JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 2, Juni 2025, Pp. 1087-1098

CRYPTO NARRATIVES SENTIMENT ANALYSIS ON BITCOIN PRICE PREDICTION USING THE NAIVE BAYES METHOD

Didik Nuryadi *[1)], Daniel H.F. Manongga[2)], Irwan Sembiring[3)]
1. Faculty of Information Technology, Satya Wacana Christian University, Indonesia
2. Faculty of Information Technology, Satya Wacana Christian University, Indonesia
3. Faculty of Information Technology, Satya Wacana Christian University, Indonesia

Article Info
Keywords: Crypto Narrative; Bitcoin; Naive Bayes Classifier; Data Mining; Sentiment.

Article history:
Received 24 Oktober 2024
Revised 19 November 2024
Accepted 28 Desember 2024
Available online 15 March 2025

DOI :
https://doi.org/10.29100/jipi.v10i2.6116

* Corresponding author.
Didik Nuryadi
E-mail address:
972023708@student.uksw.edu

ABSTRACT
Globalization affects many aspects of human life with consequences that may be positive or negative. Advances in information technology, which significantly assist many human activities, are one of the elements affected. As a new product of financial technology, cryptocurrency has revolutionized the global payment system. Bitcoin has experienced significant price increases in recent years, often caused by economic and psychological market factors. Sentiment analysis of the bitcoin crypto narrative is essential for understanding market behavior and predicting price trends because market sentiment has been proven to influence bitcoin price movements. Therefore, this research aims to investigate the crypto sentiment narrative regarding Bitcoin price movements using a sentiment analysis approach with the Naïve Bayes classification method. The dataset used in this research comes from crypto narratives that are considered to influence bitcoin price movements, which were collected from October 2022 to April 2024. This research succeeded in classifying the data tested using 10-fold cross-validation testing, with an average of 76.13%. The precision score for the positive opinion class was 63.92%, and the precision score for the negative opinion class reached 81.77%. The average recall value for the positive class was 61.69%, and for the negative class, it reached 83.12%. This data shows that Naïve Bayes is quite good at analyzing crypto sentiment narratives regarding bitcoin price movements.

.

I. INTRODUCTION

TODAY, globalization affects many aspects of human life with consequences that may be positive or negative. Advances in information technology, which significantly assist many human activities, are one of the elements affected [1]. As a new product of financial technology, cryptocurrency has revolutionized the global payment system [2]. Although cryptocurrencies can be used as a medium of exchange, unit of account, or store of value, they do not have legal tender status. With more than 10,000 crypto assets in circulation worldwide today, the cryptocurrency market has experienced rapid growth. Technological development and increasing market demand will increase the number of these assets [3].

In 2018, there were 35 million cryptocurrency users, from around 18 million in 2017. In subsequent years, even more significant spikes occurred: 101 million in 2020, 295 million in 2021, and 516 million in 2023 [4]. This phenomenon shows the growing interest in cryptocurrencies, especially Bitcoin, in the changing global economy and the importance of a deep understanding of the technology [5]. As of March 1, 2024, there were 8,965 types of coins circulating on the cryptocurrency market, according to data from Coinmarketcap. This number is expected to continue to increase. Bitcoin has market dominance at 60.14%, followed by Ethereum at 11.5%, and other coins only have 5% of the market. The high interest and demand led to the dominance of Bitcoin [6]. With a market capitalization of over $964.71 billion, Bitcoin is at the center of the cryptocurrency community [7]. The increasing popularity shows a strong interest in the potential and benefits of digital currencies [8]. In a situation like this, thoroughly understanding Bitcoin and the role it plays in the changing world economy is essential. Continuous effort is needed to assess challenges and opportunities. This includes building appropriate regulatory systems to protect consumers and keep the market safe [9]. In 2022, the price of Bitcoin experienced a drastic decline from around $47,000 at the start of the year to around $20,000 in the middle of the year. This decline was influenced by factors such as tight monetary policy from the US Federal Reserve, increasing inflation, and global economic uncertainty due to geopolitical conflicts. Instead, in early 2023, Bitcoin price began to recover and climb back

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 2, Juni 2025, Pp. 1087-1098

above $30,000 as institutional adoption increased and monetary policy eased [10].

Bitcoin has experienced substantial price increases in recent years. This is often caused by economic and psychological market factors [11]. Sentiment analysis of Bitcoin crypto narratives has become essential for understanding market behavior and forecasting price trends because market sentiment has been shown to influence Bitcoin price movements. Researchers and analysts can find the relationship between the latest Bitcoin crypto narratives and market price movements by using sentiment analysis methods [12]. The interaction between Bitcoin and these innovations is seen in various ways, such as integrating Bitcoin in DeFi platforms to facilitate liquidity or using Bitcoin as the underlying asset in NFT transactions. This reflects the further transformation of Bitcoin as a premier digital asset towards a broader ecosystem supporting new financial and creative applications on the blockchain. Thus, understanding how DeFi and NFTs interact with Bitcoin helps in seeing the future development and potential of the cryptocurrency market and the importance of adjusting investment strategies and regulations to accommodate these developments appropriately.

The Naive Bayes classification method is one of the most frequently used techniques in sentiment analysis [13]. It is simple but effective in classifying text based on its sentiment [14]. By using simple assumptions about the independence of features in a text [15], the Naive Bayes classification method is able to provide fairly accurate estimates of the sentiment contained in crypto narratives related to Bitcoin.

