

Analisis Prediktif Bitcoin dengan Metode SVM serta Pembobotan TIF-IDF Berbasis Data Narrative

Danendra Darmawansyah^{a1}, I Gusti Agung Gede Arya Kadyanan^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹danendracool08@gmail.com
²gungde@unud.ac.id

Abstract

The cryptocurrency market has experienced significant volatility in recent years, making it challenging for investors to make informed decisions. This study aims to develop a predictive model for cryptocurrency price increases using TF-IDF (Term Frequency-Inverse Document Frequency) and SVM (Support Vector Machine) based on narrative data. Narrative data, such as news articles and social media posts, can provide valuable insights into investor sentiment and market trends. The proposed model extracts relevant features from narrative data using TF-IDF and employs SVM to classify cryptocurrency price movements into positive, negative, or neutral categories. Experimental results demonstrate the effectiveness of the proposed model in predicting cryptocurrency price increases, with an accuracy of over 70%. The findings suggest that narrative data can be a valuable source of information for cryptocurrency price prediction and that TF-IDF and SVM are effective methods for analyzing narrative data.

Keywords: Cryptocurrency, Price Prediction, TF-IDF, SVM, Narrative Data

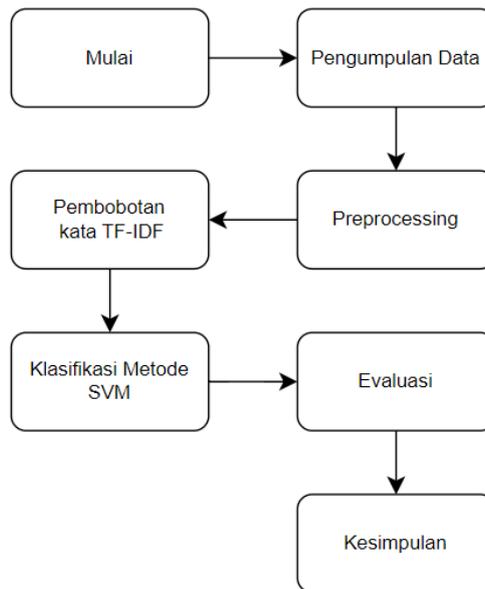
1. Pendahuluan

Fiat merupakan mata uang modern yang saat ini digunakan oleh seluruh dunia, mata uang ini dikeluarkan secara resmi oleh setiap pemerintah dari masing-masing negara. Mata uang fiat berfungsi sebagai alat tukar resmi yang menangani peran perekonomian suatu negara dari unit moneterinya seperti menyimpan nilai, menyediakan akun numerik, dan memfasilitasi pertukaran. Namun, mata uang fiat tidak terkait dengan Cadangan fisik seperti emas dan perak, sehingga mata uang tersebut berisiko kehilangan nilai akibat inflasi atau menjadi tidak ada harganya jika terjadi hiperinflasi[1]. Maka dari itu, Bitcoin hadir sebagai solusi mata uang modern yang memiliki fungsi digital seperti mata uang fiat namun juga dapat sebagai perlindungan terhadap inflasi yang bisa terjadi pada mata uang fiat. Hal ini terjadi karena token kripto memiliki jumlah yang terbatas, berbeda dengan uang fiat yang pasokannya tidak terbatas. Dunia crypto mempunyai salah satu keunikan yaitu sistemnya menganut desentralisasi yang memungkinkan setiap server saling terkoneksi dan mempunyai peranan yang sama. Crypto adalah mata uang virtual yang keamanannya dijamin dengan kriptografi. Kriptografi membuat uang kripto tidak mungkin dipalsukan atau dibelanjakan secara ganda. Jadi, meskipun digunakan secara virtual, tidak mungkin ada pemalsuan yang merugikan penggunaannya. Cryptocurrency didukung oleh teknologi bernama blockchain. Teknologi inilah yang menjamin keamanan transaksi secara online meskipun tanpa menggunakan campur tangan pihak ketiga. Uang kripto dilindungi berbagai algoritma dan enkripsi dan kriptografi yang mengacu pada teknologi blockchain [2]. Pasar mata uang kripto telah mengalami volatilitas yang signifikan dalam beberapa tahun terakhir, sehingga sulit bagi investor untuk membuat keputusan yang tepat[3]. Hal ini disebabkan oleh berbagai faktor, seperti spekulasi, regulasi, dan sentimen investor. Sentimen investor, khususnya, memainkan peran penting dalam menentukan pergerakan harga mata uang kripto. Sentimen positif dapat mendorong harga naik, sedangkan sentimen negatif dapat menyebabkan harga turun[4].

