

BAB III

HASIL ANALISIS DAN PEMBAHASAN

3.1 Dataset

Dalam penelitian ini hasil *crawling* data terdiri dari 15 kolom yang mencakup *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *id_str*, *image_url*, *in_reply_to_screen_name*, *lang*, *location*, *quote_count*, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str*, dan *username*, data yang didapatkan berjumlah 558 *tweet* yang terkait dengan peristiwa bitcoin halving yang dikumpulkan dari *Twitter*. Setelah itu data *tweet* tersebut telah melalui proses preprocessing hasil akhir data yang akan berjumlah 538, data disimpan dalam format CSV. Gambar 3.1 menunjukkan hasil pengambilan data menggunakan alat pengumpul *tweet* (*crawling*). Data pengambilan tersebut dapat dilihat pada gambar di bawah ini.

#	A	B	C	D	F	G	H	I	J	K	L	M	N	O	P	Q
1	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username	
2	1.79E+18	Thu May 02 04:15:4	0	Penambang Bitcoi	1.79E+18			in		0	0	0	https://hwit	1.72E+18	CointelegraphMY	
3	1.79E+18	Thu May 02 04:12:2	0	Kabar buruk buat	1.79E+18	https://pbs.twimg.com/media/GMI83QE		in		0	1	0	https://hwit	1.77E+18	kbursacrypto	
4	1.79E+18	Thu May 02 04:02:1	0	@TLD_Survive kali	1.79E+18	TLD_Survive		in	Sydney, New S	0	0	0	https://hwit	1.53E+18	kindo0111	
5	1.79E+18	Thu May 02 04:02:2	0	@rdmncryptoguy @	1.79E+18	rdmncryptoguy		in	Indonesia	0	0	0	https://hwit	9.42E+17	aadbitcoin	
6	1.79E+18	Thu May 02 04:02:2	0	@timothyronald2	1.79E+18	https://pbs.t	timothyronald22			0	0	0	https://hwit	1.72E+18	emba67765	
7	1.79E+18	Thu May 02 03:55:1	0	@sipalabotak @Fr	1.79E+18	sipalabotak		in	Earth	0	0	0	https://hwit	1.43E+18	ismalcbtc	
8	1.79E+18	Thu May 02 03:20:2	23	Willy Woo analis c	1.79E+18	https://pbs.twimg.com/media/GMixCb4		in		0	3	0	https://hwit	1.67E+18	akademicroptoid	
9	1.79E+18	Thu May 02 03:19:1	0	Ini mah beneran d	1.79E+18	https://pbs.twimg.com/media/GMiw0Xu		in	Indonesia	0	2	0	https://hwit	48856009	sinoypermata	
10	1.79E+18	Thu May 02 03:17:1	8	MicroStrategy Cipi	1.79E+18	https://pbs.twimg.com/media/GMivHbH		in	Desolate #328	0	1	3	https://hwit	70912579	bukangananmu	
11	1.79E+18	Thu May 02 03:16:1	0	@andricuarterus i	1.79E+18	andricuarterus		in		0	0	0	https://hwit	1.76E+18	CuanDelan61328	
12	1.79E+18	Thu May 02 03:10:1	2	Lantas apakah fen	1.79E+18	akademicroptoid		in		0	0	0	https://hwit	1.67E+18	akademicroptoid	
13	1.79E+18	Thu May 02 03:10:1	0	Hal ini justru mem	1.79E+18	akademicroptoid		in		0	1	0	https://hwit	1.67E+18	akademicroptoid	
14	1.79E+18	Thu May 02 03:10:1	18	BlackRock perusa	1.79E+18	https://pbs.twimg.com/media/GMiu2Px		in		0	1	0	https://hwit	1.67E+18	akademicroptoid	
15	1.79E+18	Thu May 02 03:08:1	0	@boovoeer @halk	1.79E+18	boovoeer		in	Yogyakarta, Ind	0	0	0	https://hwit	1.12E+18	SultanAmirr	
16	1.79E+18	Thu May 02 03:08:1	26	GM ini alasan saya	1.79E+18	https://pbs.twimg.com/media/GMiu0Pj		in		0	5	0	https://hwit	1.45E+18	andricuarterus	
17	1.79E+18	Thu May 02 03:01:1	0	@timothyronald2	1.79E+18	timothyronald22		in		0	0	0	https://hwit	1.76E+18	sopari0TG	
18	1.79E+18	Thu May 02 02:59:1	0	@meinmohhtar Bt	1.79E+18	meinmohhtar		in	Malaysia	0	0	0	https://hwit	1.77E+18	YvDudie	
19	1.79E+18	Thu May 02 02:58:1	0	bitcoin scam. jgn k	1.79E+18			in		0	0	0	https://hwit	1731723541	Oxruojers	
20	1.79E+18	Thu May 02 02:56:1	0	@SultanAmirr @h	1.79E+18	SultanAmirr		in		0	1	0	https://hwit	1.72E+18	boovoeer	
21	1.79E+18	Thu May 02 02:49:1	0	@anggaandinata #	1.79E+18	anggaandinata		in	Earth	0	0	0	https://hwit	1.43E+18	ismalcbtc	
22	1.79E+18	Thu May 02 02:44:1	0	Gak ngerti lagi am	1.79E+18	https://pbs.twimg.com/media/GMioyG9		in	Absurdistan	0	0	3	https://hwit	99705510	BNGPY	
23	1.79E+18	Thu May 02 02:44:1	0	@AnalisaCrypto K	1.79E+18	AnalisaCrypto		in		0	0	0	https://hwit	1.63E+18	Rizkydibcoin	
24	1.79E+18	Thu May 02 02:41:1	4	2 Mei 2024: Poten	1.79E+18			in	Greater Jakarta	1	3	4	https://hwit	41605719	aswadiqbal	
25	1.79E+18	Thu May 02 02:08:1	1	Makasi banyak bit	1.79E+18	https://pbs.twimg.com/media/GMigtO6		in		0	0	0	https://hwit	6882875	Agusyudiartha	
26	1.79E+18	Thu May 02 02:04:1	2	Ceramah diatas di	1.79E+18	aadbitcoin		in	Indonesia	0	0	0	https://hwit	9.42E+17	aadbitcoin	
27	1.79E+18	Thu May 02 02:04:1	4	Jadi menurut perh	1.79E+18	https://pbs.t	aadbitcoin		Indonesia	0	2	0	https://hwit	9.42E+17	aadbitcoin	
28	1.79E+18	Thu May 02 02:04:1	35	Bitcoin adalah unt	1.79E+18	https://pbs.twimg.com/ext_tw_video_th		in	Indonesia	0	1	9	https://hwit	9.42E+17	aadbitcoin	
29	1.79E+18	Thu May 02 01:57:1	0	#Bitcoin update C	1.79E+18	https://pbs.twimg.com/media/GMic33b		in	Indonesia	0	0	0	https://hwit	1.01E+18	CryptoTalk_007	
30	1.79E+18	Thu May 02 01:57:1	0	B U Y T H E D I P #B	1.79E+18			in	Sri Lanka	0	0	0	https://hwit	1.65E+18	tharusha908	
31	1.79E+18	Thu May 02 01:46:1	0	BITC 5250k MINIM	1.79E+18			in	conchain	0	0	0	https://hwit	1.74E+18	Twelecegrass	
32	1.79E+18	Thu May 02 01:34:1	1	BITCOIN BATUK DI	1.79E+18	https://pbs.twimg.com/media/GMIv5Bw		in	City Hunter	0	2	0	https://hwit	1.73E+18	JeckyJeck399826	
33	1.79E+18	Thu May 02 01:32:1	0	@ajalb_investasi	1.79E+18	ajalb_investasi		in		1	1	0	https://hwit	1.51E+18	satriaharjati	
34	1.79E+18	Thu May 02 01:25:1	0	APABILA HANYA B	1.79E+18	https://pbs.twimg.com/media/GMIWpM		in		0	0	0	https://hwit	1.38E+18	AryaArynt	
35	1.79E+18	Thu May 02 01:22:1	1	Bitcoin Ekosistem	1.79E+18			in	Blockchain	0	1	0	https://hwit	3255392544	mwisbc	
36	1.78E+18	Thu May 02 01:19:1	0	@Kalimasada97 Bt	1.79E+18	Kalimasada97		in		0	0	0	https://hwit	1.72E+18	FacebookGa74666	
37	1.79E+18	Thu May 02 01:18:1	0	BlackRock menua	1.79E+18			in		0	0	0	https://hwit	1.66E+18	PutuCryptos	