Therefore, this research aims to investigate the sentiment of crypto narratives regarding Bitcoin price movements using a sentiment analysis approach with the Naive Bayes classification method. It is hoped that the results of this research will provide deeper insight into the dynamics of the cryptocurrency market and contribute to the development of trading strategies that are more adaptive and responsive to changes in market sentiment. Thus, it is hoped that this research can become a basis for more informative and effective decision-making for stakeholders in the Bitcoin ecosystem.

II. Research Methods

In this research, the author uses crypto narrative sentiment data on Bitcoin price movements. The process of data collection, data pre-processing, word weighting, classification, and finally evaluation is part of the research process. Fig 1 shows the detailed process of this research.

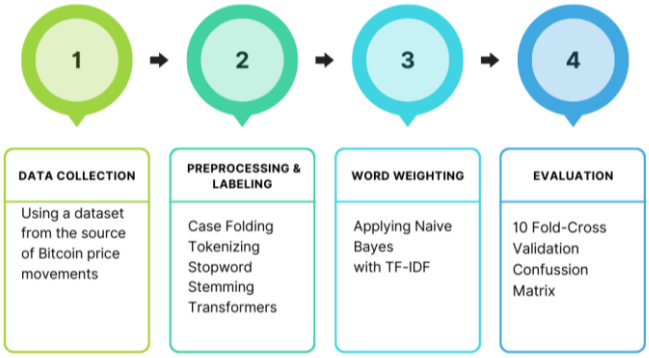

Fig. 1. Research Stages

A. Data Collection

Primary data used in this research comes from the following sources:

1) https://www.investing.com
2) https://www.idx.co.id
3) https://en.wikipedia.org/wiki/2023#References

Five attributes and one class attribute relate to our research, and the information is presented in Table 1.

TABLE I
ATTRIBUTES OF CRYPTO NARRATIVES DATA ON BITCOIN PRICE PREDICTION

1	Crypto Narratives
2	Date
3	Country
4	Source
5	Label

B. Preprocessing & Labeling

Preprocessing and labeling are crucial steps in the data classification process as they seek to remove any irrelevant information in the data to be processed and assign positive or negative labels, which can enhance the

accuracy of the classification findings. This stage is critical for a successful data classification process. The preprocessing stage consists of a number of steps, as stated in [16]:

1) *Folding* Case
   The process of changing all sentences to lowercase so that there are no differences in writing that could affect the classification process [17].
2) *Tokenizing*
   Sentence tokenization is the process of converting a sentence into individual words and removing punctuation. This is done to organize data into smaller parts that are relevant to the classification process [18].
3) *Stopword*
   This process removes important words from sentences that usually do not have special meaning and appear frequently in the language. Considered irrelevant to the classification process, these words are on the dictionary's stopword list [19].
4) *Steaming*
   The steaming process removes affixes or word endings to leave only the base word, reducing the data size and speeding up the classification process [20].
5) *Transformers*
   After carrying out Case Folding, Tokenizing, Stopword and Stemming, the next step is Labeling using a Transformer. Labeling is done last in preprocessing because the data is immaculate after going through these four stages. Transformer is a machine learning architecture that uses attention mechanisms to understand context and relationships between words in text. This model is very effective for various tasks in language processing, such as translation, text classification, and sentiment analysis. One of the main advantages of transformers in language processing applications is their ability to simultaneously consider the context of the entire sentence through attention mechanisms. Despite their advantages of high efficiency and accuracy, transformers also require significant training and implementation resources and high implementation complexity.

Five preprocessing processes are typically utilized in text analysis. There is, however, a further phase known as lemmatization. As an alternative, the stemming method is used in this work. This is due to the fact that the stemming process is executed by the Natural Language Processing (NLP) library and the dataset utilized is in English. Lemmatization, meanwhile, is challenging due to the scarcity of English-language resources. Consequently, it was decided that stemming was a suitable preprocessing technique for this study.

C. Word Weighting

During the preprocessing phase, a compilation of words or keywords is generated. Afterwards, the next step entails employing a method known as term weighting, where each word or phrase is given a weight or value to indicate its importance within the document. To evaluate the occurrence and similarity of terms in each document, term values or weights are calculated for each document [21]. The frequency of a term in a document collection is directly proportional to the value or weight ascribed to that phrase. After completing the weighting stage, the next step is to carry out the classification process using the recommended algorithm. The Tf-Idf technique is used to assign weights throughout the term weighting process. This approach is advantageous because it emphasizes terms frequent in a particular document but rare in the overall corpus, thus highlighting their discriminative power in distinguishing documents.

The term frequency (TF) technique determines the weight of a document by counting the occurrence of a certain term or phrase inside the document. The frequency of a specific term or word has a direct impact on the importance of the text in respect to that term or phrase, and vice versa. Inverse Document Frequency (IDF) is a technique employed to analyze the occurrence rate of a specific term across a corpus of textual data in comments. In the Inverse Document Frequency (IDF) measure, words with lower frequency in the document collection are considered more meaningful or useful. The computation of Idf is conducted utilizing the prescribed formula (1).

$$Idf = \log\left(\frac{jumlah\ seluruh\ dokumen\ dalam\ koleksi}{jumlah\ dokumen\ yang\ mengandung\ istilah}\right) \quad (1)$$

Tf-Idf is the product obtained by multiplying the value of the Term Frequency (Tf) with the Inverse Document Frequency (IDF). The calculation can be performed utilizing the Tf and Idf formulas.

