train	More than 144 Mio.	LSTM has a higher complexity and takes a long time, so it is difficult to train with high performance.	BERT is faster because BERT only performs one-time processing of the entire sentence or document, while LSTM must process each word in the order given.
Anping Zhao, Yu Yu	2021	Knowledge-enabled BERT for aspect-based sentiment analysis	This paper is used to find out the knowledge-enabled BERT method for NLP Sentiment.	Train		The model leverages additional information from a sentiment knowledge graph to guide the input sentence embedding with a BERT language representation model, and it can capture more explainable information to boost performance and achieve detailed sentiment analysis	The paper proposes a knowledge-enabled BERT model that utilizes a sentiment knowledge graph to improve aspect-based sentiment analysis performance.

						results.	
David C	2023	CRYPTOCURRENCY PRICE PREDICTION USING TWITTER SENTIMENT ANALYSIS	We'll implement this model for our research purposes	Train	1.902.614 Tweets	MAE = 0.007 and RMSE = 0.009.	Based on that result, its possible to predict using sentiment analysis where Forum data by itself is input.
Dan Lin, Jiajing Wu, Qi Xuan, Chi K. Tse	2022	Ethereum Transaction Tracking: Inferring Evolution of Transaction Networks Via Link Prediction	Track and follow transaction records on blockchain from network perspective	Train		The paper uses a machine learning approach to predict the likelihood of a transaction between two addresses based on the historical transaction data. The results show that the proposed method can effectively predict the evolution of transaction networks and identify the most influential nodes in the network.	The paper provides an interesting approach to tracking the evolution of transaction networks and has potential applications in various fields. However, it is important to note that the study focuses on Ethereum and may not be applicable to other blockchain networks. Overall, the paper contributes to the growing body of research on blockchain technology and its potential applications.
Hengyu Li	2022	The Applications of Cryptocurrency: Evidence from Ethereum	Investigate and discuss the basic principles and corresponding applications of Ethereum	Elliptic Curve Digital Signature Algorithm (or be known as ECDSA) in order to ensure the security of transactions	The market value of crypto-currency	Etherium still has many issues to fix ranging from security risks and privacy risks	More research is required to distinguish whether Ether can be a viable alternative that could be added to the economy or even
Abayomi Bello, Siew-Chin	2023	A BERT Framework to Sentiment Analysis of Tweets	To propose a BERT framework	BERT Training	BERT, BiLSTM, Twitter	The proposed BERT framework	The proposed BERT framework is effective in

Ng, Man-Fung Leung			for sentiment analysis of tweets		dataset	achieved state-of-the-art performance in identifying sentiment in tweets according to three categories (positive, negative, and neutral) based on the context of the writer.	sentiment analysis of tweets and can be used in various applications.
Adam Prasetya, Ferdiansyah, Yesi Novaria Kunang, Edi Surya Negara, Winoto Chandra	2021	Sentiment Analisis Terhadap Cryptocurrency Berdasarkan Comment dan Reply pada Platform Twitter	To measure the level of impact given of public opinion on crypto currency	Train with SVM and Naive Bayes method		Naive Bayes = 71.03% on Bitcoin, 75% on Eth, and 71.66% on ripple	This research creates a sentiment analysis on public opinion to see if there's any impact given to crypto currencies with SVM Algorithm. This Method performs better for accuracy in classifying than Naive Bayes.
Stefan Lessmann, Benjamin Fabian	2023	Cryptocurrency Return Prediction Using Investor Sentiment Extracted by BERT-Based Classifiers From News Articles, Reddit Posts and Tweets	To predict cryptocurrency returns using investor sentiment extracted by BERT-based classifiers from news articles, Reddit posts, and tweets.	Fine-tuning a BERT-based sentiment classifier to the specific task of extracting the beliefs, emotions, and motivations of cryptocurrency traders based on combined data from news articles, tweets, and Reddit posts.	Cryptocurrency traders	The sentiment from news articles, Reddit posts, and tweets can be used to predict cryptocurrency returns.	This paper provides a comprehensive approach to predicting cryptocurrency returns using investor sentiment extracted by BERT-based classifiers.
Duygu Ider	2022	Cryptocurrency Return	To predict	Extracting	Cryptocur	Investor	This paper

		Prediction Using Investor Sentiment Extracted by BERT-Based Classifiers from News Articles, Reddit Posts and Tweets	cryptocurrency returns using investor sentiment extracted by BERT-based classifiers from news articles, Reddit posts, and tweets.	investor sentiment using BERT-based classifiers from news articles, Reddit posts, and tweets.	currency traders	sentiment can be used to predict cryptocurrency returns.	provides a straightforward approach to predicting cryptocurrency returns using investor sentiment extracted by BERT-based classifiers.
--	--	---	---	---	------------------	--	--