Penelitian ini bertujuan untuk mengembangkan model prediksi kenaikan harga mata uang kripto menggunakan metode SVM (Support Vector Machine) dengan pembobotan TF-IDF (Term Frequency-Inverse Document Frequency) berdasarkan data naratif. Data naratif, seperti artikel berita dan posting media sosial, dapat memberikan wawasan berharga tentang sentimen investor dan tren pasar. Model yang diusulkan mengekstrak fitur yang relevan dari data naratif menggunakan TF-IDF dan menggunakan SVM untuk mengklasifikasikan pergerakan harga mata uang kripto ke dalam kategori positif atau negatif.

2. Metode Penelitian

Penelitian akan melalui beberapa tahapan yaitu mulai dari pengumpulan data, preprocessing, pembobotan kata TF-IDF, klasifikasi metode SVM, evaluasi, dan langkah terakhir adalah kesimpulan dari penelitian ini. Untuk gambaran dari tahapan penelitian dapat dilihat pada gambar 1 dibawah.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Proses pengambilan data melibatkan pengumpulan data pendukung, yang diambil selama proses tersebut. Beberapa cara yang dilakukan adalah melalui studi literatur dan mempelajari beberapa referensi yang berkaitan dengan penelitian. Dataset diambil dengan proses crawling melalui Google Colab dengan menggunakan Twitter Harvest. Dataset dalam bentuk format file csv. Untuk penjelasan lebih lanjut dapat dilihat pada table 1 dibawah.

conversation_id_str	created_at	favorite_count	full_text	id_str	image_url
1780509508629897597	Wed Apr 17 08:12:01 +0000 2024	0	\$BTC LONG market order SL 59309 TP1(10%) 69523...	1780509508629897597	NaN
1780508646490980489	Wed Apr 17 08:08:36 +0000 2024	0	Curiga banget Mei 2024 bakal berdarah-darah ma...	1780508646490980489	https://pbs.twimg.com/tweet_video_thumb/GLWjKq...
1780246978489917951	Wed Apr 17 08:03:03 +0000 2024	0	@yunepo kalo kejadian kyy blackswan tetap pal...	1780507249624142240	NaN
1780461881150304721	Wed Apr 17 06:58:47 +0000 2024	2	@EverythingAjay Ajay bhai yhe khali Crypto Mai...	1780491076660523091	NaN
1780465144666915147	Wed Apr 17 05:15:44 +0000 2024	0	Di bawah lampu kereta bawah tanah yang beriram...	1780465144666915147	https://pbs.twimg.com/media/GLV7m_tacAA_I9t.jpg

Gambar 2. Dataset yang Digunakan

Data yang digunakan diatas berjumlah 1001 baris dan 20 kolom data, didalamnya terdapat keyword untuk memperoleh hasil yang dibutuhkan pada penelitian ini. Dataset yang digunakan adalah data yang diperoleh pada tahun 2023.

2.2. Metode

2.2.1. Analisis Sentimen

Analisis sentimen bertujuan untuk mengetahui, mengekstrak, dan menjalankan informasi tekstual secara otomatis untuk menemui keterangan sentimental dalam mengungkapkan pemikiran. Sentiment analysis digunakan untuk melihat pendapat atau kesamaan opini terhadap sebuah persoalan[5].