D. Naive Bayes Classifier

The Naive Bayes classification algorithm is frequently employed in machine learning. This algorithm operates by computing the potential probabilities of each class for every data instance and subsequently choosing the class with the highest probability as the forecast. This method also includes the computation of posterior probabilities,

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 2, Juni 2025, Pp. 1087-1098

which ascertain the likelihood that a data instance belongs to a specific class, taking into account the probability distribution of the class and the observed attributes [22].

$$P(c|d) = \frac{P(c) * P(d|c)}{P(d)} \quad (2)$$

The information:
P(c|d) : posterior, namely the probability of class c given document d.
P(c) : prior or initial opportunity for the emergence of category c.
P(d|c) : Likelihood Value
P(d) : Evidence or opportunities for documents to appear d.

E. Evaluation

An essential aspect of this research involves assessment, which assesses the performance and effectiveness of the model. The confusion matrix is a significant tool for analyzing classification outcomes since it offers precise information on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values. This study will utilize ten-fold cross-validation to partition the data and evaluate the effectiveness of the model or algorithm. The resulting value in the confusion matrix is larger, suggesting that the developed model is superior and more accurate in the classification process. Accuracy, precision, and recall are three important metrics used to evaluate the performance of a model or system. An exemplary recommendation system might be considered outstanding if it has elevated levels of recall, accuracy, and precision. Formula (4) denotes the mathematical equation employed for computing the accuracy measure. Formula (5) denotes the mathematical equation employed for computing the precision value. Formula (6) denotes the mathematical equation employed to compute the recall value of the monitored group.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \text{x } 100\% \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \text{x } 100\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \text{ x } 100\% \quad (6)$$

Explanation: 1). True Positive (TP): When the predictions of existing categories and the system categories from the same comment accurately match and coincide, it is considered a true positive. 2). True Negative (TN): When the predictions made by the system and the existing categories for comment do not match, it is considered a true negative. This means that the accuracy of the system is different from the expected outcome. 3). False Positive (FP): This occurs when a comment is predicted to belong to a specific category by the system, but it is in a different category. 4). False Negative (FN): When a comment is predicted to belong to a specific category by the algorithm, but it actually does not belong to that group, it is considered a false negative.

F. Simulation of a specific case

The Naive Bayes approach is employed in a case simulation to forecast Bitcoin price changes by classifying the sentiment of crypto narratives. This simulation assumes the acquisition and analysis of crypto narratives, followed by the identification of attitudes. The purpose is to deliver information to the general public and economic actors, specifically.

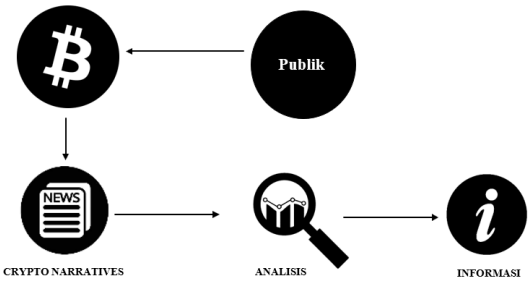

Fig. 2. Simulation of a specific case

The sentiment analysis of cryptocurrency narratives is conducted in the case scenario depicted in Fig 2. Once sentiment crypto tales are acquired and examined, the crypto narratives that are believed to impact Bitcoin price fluctuations will be determined. Subsequently, crypto tales from several sources worldwide will be gathered,

encompassing crucial details such as the content of the narratives, their dates, the countries they originate from, and the sources of these narratives. The gathered crypto tales will be meticulously preserved and, after that, subjected to comprehensive investigation. This analysis will offer compelling data that can serve as a valuable lesson for both the general public and economic actors. The findings of this investigation can assist Bitcoin participants in mitigating potential future financial setbacks.

III. Results and Analysis

A. Data Sets

The dataset utilized in this study originates from crypto narratives, which are acknowledged to exert an impact on the fluctuations of Bitcoin prices. The dataset was gathered between October 2022 and April 2024 and then annotated using the Python huggingface module. The dataset has 368 records, each consisting of 3657 words. The labeling of each record was done using the Python module called huggingface. The dataset has 368 records, each containing 3657 words, and every record has been assigned a label or category. The dataset comprises five variables. However, only the crypto narratives, date, and country factors are pertinent to the objectives of this research. The dataset utilized in this investigation is incorporated in Table 2.

TABLE II
METADATA CRYPTO NARRATIVES

Id	Crypto narratives	Country
1	Three scientists share physics Nobel prize for quantum mechanics work	Global
2	Nobel Prize in Chemistry Is Awarded to 3 Scientists for Work 'Snapping Molecules Together'	Global
3	Annie Ernaux wins the 2022 Nobel prize in literature	Global
4	Nobel peace prize 2022 awarded to human rights campaigners in Ukraine, Russia and Belarus	Global
5	A noble award for a 'popular misconception	Global
6	COP27: World on track to increase emissions 10.6% by 2030 - UN report	Global
7	What Is COP27? And Other Questions About the Big U.N. Climate Summit	Global
8	Ethiopia and Tigray sign peace agreement in bid to end devastating war	Africa
9	Denmark election: Centre-left bloc comes out on top	Denmark
10	Danish left-wing bloc retains a majority in parliament	Global
up to	…	…
368	Isfahan, Iran: Explosions Hear, Reports Say	Iran
368	Isfahan, Iran: Explosions Hear, Reports Say	Iran

B. Preprocessing & Labeling

In order to expedite and streamline the process of categorizing data, this study performed computations on a dataset consisting of 368 records that had a total of 3652 words from crypto narratives that were collected. Subsequently, a series of preprocessing and labeling phases are conducted prior to inputting the dataset into the suggested model, specifically:

1) Casefolding

The Casefolding process of changing all text to lowercase helps equalize words that should be identical but written in different forms. Table 3 shows the results of changes in the case folding stage. The first column of the table consists of the ID, the second column consists of the crypto narratives, and the third column contains the nation. The distinction between Table 2 and Table 3 is in the alteration of the first capital letters to lowercase ones. For instance, the case folding procedure is employed to convert. For instance, the term "Work" is transformed into "work". This will reduce unnecessary variations in the data. The results of the case folding process are used to convert crypto narratives into small word forms, making it easier to compose the text.