Sentiment Analysis on Ethereum Cryptocurrencies

Farrel Alexander Tjan

Computer Science

Department

School of Computer Science

Bina Nusantara University

Jakarta, Indonesia 11480

farrel.tjan@binus.ac.id

Hizkia Albertian Lukanta

Computer Science

Department

School of Computer Science

Bina Nusantara University

Jakarta, Indonesia 11480

hizkia.lukanta@binus.ac.id

Reihan Wilbert

Computer Science

Department

School of Computer Science

Bina Nusantara University

Jakarta, Indonesia 11480

reihan.wilbert@binus.ac.id

David David

Computer Science Department

School of Computer Science

Bina Nusantara University

Jakarta, Indonesia 11480

david01@binus.edu

Panji Arisaputra

Computer Science Department

School of Computer Science

Bina Nusantara University

Jakarta, Indonesia 11480

panji.arisaputra@binus.ac.id

Abstract

Sentiment analysis to predict the development of ethereum cryptocurrencies provides valuable insights into the importance of honest market analysis. This study aims to develop an analytical model that can study all market analysis provided by experts, so that it can produce analytical outputs that are easier for many people to understand. Through more than 6,000 expert analyzes from the tradingview.com forum taken using the web scraping method with beautifulsoup4, the machine is trained using a supervised learning method so that it can find out which sentences contain the right words, whether they are short or long. The results of machine learning get positive results in the form of 75-76% precision. Based on the above results, it can be concluded that the model we have created is suitable for use as a tool to assist traders in analyzing market data.

INTRODUCTION

Cryptocurrencies market prediction is the most challenging problem that all traders face. A lot of variables, indicators and analysis have been done to get the most profit out of the market. Technical Analysis and Fundamental Analysis has been used and has been proven for a long time to predict the movement of a market, a lot of indicators in technical analysis such as RSI, MACD, and Ichimoku are also very popularly used with traders as their consideration to place their entry position. Even though there are a lot of factors that influence the price of cryptocurrencies market such as macro economics, which include large scale economic factors such as interest rates, inflation, and *Gross Domestic Product* that we can't avoid.

A lot of traders like to share their own analysis and their opinion through social

media. Because there are a lot of varieties when it comes to analysis, biases and human errors in their analysis that conflict with the market condition it may mislead other people to make a false decision in the market which is crucial and fatal.

In this paper we'll be using Ethereum as our object of analysis, the second biggest market cap in cryptocurrencies. Ethereum gains a lot of attention because a lot of blockchain development was built on this technology. Ethereum is more volatile than bitcoin meaning the movement of the price is fast, unlike other coins beside Bitcoin, Ethereum is highly valued by the market.

The objective of this paper is to help traders and investors to read and determine the sentiment behind the people's analysis, market and news, so they will make a better decision whether they will be taking a long or short position in the market. To achieve this, we will implement BERT (Bidirectional Encoder Representations from Transformers).

The contributions of our papers are as follows:

- We introduce advanced techniques using an advanced language model, to help traders read the current market condition, thus helping them at making decisions in the market
- Throughout this paper, we may contribute for further studies in cryptocurrency, stock market, and global economics.

LITERATURE REVIEW

[1] This paper introduces BERT, a pre-training method for deep bidirectional transformers for language understanding. BERT is trained on large amounts of text data and can be fine-tuned for various NLP tasks, achieving state-of-the-art results on several benchmarks. [2] This paper explores the underground community of Ethereum, analyzing social network data to uncover patterns of behavior and sentiment. [3] This paper proposes a knowledge-enabled BERT model for aspect-based sentiment analysis, which

incorporates external knowledge to improve the accuracy of sentiment analysis. [4] This paper investigates the use of Twitter sentiment analysis for predicting cryptocurrency prices, using machine learning models to analyze tweet data. [5] This paper presents a method for tracking the evolution of Ethereum transaction networks using link prediction, which can help identify patterns and anomalies in transaction behavior. [6] This paper examines the applications of cryptocurrency, focusing on Ethereum and its use cases in various industries. [7] This paper proposes a BERT framework for sentiment analysis of tweets, achieving high accuracy on a benchmark dataset. [8] This paper investigates sentiment analysis of cryptocurrency based on comments and replies on Twitter, using machine learning models to classify sentiment. [9] This paper proposes a BERT-based approach for extracting investor sentiment from news articles, Reddit posts, and tweets, and uses this sentiment to predict cryptocurrency returns. [10] This paper also proposes a BERT-based approach for predicting cryptocurrency returns using investor sentiment extracted from news articles, Reddit posts, and tweets.

METHODOLOGY

This study uses a patent intellectual property from BERT. The research uses the topic of artificial intelligence in natural language processing (NLP), sentiment from a website. Researchers used 70 patent documents and 45 simple patent, intellectual property families from the Lens database (<https://www.lens.org/>) through online searches in March 2023. The keyword search queries for patents in Lens in the title, abstract and claims are Patents (BERT AND (NLP AND sentiment)). Data was not limited because patent data used from 2019 to 2023. The study analyzed the patent landscape, namely the number of annual patent growth, legal status, top applicants, top owners, top CPC classification codes, patent documents by Jurisdiction, top investor, top agents & attorneys, and top cited patents. The Lens site

provides an analysis function that displays patent landscape information on a selected topic. Researchers use this service to analyze and visualize NLP sentiment with BERT.