2.2.2. Text Mining

Text mining merupakan proses ekstraksi berita dari data asal yang belum terstruktur. Data yang belum terstruktur akan diolah memakai Teknik serta metode tertentu membuat berita yang bermanfaat untuk pengguna. Text mining adalah Teknik yang dipergunakan untuk menangani persoalan classification, clustering, information extraction, dan information retrieval [6].

2.2.3. Preprocessing

Preprocessing adalah proses untuk menormalkan istilah. Hal ini dilakukan agar data yang diterima dengan baik. Mengumpulkan data opini dari media sosial Twitter tidak boleh identik dengan kata baku, kata yang tidak ada dalam kamus, atau bahasa daerah yang digunakan atau dihilangkan. Untuk mengembalikan sejumlah teks ke teks alami dengan mengeliminasi ekspresi atipikal agar dapat meminimalkan noise pada tahap selanjutnya, diperlukan pre-processing atau normalisasi untuk mengatasi hal ini[7].

2.2.4. Term Frequency-Inverse Document Frequency

Salah satu proses pembobotan setiap kata agar bisa mengoptimalkan kemampuan analisis sentiment pada proses text mining. Term Frequency (tf(w,d)) dianggap memiliki proporsi kepentingan sesuai total kemunculannya dalam teks atau dokumen. Inverse Document Frequency (IDF) merupakan metode pembobotan token yang berfungsi untuk memonitor kemunculan token dalam himpunan teks. Berikut merupakan rumus dalam mencari TF-IDF:

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

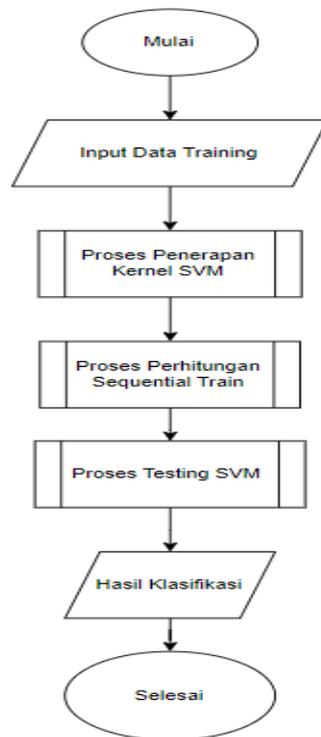
$$TF = \frac{\text{Number of times the term appears in the document}}{\text{Total number of terms in the document}}$$

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|} \quad (2)$$

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (3)$$

2.2.5. Support Vector Machine

SVM adalah Teknik untuk memprediksi klasifikasi dan regresi. SVM merupakan algoritma pembelajaran mesin yang bekerja berdasarkan prinsip Structural Risk Minimization (SRM) dengan tujuan untuk menemukan hyperlane optimal yang memisahkan dua kelas dalam ruang input. Untuk alur penggunaan algoritma SVM dapat dilihat pada gambar 3 dibawah ini.



Gambar 3. Alur Algoritma SVM

2.2.6. Evaluasi

Evaluasi bertujuan untuk menilai hasil uji coba system yang dibuat apakah telah sesuai dengan hasil yang diinginkan. Pemilihan model dan pencarian parameter terbaik menggunakan Grid Search, yang bertujuan untuk mengoptimalkan kinerja model SVM (Support Vector Machine) dengan TF-IDF sebagai fitur representasi. Proses ini terdiri dari beberapa langkah yaitu mendefinisikan parameter grid, melakukan pencarian parameter terbaik dengan k-fold cross validation, serta melatih dan mengevaluasi model dengan parameter terbaik yang ditemukan.