Gambar 3. 1 Hasil Crawling

Untuk mempermudah analisis data pada tahap selanjutnya, hanya atribut *full_text* yang digunakan. Berikut adalah Tabel 3.1 yang menampilkan atribut dataset yang digunakan dalam penelitian ini.

Tabel 3. 1 Dataset

<i>No</i>	<i>Full_text</i>
1	Penambang Bitcoin @RiotPlatforms melaporkan suku rekod dengan pendapatan bersih sebanyak \$211.8 juta tetapi tidak mencapai anggaran pendapatan analis. https://t.co/VW3jSxLusT
2	Kabar buruk buat para pecinta kripto! Harga Bitcoin merosot karena prospek suku bunga yang tetap tinggi. Koreksi bulanan terdalam sejak kolapsnya FTX membuat aset digital turun sekitar 16 persen di bulan April 2024. https://t.co/eTrWAZI9JH
3	@TLD_Survive kalau saya dikasih 100juta dalam kondisi ini uang nganggur ya. saya all in beli bitcoin saat sentiment lagi fear mungkin beberapa bulan kedepan saat harga sedang dibottomnya tunggu 2025 atau 2027 and earn money from that
4	@rdmcriptoguy @crypto_radiz @yanzero_ Berarti sekilas aja smart contract bug risk ama liquidity risk aja palingan.. Risk pasti ga 0% tp mitigasi resiko mestinya bisa.. Udh pas lah bro @yanzero_ lah yg lbh paham masalah defi ini dbanding aye..
5	@timothyronald22 GN akibat ikutin borong bitcoin konsisten https://t.co/j4pAyYrFZr
.....
557	Sebagian orang menanti dan merayakan Bitcoin Halving seperti turunnya Nabi Isa. *random rant udah lama nyimpen tweet ini
558	Kalau tahun ini bakalan keluar ntar bulan Januari klw habis itu uang disetor ke rdo tunggu bitcoin hancur luluh lantah dan bearish lalu beli bitcoin tak lupa sebagian uangnya modal untuk buka usaha dan buat pengembangan kompetensi diri sendiri