TABLE III
CASEFOLDING

Id	Crypto narratives	Country
1	three scientists share physics nobel prize for quantum mechanics work	global

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 2, Juni 2025, Pp. 1087-1098

2	nobel prize in chemistry is awarded to 3 scientists for work 'snapping molecules together'	global
3	annie ernaux wins the 2022 nobel prize in literature	global
4	nobel peace prize 2022 awarded to human rights campaigners in ukraine, russia and belarus	global
5	a noble award for a 'popular misconception	global
6	cop27: world on track to increase emissions 10.6% by 2030 - un report	global
7	what is cop27? and other questions about the big u.n. climate summit	global
8	ethiopia and tigray sign peace agreement in bid to end devastating war	africa
9	denmark election: centre-left bloc comes out on top	denmark
10	danish left-wing bloc retains a majority in parliament	global
up to	…	…
368	isfahan, iran: explosions hear, reports say	iran

2) Tokenizing

The tokenizing process breaks down text into individual tokens (words or phrases), allowing for a more detailed analysis of each text component. Tokenization makes it easier to identify patterns and word frequencies. Table 4 displays the results of the tokenizing stage, with a column containing crypto narratives. The difference between Table 3 and Table 4 is the change in the comma (",") used as a separator between words in each phrase. The result of this tokenizing technique is to divide sentences into separate words, which makes it easier to eliminate unimportant keywords.

TABLE IV
TOKENIZING

Id	Crypto narratives	Country
1	'three', 'scientists', 'share', 'physics', 'nobel', 'prize', 'for', 'quantum', 'mechanics', 'work'	global
2	'nobel', 'prize', 'in', 'chemistry', 'is', 'awarded', 'to', 'scientists', 'for', 'work', 'snapping', 'molecules', 'together'	global
3	'annie', 'ernaux', 'wins', 'the', 'nobel', 'prize', 'in', 'literature'	global
4	'nobel', 'peace', 'prize', 'awarded', 'to', 'human', 'rights', 'campaigners', 'in', 'ukraine', 'russia', 'and', 'belarus'	global
5	'noble', 'award', 'for', 'popular', 'misconception'	global
6	'cop', 'world', 'on', 'track', 'to', 'increase', 'emissions', 'by', 'un', 'report'	global
7	'what', 'is', 'cop', 'and', 'other', 'questions', 'about', 'the', 'big', 'un', 'climate', 'summit'	global
8	'ethiopia', 'and', 'tigray', 'sign', 'peace', 'agreement', 'in', 'bid', 'to', 'end', 'devastating', 'war'	africa
9	'denmark', 'election', 'centreleft', 'bloc', 'comes', 'out', 'on', 'top'	denmark
10	'danish', 'leftwing', 'bloc', 'retains', 'majority', 'in', 'parliament'	global
up to	…	…
368	'isfahan', 'iran', 'explosions', 'hear', 'reports', 'say'	iran

3) Stopword

The Stopword process removes common words that do not provide significant information about sentiment. This helps reduce noise in the data and improves analysis efficiency. Table 5. The differentiating characteristic between Table 4 and Table 5 is in the use of shared terminology such as "for", "in", "is", "to", "the", "and", "on", "by", "what", "other", "about", "out", and so forth. These words have been removed from the crypto narratives sentence, which is significant for data analysis and classification. Thus, only words that are considered essential and relevant will be used in the next stage.

TABLE V

STOPWORD

Id	Crypto narratives	Country
1	three scientists share physics nobel prize quantum mechanics work	global
2	nobel prize chemistry awarded scientists work snapping molecules together	global
3	annie ernaux wins nobel prize literature	global
4	nobel peace prize awarded human rights campaigners ukraine russia belarus	global
5	noble award popular misconception	global
6	cop world track increase emissions un report	global
7	cop questions big un climate summit	global
8	ethiopia tigray sign peace agreement bid end devastating war	africa
9	denmark election centreleft bloc comes top	denmark
10	danish leftwing bloc retains majority parliament	global
up to	…	…
368	isfahan iran explosions hear reports say	iran

4) Stemming

The stemming process reduces words to their basic form, allowing the grouping of words with the same meaning but different forms. The stemming stage in this research uses the NLTK Porter Stemmer library, as seen in Table 6. The difference between Table 5 and Table 6 is in the exact language used, such as the application of the NLTK Porter Stemmer library for this purpose. The distinction between Table 5 and Table 6 is characterized by the substitution of certain words, such as "scientists", "physics", "mechanics", "chemistry", "awarded", "snapping", "molecules", and "wins". These words have been changed into crypto narrative sentences. The aim of stemming is to remove affixes and word endings to obtain the appropriate form of the primary word. The stemming results make all words with the same root uniform, making it easier to classify and analyze the data. It helps simplify the model and increase accuracy.