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

$$\text{Precision (P)} = \frac{TP}{(TP+FP)} \quad (5)$$

$$\text{Recall (R)} = \frac{TP}{(TP+FN)} \tag{6}$$

$$F - 1 \text{ Score} = 2 \times \frac{(R \times P)}{(R+P)} \tag{7}$$

Keterangan:

- TP : True Positives
- TN : True Negatives
- FN : False Negatives
- FP : False Positives

3. Hasil dan Pembahasan

Penelitian ini menggunakan metode SVM dengan pembobotan TF-IDF untuk menganalisa tren harga pada token Bitcoin berbasis data naratif yang diambil melalui proses crawling menggunakan Twitter Harvest. Dataset yang digunakan dibagi menjadi 80% data training dan 20% data testing. Dataset yang diperoleh telah melewati preprocessing agar data dapat dibaca sebagai data positif atau negatif. Setelah itu dilakukan K-fold cross validation untuk menilai performa model dari data yang diberikan.

3.1. Preprocessing

3.1.1. Labeling

Data yang telah diperoleh melalui crawling akan dilanjutkan dengan proses labeling yang bertujuan untuk dapat menentukan sentimen terhadap token Bitcoin.

	in_reply_to_screen_name	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
^ETQIW.jpg	NaN	in	Indonesia	0	0	0	https://x.com/juragantrading/status/1791080173779165622	1481611259472936962	juragantrading
	blockvestglobal	in	Jakarta	0	2	0	https://x.com/blockvestglobal/status/1792060832303776177	1114710404012228608	blockvestglobal
	NaN	in	Indonesia	2	53	26	https://x.com/anggaandinata/status/1792727449677246594	134843989	anggaandinata
sQAAbQwA.jpg	NaN	in	Indonesia	0	2	4	https://x.com/PintuID/status/1789924036836835797	763282919817289728	PintuID
	hidatulfikri	in	Tangerang, Indonesia	0	1	0	https://x.com/hidatulfikri/status/1790372432961851725	625625730	hidatulfikri

Gambar 4. Sebelum Labeling

tweet_url	user_id_str	username	cleaning	hapusEmoji	replaceTOM	caseFolding	label	sentiment	score
/status/17900414611...	1484413687708676102	cryptodailyid	BREAKING naik diatas \$	BREAKING naik diatas \$	BREAKING naik diatas \$	BREAKING naik diatas \$	POSITIVE	{'label': 'POSITIVE', 'score': 0.994749128818...}	0.994749
/id/status/1790994337...	1702269721767190528	coinedition_id	mempertahankan perdagangan sideways dengan ant...	mempertahankan perdagangan sideways dengan ant...	mempertahankan perdagangan sideways dengan ant...	mempertahankan perdagangan sideways dengan ant...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.997230827808...}	0.997231
/status/17917046919...	1107742918952083462	bitcoin_brown	iamyessouareno Muh Wakanda	iamyessouareno Muh Wakanda	iamyessouareno Muh Wakanda	iamyessouareno Muh Wakanda	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.976173818111...}	0.976174
1792212896468119842	439326281	ukikahir	Bitcoin adalah penipuan yang pada akhirnya aka...	Bitcoin adalah penipuan yang pada akhirnya aka...	Bitcoin adalah penipuan yang pada akhirnya aka...	Bitcoin adalah penipuan yang pada akhirnya aka...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.999657988548...}	0.999658
/status/17914657650...	1133128942209028096	lbrahim632931	tonpuppets UQBYt e Z UwaGYhffuUV oFNGikGUQ vb...	tonpuppets UQBYt e Z UwaGYhffuUV oFNGikGUQ vb...	tonpuppets UQBYt e Z UwaGYhffuUV oFNGikGUQ vbri...	tonpuppets UQBYt e Z UwaGYhffuUV oFNGikGUQ vbri...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.995218217372...}	0.995218
/status/1792699716048...	238091066	InvestorID	Harga Bitcoin Meroket ke Level US\$ Ribu ETH ...	Harga Bitcoin Meroket ke Level US\$ Ribu ETH ...	Harga Bitcoin Meroket ke Level US\$ Ribu ETH ...	Harga Bitcoin Meroket ke Level US\$ Ribu ETH ...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.993190705776...}	0.993191
036/status/179119805...	1390989451099197441	Bustomi92248936	Senator Cynthia Lummis menyerukan persetujuan ...	Senator Cynthia Lummis menyerukan persetujuan ...	Senator Cynthia Lummis menyerukan persetujuan ...	Senator Cynthia Lummis menyerukan persetujuan ...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.881730556488...}	0.881731
/status/179282947928...	1495570216847241217	majalabs_xyz	Bitcoin mencapai US\$ pada Mei ...	Bitcoin mencapai US\$ pada Mei ...	Bitcoin mencapai US\$ pada Mei naik % dalam ...	Bitcoin mencapai US\$ pada Mei naik % dalam ...	NEGATIVE	{'label': 'NEGATIVE', 'score': ...}	0.995741