3.2 Labeling

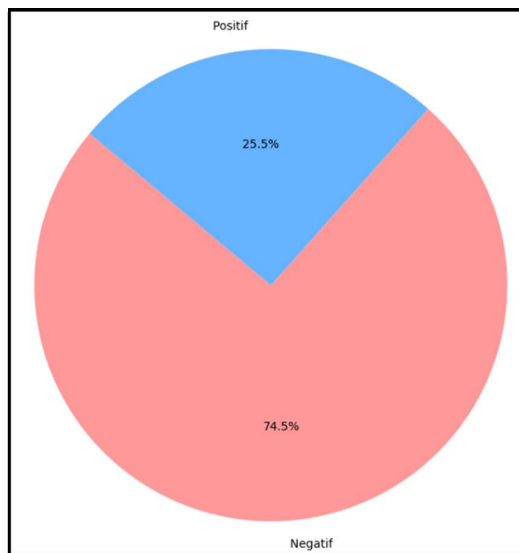
Dataset kemudian akan melalui tahap pemberian label, pelabelan sebanyak 538 data dilakukan secara manual oleh *expert* di bidang bahasa dengan mengkategorikan tweet menjadi dua kategori yaitu positif dan negatif berdasarkan sentimen. Tabel 3.2 merupakan hasil text yang telah diberikan label sebagai berikut :

Tabel 3. 2 Labeling Dataset

<i>No</i>	<i>Full_text</i>	<i>Sentimen</i>
1	Penambang Bitcoin @RiotPlatforms melaporkan suku rekod dengan pendapatan bersih sebanyak \$211.8 juta tetapi tidak mencapai anggaran pendapatan analis. https://t.co/VW3jSxLusT	Negatif
2	Kabar buruk buat para pecinta kripto! Harga Bitcoin merosot karena prospek suku bunga yang tetap tinggi. Koreksi bulanan terdalam sejak kolapsnya FTX membuat aset digital turun sekitar 16 persen di bulan April 2024. https://t.co/eTrWAZI9JH	Negatif

<i>No</i>	<i>Full_text</i>	<i>Sentimen</i>
3	@TLD_Survive kalau saya dikasih 100juta dalam kondisi ini uang nganggur ya. saya all in beli bitcoin saat sentiment lagi fear mungkin beberapa bulan kedepan saat harga sedang dibottomnya tunggu 2025 atau 2027 and earn money from that	Positif
4	@timothyronald22 GN akibat ikutin borong bitcoin konsisten https://t.co/j4pAyYrFZr	Positif
5	bitcoin scam. jgn beli. ponzi	Negatif
.....
557	Sebagian orang menanti dan merayakan Bitcoin Halving seperti turunnya Nabi Isa. *random rant udah lama nyimpen tweet ini	Negatif
558	Kalau tahun ini bakalan keluar ntar bulan Januari klw habis itu uang disetor ke rdo tunggu bitcoin hancur luluh lantah dan bearish lalu beli bitcoin tak lupa sebagian uangnya modal untuk buka usaha dan buat pengembangan kompetensi diri sendiri	Negatif

Dari total dataset yang telah diberikan label, terdapat 25.5% *tweet* yang memiliki sentimen positif, sedangkan 74.5% *tweet* lainnya memiliki sentimen negatif. Visualisasi distribusi data sentimen dapat dilihat pada gambar 3.2.



Gambar 3. 2 Visualisasi Persentase Sentimen

Setelah dataset terlabeli, maka proses selanjutnya adalah preprocessing, dimana proses preprocessing tersebut terbagi menjadi beberapa tahap, yaitu case folding, cleansing, tokenizing, stopword removal, stemming dan delete duplicates.

3.3 Pre-processing

Proses pembersihan atau *preprocessing* bertujuan untuk membersihkan data. Pada tahapan ini, data hasil *crawling* akan diolah dengan memilih data yang relevan untuk digunakan dan menghapus data yang tidak berguna, sehingga data menjadi lebih terstruktur dan siap untuk diproses lebih lanjut. Tahap *preprocessing* terdiri dari *case folding*, *cleansing*, *tokenizing*, *stopword removal*, *stemming* dan *delete duplicates* untuk menghapus dataset yang muncul dua kali atau lebih. Berikut hasil teks setelah *preprocessing* sebagai berikut:

a. Case folding

berdasarkan pada tabel 3.3, semua huruf kapital dalam data tweet telah diubah menjadi huruf kecil. Perubahan ini dilakukan untuk mempermudah proses pembacaan oleh mesin terhadap korpus dan mengurangi waktu yang dibutuhkan.