TABLE VI
STEMMING

Id	Crypto narratives	Country
1	three **scientist** share **physic** nobel prize quantum **mechanic** work	global
2	nobel prize **chemistri award scientist** work snap **molecul** together	global
3	annie ernaux **win** nobel prize **literatur**	global
4	nobel **peac** prize **award** human right **campaign** ukrain russia belaru	global
5	**nobl** award popular **misconcept**	global
6	cop world track **increas emiss** un report	global
7	cop **question** big un **climat** summit	global
8	ethiopia tigray sign **peac** agreement bid end **devast** war	africa
9	denmark elect centreleft bloc come top	denmark
10	danish **leftw** bloc **retain major** parliament	global
up to	…	…
368	isfahan iran **explos** hear report say	iran

5) Transformers

The labeling process simplifies and speeds up the system's work. Using manual labeling will take quite a long time and require linguistic experts who are qualified in their field. Therefore, in this research, we utilize the Transformer model. The Transformer model can categorize text as positive or negative. This research uses resources from Huggingface to carry out this process. The model used is DistilBERT, which has been finetuned

on the SST-2 dataset for sentiment analysis. This model was chosen because it has superior performance with high accuracy, precision, recall on the results of the evaluation carried out in natural language understanding tasks. Table 7 shows the metadata resulting from labeling. The labeling process here uses the Python programming language.

TABLE VII
TRANFORMERS

Id	Crypto narratives	Label
1	Three scientists share physics Nobel prize for quantum mechanics work	Positive
2	Nobel Prize in Chemistry Is Awarded to 3 Scientists for Work 'Snapping Molecules Together'	Positive
3	Annie Ernaux wins the 2022 Nobel prize in literature	Positive
4	Nobel peace prize 2022 awarded to human rights campaigners in Ukraine, Russia and Belarus	Positive
5	A noble award for a 'popular misconception	Positive
6	COP27: World on track to increase emissions 10.6% by 2030 - UN report	Negative
7	What Is COP27? And Other Questions About the Big U.N. Climate Summit	Negative
8	Ethiopia and Tigray sign peace agreement in bid to end devastating war	Negative
9	Denmark election: Centre-left bloc comes out on top	Positive
10	Danish left-wing bloc retains a majority in parliament	Positive
up to	…	…
368	Isfahan, Iran: Explosions Hear, Reports Say	Negative

The preprocessing and labeling processes have effectively eliminated unwanted interference, standardized the representation of words in documents, and assigned positive and negative labels to crypto narratives. Consequently, the suggested classification technique allows for the compression of documents that were previously longer, making them more organized and suitable for additional research. Once the data has completed the preprocessing and labeling steps, the classification and evaluation procedure may be conducted with more efficiency and accuracy. This technique aims to determine the presence of sentiment related to crypto narratives around Bitcoin price fluctuations.

C. Word Weighting

After completing the preprocessing stage for a dataset of 368 entries, the terms are transformed into vector data using the Tf*Idf multiplication technique. Table 8 displays the list of terms. This vector data will serve as the foundation for the subsequent procedure of determining the sentiment of crypto narratives regarding Bitcoin price fluctuations.

TABLE VIII
PREPROCESSING TERMS

No	Term	No	Term
1	about	9	agreement
2	accept	10	allow
3	accord	11	america
4	actor	12	among
5	admit	13	animos
6	africa	14	answer
7	after	…	…
8	against	1377	zambia

In order to continue, it is imperative to compute the multiplication of the Term Frequency (Tf) and the Inverse Document Frequency (IDF). The results of the Tf*idf method are presented in Table 9. The purpose of this technique is to allocate a weight or value that signifies the importance of each word or term inside the document in the dataset used. Consequently, it will be more convenient to categorize and locate the emotion of crypto narratives around Bitcoin price fluctuations.

TABLE IX

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 2, Juni 2025, Pp. 1087-1098

TF-IDF RESULTS

Id	about	accept	accord	actor	s/d	zambia
1	0	0	0	0	.	0
2	0	0	0	0	.	0
3	0	0	0	0	.	0
4	0	0	0	0.421	.	0
5	0	0	0	0.500	.	0
6	0	0	0	0	.	0
7	0.357	0	0	0	.	0
8	0	0	0	0	.	0.212
9	0	0	0.455	0	.	0
10	0	0	0	0	.	0
s/d	…	…	…	…	.	0
368	0	0	0	0	.	0

D. Naive Bayes Classifier

Once the preprocessing and labeling stages are finished and the Tf*Idf multiplication is performed to transform words into data vectors, the subsequent step involves conducting testing using the Naïve Bayes method for sentiment analysis of crypto narratives in this research. The research was conducted ten times using various threshold values, specifically 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. The objective of this test is to determine how the classification model categorizes the sentiment of crypto narratives as either influential or non-influential to Bitcoin price fluctuations, using various threshold levels.