Gambar 5. Sesudah Labeling

Proses ini akan mengeluarkan hasil sentiment dari dataset yang sudah diperoleh. Proses labeling menggunakan Transformer Python untuk mendapatkan hasil sentiment positif atau negatif.

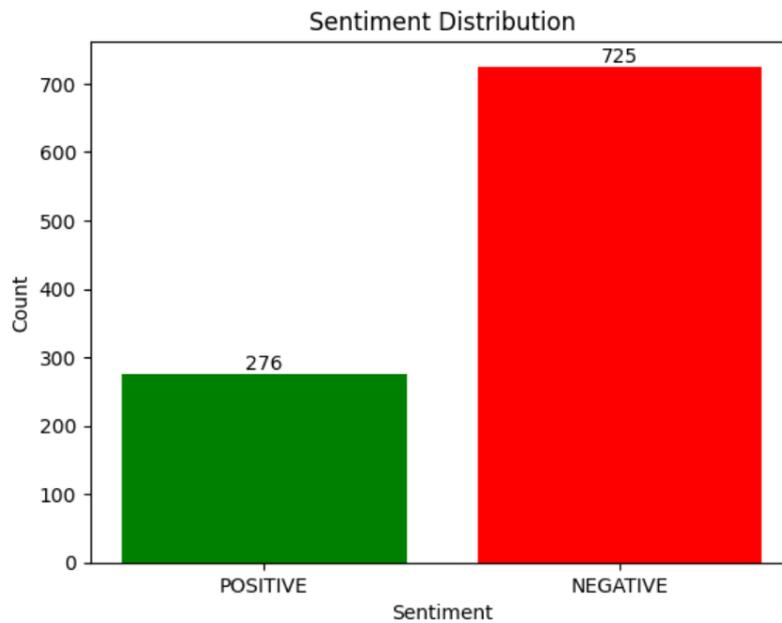
3.1.2. Cleansing pipeline

Pada proses ini, dataset yang digunakan akan melalui proses pembersihan kata agar tersisa kata penting yang dapat dianalisa sentimennya. Bagian ini meliputi cleaning karakter, emoji, Stopword, caseFolding, Tokenizing, formalisasi, dan stemming.

cleaning	hapusEmoji	replaceTOM	caseFolding	label	sentiment	score	tokenizing	formalisasi
Dulu Bitcoin masih rb rupiah orang masih ga m...	Dulu Bitcoin masih rb rupiah orang masih ga m...	Dulu Bitcoin masih rb rupiah orang masih ga m...	dulu bitcoin masih rb rupiah orang masih ga m...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.998764753341...}	0.998765	[dulu, bitcoin, masih, rb, rupiah, orang, masi...	[dulu, bitcoin, masih, rb, rupiah, orang, masi...
Market Overview Bitcoin hari ini diperdagangka...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.989911854267...}	0.989912	[market, overview, bitcoin, hari, ini, diperda...	[market, overview, bitcoin, hari, ini, diperda...			
klo negara eropa mata uangnya ganti euro buat...	klo negara eropa mata uangnya ganti euro buat...	klo negara eropa mata uangnya ganti euro buat...	klo negara eropa mata uangnya ganti euro buat...	NEGATIVE	{'label': 'NEGATIVE', 'score': 0.979081392288...}	0.979081	[klo, negara, eropa, mata, uangnya, ganti, eur...	[klo, negara, eropa, mata, uangnya, ganti, eur...
ylad bRey dari namanya aja udh beda bitcoin...	ylad bRey dari namanya aja udh beda bitcoin...	ylad bRey dari namanya aja udh beda bitcoin d...	ylad brey dari namanya aja udh beda bitcoin d...	POSITIVE	{'label': 'POSITIVE', 'score': 0.966544508934...}	0.966545	[ylad, brey, dari, namanya, aja, udh, beda, bi...	[ylad, brey, dari, namanya, aja, udah, beda, b...