Tabel 3. 3 Hasil case folding

No	Case Folding
1	penambang bitcoin @riotplatforms melaporkan suku rekod dengan pendapatan bersih sebanyak \$211.8 juta tetapi tidak mencapai anggaran pendapatan analis. https://t.co/vw3jsxlust
2	kabar buruk buat para pecinta kripto! harga bitcoin merosot karena prospek suku bunga yang tetap tinggi. koreksi bulanan terdalam sejak kolapsnya ftx membuat aset digital turun sekitar 16 persen di bulan april 2024. https://t.co/etrwazl9jh
3	@tld_survive kalau saya dikasih 100juta dalam kondisi ini uang nganggur ya. saya all in beli bitcoin saat sentiment lagi fear mungkin beberapa bulan kedepan saat harga sedang dibottomnya tunggu 2025 atau 2027 and earn money from that
4	@rdmcriptoguy @crypto_radiz @yanzero_ berarti sekilas aja smart contract bug risk ama liquidity risk aja palingan.. risk pasti ga 0% tp mitigasi resiko mestinya bisa.. udh pas lah bro @yanzero_ lah yg lbh paham masalah defi ini dbanding aye..
5	@timothyronald22 gn akibat ikutin borong bitcoin konsisten https://t.co/j4payyrfzr
.....
557	sebagian orang menanti dan merayakan bitcoin halving seperti turunnya nabi isa. *random rant udah lama nyimpen tweet ini
558	kalau tahun ini bakalan keluar ntar bulan januari klw habis itu uang disetor ke rdo tunggu bitcoin hancur luluh lantah dan bearish lalu beli bitcoin tak lupa sebagian uangnya modal untuk buka usaha dan buat pengembangan kompetensi diri sendiri

b. Cleansing

Cleansing telah dilakukan pada Tabel 3.4, Cleansing merupakan penghapusan karakter yang tidak diperlukan dari teks. Langkah ini dilakukan untuk membersihkan teks dari karakter-karakter yang tidak relevan dan dapat mengganggu proses analisis.

Tabel 3. 4 Hasil Cleansing

<i>No</i>	<i>Cleansing</i>
1	penambang bitcoin melaporkan suku rekod dengan pendapatan bersih sebanyak juta tetapi tidak mencapai anggaran pendapatan analis
2	kabar buruk buat para pecinta kripto harga bitcoin merosot karena prospek suku bunga yang tetap tinggi koreksi bulanan terdalam sejak kolapsnya ftx membuat aset digital turun sekitar persen di bulan april
3	kalau saya dikasih juta dalam kondisi ini uang nganggur ya saya all in beli bitcoin saat sentiment lagi fear mungkin beberapa bulan kedepan saat harga sedang dibottomnya tunggu atau and earn money from that
4	berarti sekilas aja smart contract bug risk ama liquidity risk aja palingan risk pasti ga tp mitigasi resiko mestinya bisa udh pas lah bro lah yg lbh paham masalah defi ini dbanding aye
5	gn akibat ikutin borong bitcoin konsisten
.....
557	sebagian orang menanti dan merayakan bitcoin halving seperti turunnya nabi isa random rant udah lama nyimpen tweet ini
558	kalau tahun ini bakalan keluar ntar bulan januari klw habis itu uang disetor ke rdo tunggu bitcoin hancur luluh lantah dan bearish lalu beli bitcoin tak lupa sebagian uangnya modal untuk buka usaha dan buat pengembangan kompetensi diri sendiri

c. *Tokenizing*

Tokenizing telah dilakukan seperti pada Tabel 3.5 Data tweet yang awalnya berbentuk kalimat telah dipecah menjadi kata-kata individu. Proses *tokenizing* dilakukan untuk memudahkan tahap transformasi sehingga pemrosesan dilakukan berdasarkan kata demi kata, bukan berdasarkan kalimat.

Tabel 3. 5 Hasil Tokenizing

<i>No</i>	<i>Tokenizing</i>
1	['penambang', 'bitcoin', 'melaporkan', 'suku', 'rekod', 'dengan', 'pendapatan', 'bersih', 'sebanyak', 'juta', 'tetapi', 'tidak', 'mencapai', 'anggaran', 'pendapatan', 'analisis']
2	['kabar', 'buruk', 'buat', 'para', 'pecinta', 'kripto', 'harga', 'bitcoin', 'merosot', 'karena', 'prospek', 'suku', 'bunga', 'yang', 'tetap', 'tinggi', 'koreksi', 'bulanan', 'terdalam', 'sejak', 'kolapsnya', 'ftx', 'membuat', 'aset', 'digital', 'turun', 'sekitar', 'persen', 'di', 'bulan', 'april']
3	['kalau', 'saya', 'dikasih', 'juta', 'dalam', 'kondisi', 'ini', 'uang', 'nganggur', 'ya', 'saya', 'all', 'in', 'beli', 'bitcoin', 'saat', 'sentiment', 'lagi', 'fear', 'mungkin', 'beberapa', 'bulan', 'ke depan', 'saat', 'harga', 'sedang', 'dibottomnya', 'tunggu', 'atau', 'and', 'earn', 'money', 'from', 'that']
4	['berarti', 'sekilas', 'aja', 'smart', 'contract', 'bug', 'risk', 'ama', 'liquidity', 'risk', 'aja', 'palingan', 'risk', 'pasti', 'ga', 'tp', 'mitigasi', 'resiko', 'mestinya', 'bisa', 'udh', 'pas', 'lah', 'bro', 'lah', 'yg', 'lbh', 'paham', 'masalah', 'defi', 'ini', 'dbanding', 'aye']
5	['gn', 'akibat', 'ikutin', 'borong', 'bitcoin', 'konsisten']
.....
557	['sebagian', 'orang', 'menanti', 'dan', 'merayakan', 'bitcoin', 'halving', 'seperti', 'turunnya', 'nabi', 'isa', 'random', 'rant', 'udah', 'lama', 'nyimpen', 'tweet', 'ini']
558	['kalau', 'tahun', 'ini', 'bakalan', 'keluar', 'ntar', 'bulan', 'januari', 'klw', 'habis', 'itu', 'uang', 'disetor', 'ke', 'rdo', 'tunggu', 'bitcoin', 'hancur', 'luluh', 'lantah', 'dan', 'bearish', 'lalu', 'beli', 'bitcoin', 'tak', 'lupa', 'sebagian', 'uangnya', 'modal', 'untuk', 'buka', 'usaha', 'dan', 'buat', 'pengembangan', 'kompetensi', 'diri', 'sendiri']