The categorization of crypto tales sentiment using Naïve Bayes yielded an accuracy of 75.97% in the first test, where the sentiment value was set to 1. The accuracy of predicting positive opinions is 65.71%, and for negative opinions it is 67.74%. Furthermore, the system's efficacy in obtaining information also fluctuates, namely with a success rate of 69.70% for positive opinions and 63.64% for negative views. The purpose of this study is to discover possible enhancements to the model and enhance the accuracy of sentiment classification for crypto narratives in relation to Bitcoin price fluctuations.

	true POSITIVE	true NEGATIVE	class precision
pred. POSITIVE	73	42	65.71%
pred. NEGATIVE	45	202	67.74%
class recall	69.70%	63.64%	

Fig. 3. First test of the Naïve Bayes Model, with a value of 1

E. Evaluation

The test results indicate that the accuracy of each categorization model varies based on the test value utilized, which falls within the range of 1 to 10. Table 11 presents a comparison of the data obtained from 10 tests, displaying the accuracy value of the Naïve Bayes algorithm for each test threshold value. The table demonstrates that the accuracy of the two methods varies, and the performance of the classification model can be influenced by the utilization of different threshold testing values. This evaluation offers valuable information regarding the efficiency and efficacy of the strategy in categorizing messages pertaining to the sentiment of crypto narratives on Bitcoin price fluctuations.

TABLE X
COMPARISON OF CLASSIFICATION MODEL TESTING ACCURACY VALUES

Model	Testing	Accuracy	Classification
Naïve Bayes Classifier	1	75,97%	Fair
	2	72,93%	Fair
	3	75,69%	Fair
	4	75,69%	Fair
	5	76,24%	Fair
	6	79,01%	Fair
	7	77,07%	Fair
	8	75,97%	Fair
	9	76,52%	Fair
	10	76,24%	Fair

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 2, Juni 2025, Pp. 1087-1098

Average	**76,13%**	**Fair**

Table 10 functions as a performance evaluation metric to verify the effectiveness of the built classification model. Table 11 displays the mean values of precision and recall for this classification technique. The evaluation findings for the Naïve Bayes technique indicate that the average precision for the negative opinion class is 81.77%. In comparison, the recall for the negative opinion class is 83.12%.

TABLE XI
MEASUREMENT EVALUATION

Model	Testing	Precision		Recall	
		+	-	+	-
naïve bayes	1	63,48%	81,78%	61,86%	82,79%
	2	58,20%	80,42%	60,17%	79,10%
	3	63,39%	81,20%	60,17%	83,20%
	4	62,93%	81,71%	61,86%	82,38%
	5	64,55%	81,35%	60,17%	84,02%
	6	68,75%	83,60%	65,25%	85,66%
	7	65,22%	82,59%	63,56%	83,61%
	8	63,48%	81,78%	61,86%	82,79%
	9	64,86%	81,67%	61,02%	84,02%
	10	64,29%	81,60%	61,02%	83,61%
Average		**63,92%**	**81,77%**	**61,69%**	**83,12%**

Based on the research results, it can be concluded that using the Naïve Bayes method is quite good in analyzing the sentiment of crypto narratives on Bitcoin price movements. The Naïve Bayes method achieved an average accuracy of 76.13%. In addition, the average precision value for the positive opinion class shows a high classification rate of 63.92%, while the precision for the negative opinion class reaches 81.77%. These results show that classification using this method is effective compared to previous research, which used the Naive Bayes method to analyze public sentiment towards crypto coins on Twitter in Indonesia. Previous research using dataset 234 and carrying out five tests produced an average accuracy of 68.51% [23]. Therefore, the Naïve Bayes approach is strongly advised for sentiment analysis of Bitcoin price changes in crypto narratives. In addition to achieving a high classification rate for both classes, this method is characterized by its simplicity and excellent overall performance.

Id	label	prediction(la...	.. ..	text	Tanggal
3	POSITIVE	POSITIVE	. .	anni ernaux nobel prize literatur	Oct 6, 2022
9	POSITIVE	POSITIVE	. .	denmark elect centreleft bloc come	Nov 2, 2022
10	POSITIVE	POSITIVE	. .	danish leftw bloc retain major parliament	Nov 2, 2022
11	NEGATIVE	NEGATIVE	. .	offici version ethiopia peac agreement	Nov 4, 2022
12	NEGATIVE	POSITIVE	. .	bankrupt stun revers crypto exchang	Nov 11, 2022
13	NEGATIVE	NEGATIVE	. .	russian forc retreat from ukrainian citi kher...	Nov 11, 2022

Fig. 4. The example ID is indicated as a crypto narratives sentiment towards Bitcoin price movements.

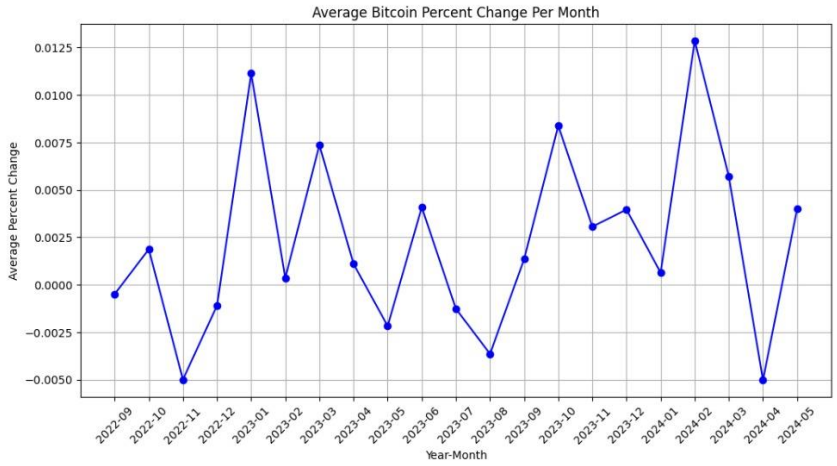

Fig. 5. Bitcoin chart 2022 to 2024.