Gambar 6. Hasil Cleansing Pipeline

Gambar diatas merupakan hasil dari proses cleansing pipeline pada dataset. Diperoleh hasil teks setelah cleaning, clear emoji, replaceTOM, caseFolding, tokenizing, dan formalisasi.



Gambar 8. Grafik Sentimen

Gambar diatas merupakan hasil visualisasi data dalam bentuk grafik yang berfokus pada sentimen untuk dibandingkan agar dapat dianalisa dengan baik terkait tren token Bitcoin. Terlihat bahwa tren negatif lebih banyak daripada tren positif yang berarti tren token Bitcoin belum menunjukkan hasil positif pada investor crypto pada tahun 2023 sesuai dengan dataset yang diambil.

3.4. Evaluasi dan Perbandingan Data dengan Market

3.4.1. Hasil Evaluasi

Berikut merupakan hasil evaluasi yang menghasilkan nilai akhir seperti precision, recall, f1-score, dan akurasi k-fold cross validation. Pada penelitian ini hasil yang didapatkan dari SVM dengan TF-IDF dapat dilihat pada gambar 8 dibawah ini.

```
➔ Parameter terbaik: {'C': 1, 'gamma': 'scale', 'kernel': 'linear'}  
Akurasi terbaik: 0.7762499999999999  
  
Classification Report for TF-IDF with Best Parameters:  
              precision    recall  f1-score   support  
  
  NEGATIVE      0.77      0.92      0.83        143  
  POSITIVE      0.60      0.31      0.41         58  
  
 accuracy              0.74        201  
 macro avg              0.68      0.61      0.62        201  
 weighted avg          0.72      0.74      0.71        201
```

Gambar 9. Hasil SVM dengan TF-IDF

Dari hasil SVM dengan TF-IDF sehingga mendapatkan evaluasi berupa nilai C terbaik ada pada nilai 1, lalu gamma yang digunakan yaitu pada scale, dan kernel yang digunakan yaitu linear. Sehingga dari testing yang dilakukan didapatkan nilai precision 77%, recall 92%, dan f1-score

83% pada sentiment negatif. Sedangkan pada sentiment positif menghasilkan nilai precision 60%, recall 31%, dan f1-score 41%. Sehingga hasil akurasi akhir yang didapatkan yaitu sebesar 74%.

3.4.2. Perbandingan Data dengan Market

Setelah mendapatkan hasil akhir dari analisa yang dilakukan dengan menggunakan metode SVM serta pembobotan TF-IDF diperoleh hasil negatif yang melambung tinggi jika dibandingkan dengan hasil positif. Hal ini berbanding lurus dengan market crypto Bitcoin yang terjadi pada tahun 2023 sesuai dengan data yang diperoleh. Bitcoin mengalami penurunan yang signifikan terjadi pada tahun 2023 hingga mengalami 50% penurunan harga jual. Data market crypto Bitcoin dapat dilihat pada gambar 9 dibawah ini.