d. Stopword Removal

Kata-kata sambung seperti "yang", "dan", "di", "dari" telah dihilangkan melalui proses stopword, terlihat pada Tabel 3.6. Kata-kata tersebut tidak muncul lagi. Proses stopword bertujuan untuk menyaring kata-kata yang tidak relevan dalam klasifikasi, seperti kata penghubung "yang", "dan", "di", dan lainnya. Proses ini dilakukan dengan bantuan library Sastrawi yang tersedia dalam bahasa pemrograman Python.

Tabel 3. 6 Hasil Stopword Removal

<i>No</i>	<i>Stopword removal</i>
1	['penambang', 'bitcoin', 'melaporkan', 'suku', 'rekod', 'pendapatan', 'bersih', 'juta', 'mencapai', 'anggaran', 'pendapatan', 'analisis']
2	['kabar', 'buruk', 'pecinta', 'kripto', 'harga', 'bitcoin', 'merosot', 'prospek', 'suku', 'bunga', 'koreksi', 'bulanan', 'terdalam', 'kolapsnya', 'ftx', 'aset', 'digital', 'turun', 'persen', 'april']
3	['dikasih', 'juta', 'kondisi', 'uang', 'nganggur', 'ya', 'all', 'in', 'beli', 'bitcoin', 'sentiment', 'fear', 'ke depan', 'harga', 'dibottomnya', 'tunggu', 'and', 'earn', 'money', 'from', 'that']

<i>No</i>	<i>Stopword_removal</i>
4	['sekilas', 'aja', 'smart', 'contract', 'bug', 'risk', 'ama', 'liquidity', 'risk', 'aja', 'palingan', 'risk', 'ga', 'tp', 'mitigasi', 'resiko', 'mestinya', 'udh', 'pas', 'bro', 'yg', 'lbh', 'paham', 'defi', 'dbanding', 'aye']
5	['gn', 'akibat', 'ikutin', 'borong', 'bitcoin', 'konsisten']
.....
557	['orang', 'merayakan', 'bitcoin', 'halving', 'turunnya', 'nabi', 'isa', 'random', 'rant', 'udah', 'nyimpen', 'tweet']
558	['ntar', 'januari', 'klw', 'habis', 'uang', 'disetor', 'rdo', 'tunggu', 'bitcoin', 'hancur', 'luluh', 'lantah', 'bearish', 'beli', 'bitcoin', 'lupa', 'uangnya', 'modal', 'buka', 'usaha', 'pengembangan', 'kompetensi']

e. *Stemming*

Tahap selanjutnya dalam *preprocessing* adalah *stemming*. Proses ini melibatkan pengubahan kata-kata menjadi bentuk dasarnya dengan menghilangkan imbuhan awalan, sisipan, akhiran. Tujuannya adalah untuk menyeragamkan bentuk kata sehingga variasi kata yang memiliki makna serupa dapat dikenali sebagai entitas yang sama. Berikut adalah tabel 3.7 hasil text setelah *stemming*:

Tabel 3. 7 Hasil Stemming

<i>No</i>	<i>Stemming</i>
1	tambang bitcoin lapor suku rekod dapat bersih juta capai anggar dapat analis
2	kabar buruk cinta kripto harga bitcoin merosot prospek suku bunga koreksi bulan dalam kolaps ftx aset digital turun persen april
3	kasih juta kondisi uang nganggur ya all in beli bitcoin sentiment fear depan harga dibottomnya tunggu and earn money from that
4	kilas aja smart contract bug risk ama liquidity risk aja paling risk ga tp mitigasi resiko mesti udh pas bro yg lbh paham defi dbanding aye
5	gn akibat ikutin borong bitcoin konsisten
.....
557	orang raya bitcoin halving turun nabi isa random rant udah nyimpen tweet
558	ntar januari klw habis uang setor rdo tunggu bitcoin hancur luluh lantah bearish beli bitcoin lupa uang modal buka usaha kembang kompetensi

f. *Delete Duplicates*

Dataset yang sebelumnya berjumlah 558, setelah melewati proses delete duplicate, mengalami pengurangan sebanyak 20 baris. Penghapusan ini dilakukan karena baris-baris