The acquired categorization results are displayed in Fig 4. The image displays five columns of data: the first column presents the ID, the second column presents the label, the third column presents the classification prediction, the fourth column presents text or cryptographic tales about Bitcoin, and the fifth column presents the date. The data in the graphic provides a clear example that there are comments classified as negative. The research findings suggest that crypto narratives labeled as ID 11 have a pessimistic tone, explicitly referring to November 2022. According to Figure 5, Bitcoin prices had a decrease in November 2022. This information can serve as a point of reference for analyzing the market value of Bitcoin on a specific date, month, and year, as well as for

monitoring the fluctuations in Bitcoin prices. Fig 5 displays a graph illustrating the notable volatility of Bitcoin's price over some time. The graph clearly shows the up-down patterns that correspond to periodic price variations. Research on sentiment in crypto narratives indicates that fluctuations in the price of Bitcoin are driven by the emergence of positive and negative attitudes within the crypto community. There is a robust association between fluctuations in sentiment and price movements, whereby price declines and gains frequently transpire subsequent to the appearance of positive or negative sentiment within a given month. During periods of prevailing optimistic emotion, Bitcoin prices typically undergo substantial gains, presenting investors with profitable possibilities. In contrast, during periods of widespread negative emotion, Bitcoin prices might significantly decrease, causing worry among holders of crypto assets. Investors can enhance their decision-making and anticipate market fluctuations by analyzing Bitcoin price trends and constantly tracking crypto sentiment narratives. This underscores the significance of sentiment analysis in trading and investment strategies inside the ever-changing cryptocurrency market.

Several challenges were encountered during the analysis of English text in this research. This is due to the existence of numerous comprehensive English corpora. In addition, crypto tales frequently employ colloquial language. This impediment hinders the achievement of the appropriate level of accuracy in the findings obtained from the data mining process.

IV. Conclusion

This study conducted comparative testing using 10-fold cross-validation to analyze crypto narratives on Bitcoin price movements through a categorization method. Uniformly, the pre-processing and data labeling processes are performed on all data. Following the pre-processing and labeling stages, the data undergoes processing using the Naïve Bayes approach to assess the performance of 10-fold cross-validation and determine the average accuracy. The analysis results indicate that the Naïve Bayes technique exhibits variations in accuracy and recall, particularly in the context of negative opinion sentiment, where it achieves a relatively high score. The Naïve Bayes approach yields a mean accuracy of 76.13%. Hence, the Naïve Bayes approach is the most suitable option for doing sentiment analysis on crypto narratives related to Bitcoin price fluctuations within the scope of this study.

In addition, when analyzing crypto narratives related to Bitcoin price movements, it is observed that crypto narratives labeled ID 11 exhibit a pessimistic emotion in November 2024. During that month, the price of Bitcoin also witnessed a decrease. This information can offer valuable perspectives in examining Bitcoin market trends and dynamics.

V. Suggestion

In future studies, it is anticipated that barriers in English language text analysis can be overcome by expanding the corpus with additional relevant and current text data. Furthermore, incorporating text sources from different domains associated with crypto narratives can also contribute to this endeavor. Furthermore, incorporating frequently updated jargon and jargon dictionaries, employing advanced preprocessing methods, and utilizing intricate machine learning or deep learning models can enhance the precision of sentiment analysis. Using datasets that have been labeled by experts in the field can also be a valuable measure to strengthen the accuracy of predictions. Therefore, it is anticipated that additional research can yield more precise outcomes in the sentiment analysis of crypto narratives regarding Bitcoin price fluctuations.

References

[1] A. Wijayanto, I. Riadi, Y. Prayudi, And T. Sudinugraha, "Network Forensics Against Address Resolution Protocol Spoofing Attacks Using Trigger, Acquire, Analysis, Report, Action Method," *Regist. J. Ilm. Teknol. Sist. Inf.*, Vol. 8, No. 2, Pp. 156–169, 2022, Doi: 10.26594/Register.V8i2.2953.

[2] S. Alam, M. Jamil, And A. Syamsir, "Digital Currency In Indonesia (Prospects And Challenges In Inclusive Financial Reviews)," *J. Ad'ministrare*, Vol. 9, No. 2, P. 515, 2022, Doi: 10.26858/Ja.V9i2.39498.

[3] R. Iyer, "New Evidence On Spillovers Between Crypto Assets And Financial Markets," *Imf Work. Pap.*, Vol. 2023, No. 213, P. 1, 2023, Doi: 10.5089/9798400256622.001.

[4] R. Shewale, "Blockchain Statistics," *Demandsage*, 2024. Https://Www.Demandsage.Com/Blockchain-Statistics/

[5] M. Campbell-Verduyn, *Bitcoin And Beyond: Cryptocurrencies, Blockchains, And Global Governance*. 2018. Doi: 10.4324/9781315211909.

[6] O. Sattarov, H. S. Jeon, R. Oh, And J. D. Lee, "Forecasting Bitcoin Price Fluctuation By Twitter Sentiment Analysis," *2020 Int. Conf. Inf. Sci. Commun. Technol. Icisct 2020*, Vol. 02, No. 1, Pp. 1–12, 2020, Doi: 10.1109/Icisct50599.2020.9351527.

[7] H. Jun Zhang, Y. Hui Chen, And H. Zhuo, "Unit Middleware For Implementation Of Human–Machine Interconnection Intelligent Ecology Construction," *J. Big Data*, Vol. 10, No. 1, 2023, Doi: 10.1186/S40537-023-00787-4.