BTC/IDR 1.097.018.000 -28.690.032 (-2,55%)						
Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%
01/12/2023	663.249.984	585.713.024	693.315.008	585.299.968	1,37K	+13.24%
01/11/2023	585.710.976	550.425.984	596.374.016	540.000.000	1,07K	+6.41%
01/10/2023	550.425.984	418.384.000	555.721.024	418.311.008	0,99K	+31.56%
01/09/2023	418.384.000	398.100.992	421.753.984	386.000.000	0,51K	+5.12%
01/08/2023	398.000.000	440.576.992	455.489.984	392.000.000	0,65K	-9.66%
01/07/2023	440.576.992	459.832.000	470.553.984	436.000.000	0,52K	-4.19%
01/06/2023	459.832.000	407.459.008	467.984.000	372.000.000	0,79K	+12.85%
01/05/2023	407.459.008	431.268.992	434.716.992	385.000.000	0,64K	-5.52%
01/04/2023	431.268.992	427.999.008	454.000.000	408.000.000	0,78K	+0.76%
01/03/2023	427.999.008	354.999.008	440.644.000	305.716.992	1,32K	+20.56%
01/02/2023	354.999.008	347.192.000	381.500.000	326.603.008	1,01K	+2.25%

Gambar 10. Harga Jual Bitcoin Tahun 2023

Sesuai dengan gambar data diatas bahwa pada tahun 2023 menunjukkan harga jual Bitcoin sedang dalam kondisi Bearish Market. Bearish Market merupakan kondisi pasar melemah diakibatkan banyak investor yang menjual aset Bitcoin mereka sehingga harga jual mengalami penurunan.

4. Kesimpulan

Berdasarkan hasil dari penelitian yang telah dilaksanakan, metode SVM dengan pembobotan TF-IDF merupakan metode yang baik untuk digunakan dalam analisis prediktif tren market Bitcoin berbasis data naratif. Hal ini dibuktikan dengan hasil perbandingan data antara hasil pengujian dengan data real yang diperoleh. Data naratif merupakan salah satu bentuk analisa market crypto yang sering digunakan, sehingga metode ini sesuai dengan analisa data naratif yang dijalankan. Metode ini menghasilkan nilai akurasi 74% yang berarti sudah baik untuk digunakan dalam analisa. Pada penelitian selanjutnya diharapkan dapat meningkatkan akurasi dari data yang diperoleh sehingga mendapatkan analisa yang lebih baik terhadap tren market crypto Bitcoin.

Daftar Pustaka

- [1] J. Chen, "Uang Fiat: Apa Artinya, Cara Kerja, Contoh, Kelebihan & Kekurangan", Investopedia, 28 Februari 2024, [Online]. Available: <https://www.investopedia.com/terms/f/fiatmoney.asp> [Diakses: 11 Mei 2024].

- [2] A. L. Hidayah, "Yuk, Berkenalan dengan Kripto!", Kementerian Keuangan Republik Indonesia, 11 April 2023, [Online]. Available: <https://www.djkn.kemenkeu.go.id/kpknl-tangerang1/baca-artikel/16059/Yuk-Berkenalan-dengan-Kripto.html> [Diakses: 11 Mei 2024].
- [3] M. M. Karim, M. H. Ali, L. Yarovaya, M. H. Uddin, S. Hammoudeh, "Return-volatility relationships in cryptocurrency markets: Evidence from asymmetric quantiles and non-linear ARDL approach", *International Review of Financial Analysis*, vol. 90, November 2023.
- [4] E. Akyildirim, A. F. Aysan, O. Cepni, O. Serbest, "Sentiment matters: the effect of news-media on spillovers among cryptocurrency returns", *The European Journal of Finance*, Juli 2022.
- [5] M. W. Berry and J. Kogan, "Text Mining Application and Theory", Wiley; 1st Edition, 2010
- [6] R. Feldman and J. Sanger, "The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data", Cambridge University Press; 1st edition, 2006.
- [7] F. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia", Institute for Logic, Language and Computation: Universiteit van Amsterdam, 2003.