tersebut muncul dua kali atau lebih dalam dataset, yang dapat menyebabkan bias dalam analisis data dan penurunan performa model. Setelah proses penghapusan duplikasi, jumlah akhir dataset menjadi 538 baris, memastikan bahwa setiap entri dalam dataset adalah unik dan representatif. Dengan demikian, dataset yang telah dibersihkan dari duplikasi ini akan memberikan dasar yang lebih kuat dan akurat untuk penelitian lebih lanjut.

```
DATA DUPLIKAT YANG TELAH DI HAPUS:
full_text
62 @DappRadar Setelah halving Bitcoin tetap menja...
203 @andricuanterus Bang kan kata lu bitcoin harus...
209 @andricuanterus Bang kan kata lu bitcoin harus...
220 Halving Bitcoin Ancam Bisnis Penambang Kecil d...
228 @anggaandinata Demand untuk bitcoin semakin me...
234 Cek suara dulu Yang kemarin: - bilang kalau tu...
249 Analis Prediksi Nasib Bitcoin Usai Halving Kee...
262 Prediksi Prospek Pasar Cryptocurrency Indonesia...
301 Selasa 30 April 2024 10 hari setelah halving #...
344 Halving Selesai Harga Bitcoin Akan Tetap Naik ...
367 Trader pasca halving menjadi seperti: #Bitcoin...
406 Terjemahan Persediaan Bitcoin sedang menuju go...
448 @DappRadar Setelah halving Bitcoin tetap menja...
454 Halving Bitcoin Ancam Bisnis Penambang Kecil d...
456 @anggaandinata Demand untuk bitcoin semakin me...
457 Cek suara dulu Yang kemarin: - bilang kalau tu...
458 Analis Prediksi Nasib Bitcoin Usai Halving Kee...
460 Prediksi Prospek Pasar Cryptocurrency Indonesia...
463 Selasa 30 April 2024 10 hari setelah halving #...
465 Halving Selesai Harga Bitcoin Akan Tetap Naik ...
```

Gambar 3. 3 Delete Duplicates

3.4 Wordcloud

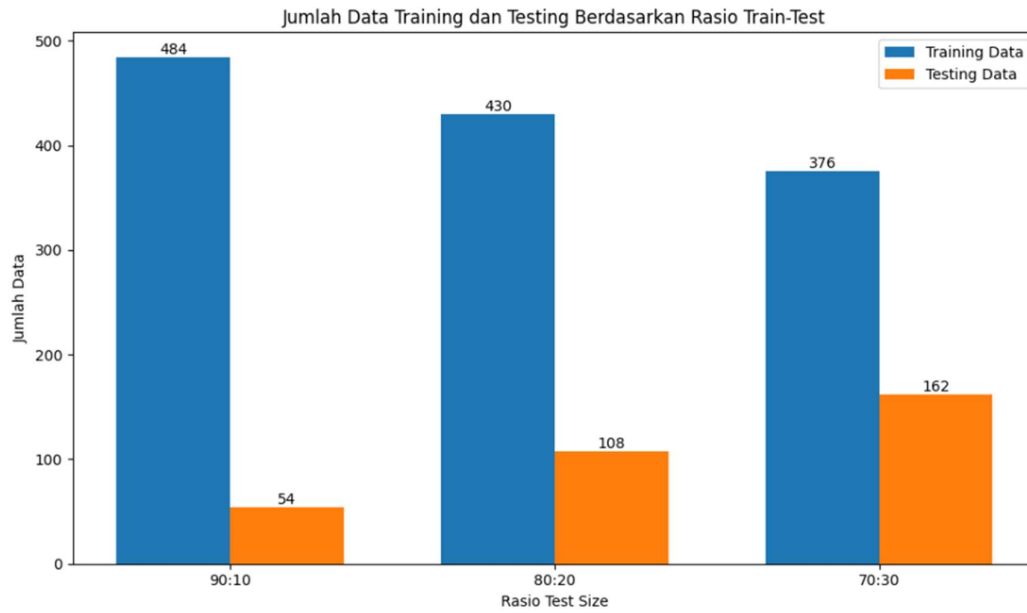
Seberapa sering suatu kata muncul dalam data akan menentukan ukuran kata tersebut dalam *wordcloud*. Semakin sering kata itu muncul dalam data, semakin besar pula ukurannya dalam *wordcloud*. Wordcloud yang ditampilkan memberikan visualisasi frekuensi kata dalam teks terkait peristiwa Bitcoin halving. Kata-kata seperti "turun" dan "naik" di tampilkan dengan ukuran besar, menunjukkan bahwa diskusi sering berfokus pada fluktuasi harga. Selain itu, kata-kata seperti "nih", "kripto", "harga", dan "peristiwa" sering muncul, menunjukkan topik utama yang dibahas adalah harga kripto dan peristiwa terkait. Kata-kata seperti "analisis" dan "investor" menunjukkan adanya diskusi teknis dan mendalam mengenai fenomena ini. Visualisasi *wordcloud* ditampilkan pada gambar 3.4.

332	0.050000	5.380239	0.236173	Bulan
2013	0.050000	5.023564	0.220516	Suku
337	0.050000	4.928254	0.216333	Bunga
1109	0.050000	4.687092	0.205747	Koreksi
521	0.050000	4.618099	0.202718	Digital
128	0.050000	3.892162	0.170852	Aser
116	0.050000	3.688563	0.161915	April
2185	0.050000	3.374906	0.148146	Turun
1117	0.050000	3.318816	0.145684	Kripto
803	0.050000	3.092043	0.135730	Harga
268	0.050000	1.415353	0.062129	bitcoin

Hasil perhitungan TF-IDF ini menunjukkan bobot pentingnya kata-kata dalam dokumen dataset. Array position menunjukkan posisi term didalam index dokumen, term seperti "buruk","kolaps" memiliki nilai IDF tertinggi sebesar 6.633002, menunjukkan bahwa kata tersebut sangat jarang muncul dalam dokumen. Nilai TF-IDF yang tinggi, seperti pada kata "buruk" dan "kolaps" dengan nilai 0.291165, menunjukkan bahwa meskipun frekuensi kemunculannya rendah, kata tersebut sangat penting didalam dokumen. Term-term ini memiliki signifikansi tinggi dalam klasifikasi sentimen pada dokumen.