Web Scraping

In this experiment, we use the web scraping method with the BeautifulSoup library. We combined the html of the website with this BeautifulSoup. After that, we input the data via pandas into csv.

Data Analysis

We use replace to clean the dataset from backslashes and numerical values. For Data Preprocessing we use the tokenizer from the BERT model pre-trained to tokenize. With tokenization we can also do segmentation, so that machines can learn more quickly with the labels given. Then we use the BERT method to train the machine.

EXPERIMENT

Experimental Setup

We use Google Colab Pro to do all of these experiments, because it's faster and can be done anywhere without having to meet each other. For web scraping, we did it on a laptop with Intel I7 and a Windows 10 system implemented in Python BeautifulSoup using Jupyter Notebook.

Link:

<https://colab.research.google.com/drive/1Ic9qiTeDIGS4Qz5LIUXrxkd4ajF5cGtK>

Dataset

For Dataset, we use the web scraping method with the BeautifulSoup library. We combined the html of the website with this BeautifulSoup.

```
#loop through the pages
for page in range(1,500):
    if(page > 1):
        url = "https://www.tradingview.com/symbols/ETHUSD/ideas/" + "page-"
        + str(page) + "/"
    else:
        url = "https://www.tradingview.com/symbols/ETHUSD/ideas/"

    content = requests.get(url)
    soup = BeautifulSoup(content.text, 'html.parser')

    items = soup("div", class_="tv-widget-idea js-userlink-popup-anchor")
    for item in items:
        if(item.find("span", class_="content-TRXznVu1 badge-idea-content-
fwzOPd3k") != None):
            titles.append(item.find("a", class_="tv-widget-idea__title
apply-overflow-tooltip js-widget-idea__popup").get_text())
            action.append(item.find("span", class_="content-TRXznVu1 badge-
idea-content-fwzOPd3k").get_text())
            desc.append(item.find("p", class_="tv-widget-idea__description-
row tv-widget-idea__description-row--clamped js-widget-
idea__popup").get_text())
```

Then, we input the data with pandas into csv.

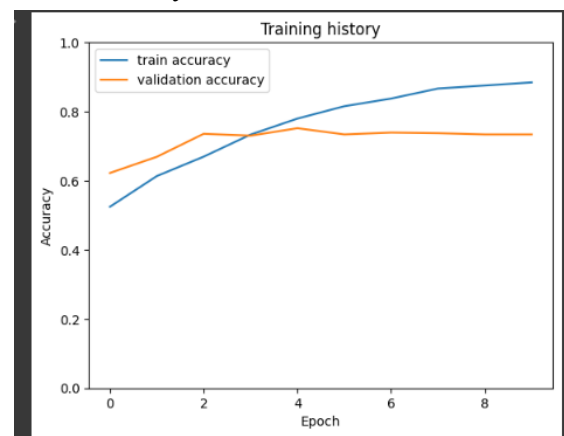
Evaluation Metrics

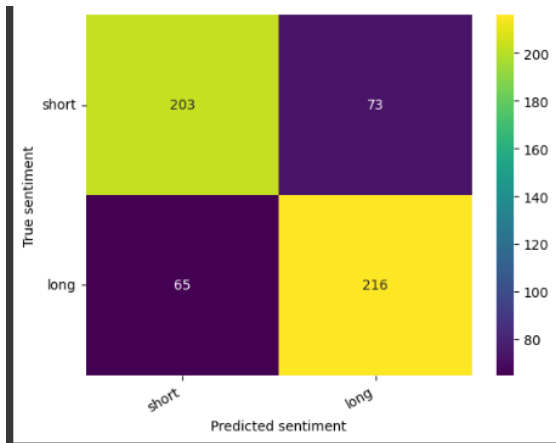
To Evaluate the performance of our proposed approach, we used the following evaluation metrics:

1. Accuracy: The percentage of correctly classified samples
2. Precision : The ratio of true positive samples

Result

In this project, our model almost experienced overfitting because our training data experienced a constant increase above 90%, but the validation that we did stagnated and even tended to decrease in epoch 5 and above. But on the accuracy of the search we did, we got an accuracy value above 75%. This is good, because with this accuracy value, we can ensure that this model is ready for use by the community.





	precision	recall	f1-score	support
short	0.76	0.74	0.75	276
long	0.75	0.77	0.76	281
accuracy			0.75	557
macro avg	0.75	0.75	0.75	557
weighted avg	0.75	0.75	0.75	557

CONCLUSION

The accuracy of our project is up to 75% and above, so it can be concluded that this project is successful and can be reused.