[8] D. Horváth, "Money In The Digital Age: Exploring The Potential Of Central Bank Digital Currency With A Focus On Social Adaptation And Education," *Sustain. Futur.*, Vol. 6, No. June, 2023, Doi: 10.1016/J.Sftr.2023.100136.

[9] L. Liestyowati, E. Sudarmanto, H. Ramadhani, S. Rijal, And T. W. Nurdiani, "Tren Investasi Aset Digital: Studi Tentang Perilaku Investor Muda Terhadap Cryptocurrency Di Tengah Perubahan Pasar Keuangan Di Kota Bandung," *J. Akunt. Dan Keuang. West Sci.*, Vol. 2, No. 03, Pp. 142–149, 2023, Doi: 10.58812/Jakws.V2i03.639.

[10] "Are Crypto Markets Correlated With Macroeconomic Factors?," *S&P Global*, 2023. Https://Www.Spglobal.Com/En/Research-Insights/Special-Reports/Are-Crypto-Markets-Correlated-With-Macroeconomic-Factors (Accessed Jul. 18, 2024).

[11] F. Kjaerland, M. Meland, A. Oust, And V. Øyen, "How Can Bitcoin Price Fluctuations Be Explained?," *Int. J. Econ. Financ. Issues*, Vol. 8, No. 3, Pp. 323–332, 2018, [Online]. Available: Http:Www.Econjournals.Com

[12] E. T. Yolanda, L. Junaedi, And A. Bimo Gumelar, "Analisis Sentimen Pergerakan Harga Mata Uang Kripto (Cryptocurrency) Menggunakan Textblob-Nltk (Natural Language Toolkit)," *Jlk*, Vol. 5, No. 2, Pp. 44–50, 2022, [Online]. Available: Https://T.Co/Kbvjupjoyu

[13] W. Purbaratri, H. D. Purnomo, D. Manongga, And I. Setyawan, "Sentiment Analysis Of Electronic Government Services Utilizing The Naive Bayes Algorithm," *Matrik J. Manajemen, Tek. Inform. Dan Rekayasa Komput.*, Vol. 23, No. 2, Pp. 441–452, 2024, Doi: 10.30812/Matrik.V23i2.3272.

[14] K. A. Baihaqi *Et Al.*, "A Comparison Support Vector Machine, Logistic Regression And Naïve Bayes For Classification Sentimen Analisys User Mobile App," *Int. J. Artif. Intell. Res.*, Vol. 7, No. 1, P. 64, 2023, Doi: 10.29099/Ijair.V7i1.962.

[15] M. F. P. Alam, D. Manongga, I. Sembiring, And W. Sulistyo, "Sentiment Analysis For User Review Polri Hospital Registration App," *Proceeding Perbanas Int. Semin. Econ. Business, Manag. Account. It*, Pp. 143–152, 2023.

[16] Rsa, "2016: Current State Of Cybercrime," P. 7, 2016, [Online]. Available: Https://Www.Rsa.Com/Content/Dam/Rsa/Pdf/2016/05/2016-Current-State-Of-Cybercrime.Pdf

[17] A. Muhariya, A. Riadi, And I. Prayudi, "Cyberbullying Analysis On Instagram Using K-Means Clustering," *Juita J. Inform.*, Vol. 10, No. 2, Pp. 261–271, 2022, Doi: 10.30595/Juita.V10i2.14490.

[18] I. Ruthven And M. Lalmas, "A Survey On The Use Of Relevance Feedback For Information Access Systems," *Knowl. Eng. Rev.*, Vol. 18, No. 2, Pp. 95–145, 2003, Doi: 10.1017/S0269888903000638.

[19] S. K. Sahu, S. Sarangi, And S. K. Jena, "A Detail Analysis On Intrusion Detection Datasets," *Souvenir 2014 Ieee Int. Adv. Comput. Conf. Iacc 2014*, Pp. 1348–1353, 2014, Doi: 10.1109/Iadcc.2014.6779523.

[20] A. Muhariya, I. Riadi, Y. Prayudi, And I. A. Saputro, "Utilizing K-Means Clustering For The Detection Of Cyberbullying Within Instagram Comments," *Ing. Des Syst. D'information*, Vol. 28, No. 4, Pp. 939–949, Aug. 2023, Doi: 10.18280/Isi.280414.

[21] P. Yugianus, H. S. Dachlan, And R. N. Hasanah, "Pengembangan Sistem Penelusuran Katalog Perpustakaan Dengan Metode Rocchio Relevance Feedback," *J. Eeccis*, Vol. 7, No. 1, Pp. 47–52, 2013.

[22] H. M. A. Ishara Amali And S. Jayalal, "Classification Of Cyberbullying Sinhala Language Comments On Social Media," *Mercon 2020 - 6th Int. Multidiscip. Moratuwa Eng. Res. Conf. Proc.*, Pp. 266–271, 2020, Doi: 10.1109/Mercon50084.2020.9185209.

[23] Z. C. Waskita And R. R. Santika, "Analisis Sentimen Masyarakat Terhadap Endemi Pada Twitter Menggunakan Metode Naive Bayes," *Pros. Semin. Nas. Mhs. Fak. Teknol. Inf.*, Vol. 01, No. September, P. 01, 2022, [Online]. Available: Https://Sipora.Polije.Ac.Id/16786/%0ahttps://Sipora.Polije.Ac.Id/16786/4/Daftar Pustaka.Pdf