3.6 Split Data

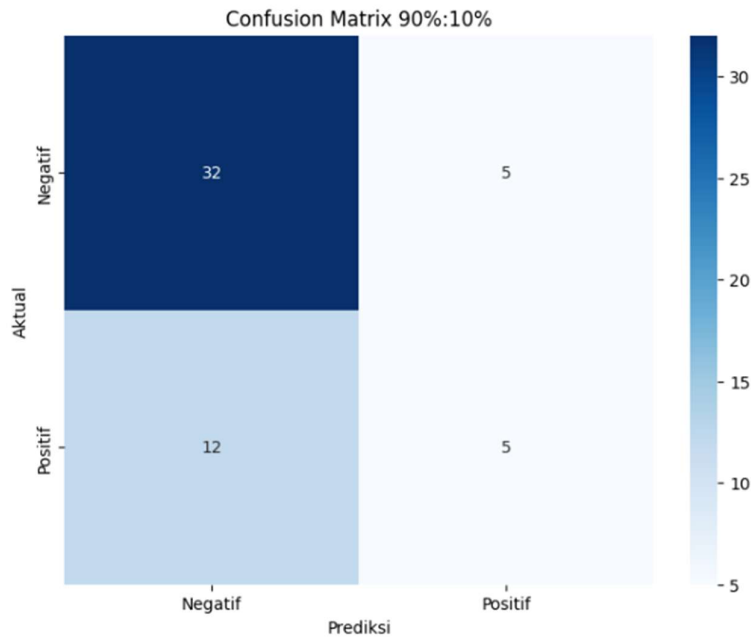
Pada penelitian ini, 538 data akan dibagi dengan beberapa rasio, yaitu 90:10,80:20, 70:20. Dari total data tersebut digunakan untuk Pembagian data dengan rasio 90:10 data training berjumlah 484, dan data testing berjumlah 54. Untuk 80:20 jumlah data training 430 dan data testing berjumlah 108 data. Terakhir untuk 70:30 jumlah data training 376 data dan 162 digunakan sebagai data testing. Visualisasi split data ditampilkan pada gambar 3.6.



Gambar 3. 5 Split Dataset

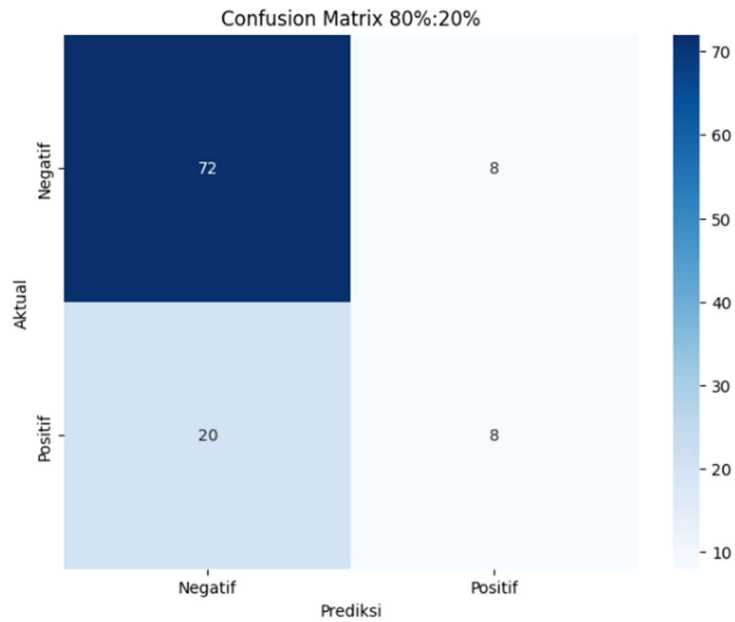
3.7 Klasifikasi

Model Naïve Bayes yang digunakan dalam penelitian ini telah diuji dengan berbagai rasio pembagian data, yaitu 90:10, 80:20, dan 70:30, untuk menentukan rasio yang paling optimal dalam klasifikasi sentimen. Berikut hasil klasifikasi ditampilkan melalui Confusion Matrix.



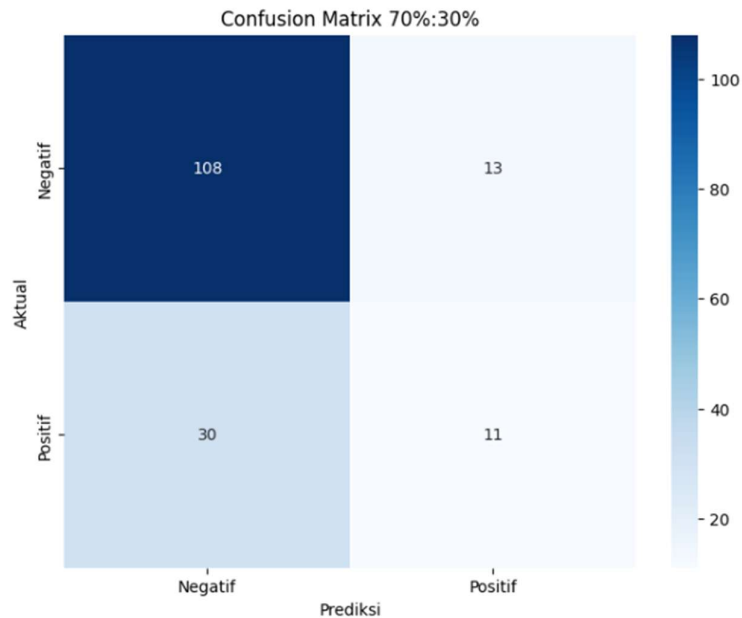
Gambar 3. 6 Confusion Matrix Rasio 90:10

Pada gambar 3.7, pengujian rasio 90:10, model Naive Bayes dapat mengidentifikasi 32 true negatives dan 5 true positives, namun juga terdapat 12 false negatives dan 5 false positives. Hasil tersebut menunjukkan bahwa meskipun model cenderung efektif dalam mengklasifikasikan kasus negatif, namun kemampuannya dalam mengenali kasus positif model belum maksimal.



Gambar 3. 7 *Confusion Matrix 80:20*

Pada gambar 3.8, rasio pengujian 80:20, terdapat peningkatan jumlah true positives menjadi 8, namun false negatives juga meningkat menjadi 20. Hal ini mengindikasikan bahwa dengan memperluas data uji, model mampu meningkatkan kemampuannya dalam mengklasifikasikan kasus positif dengan lebih baik daripada sebelumnya.

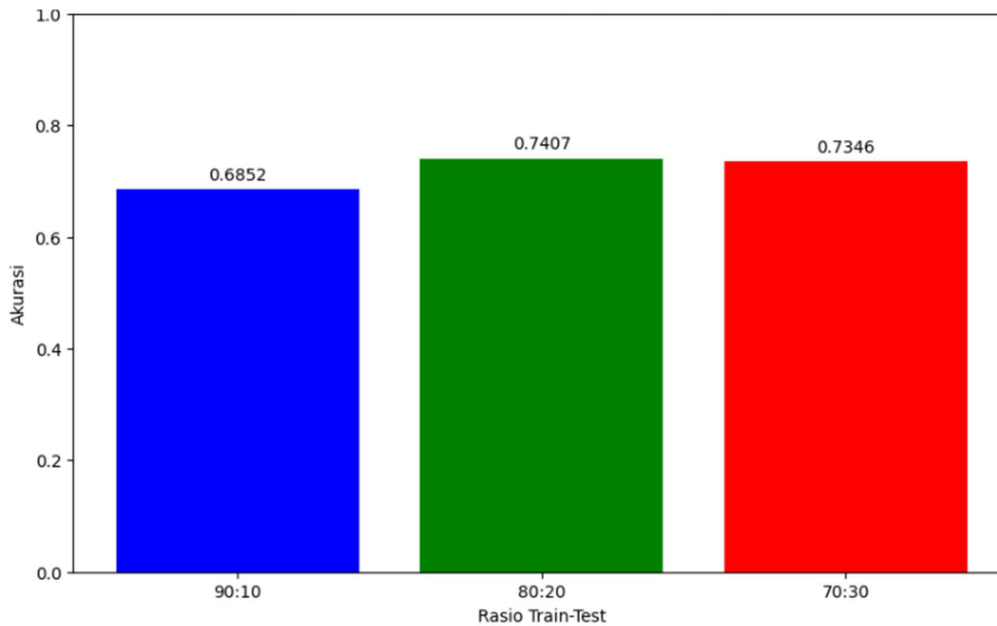


Gambar 3. 8 Confusion Matrix Rasio 70:30

Pada gambar 3.8, pengujian dengan rasio 70:30, model berhasil meningkatkan true positives menjadi 11 dengan false positives rendah sebanyak 13, namun terdapat peningkatan false negatives menjadi 30. Meskipun demikian, penambahan data uji menunjukkan potensi untuk meningkatkan kemampuan model dalam mengenali kasus positif dengan lebih baik.

3.8 Evaluasi

Setelah dilakukan pengujian dengan 3 model rasio yaitu 90:10,80:20,70:30 didapatkan akurasi untuk masing masing rasio sebesar 68.5% untuk pengujian menggunakan 90:10, dan 74% untuk pengujian dengan rasio 80:20, dan 73.4% untuk pengujian dengan rasio 70:30, dengan berbagai model rasio pengujian tersebut, menunjukkan bahwa model Naive Bayes yang dilatih dengan teknik teks *preprocessing* yang benar serta pembobotan TF-IDF memberikan performa yang cukup baik. Gambar 3.10 menunjukkan perbandingan akurasi model dalam melakukan klasifikasi dengan rasio yang telah di tentukan:



Gambar 3. 9 Perbandingan Accuracy Naive Bayes

Hasil evaluasi menunjukkan bahwa model *Naive Bayes* dengan rasio pengujian 70:30 mencapai akurasi sebesar 0.7346. Ini berarti model berhasil mengklasifikasikan sekitar 73.46% data dengan benar. Pada rasio pengujian 80:20, model menunjukkan peningkatan performa dengan akurasi sebesar 0.7407, mengindikasikan bahwa sekitar 74.07% data diklasifikasikan dengan benar. Sementara pada rasio pengujian 90:10, akurasi model menurun menjadi 0.6852, berarti model mampu mengklasifikasikan sekitar 68.52% data dengan benar. Perbedaan akurasi ini menunjukkan bahwa performa model sedikit bervariasi tergantung pada rasio pembagian data latih dan data uji. Pada rasio 70:30 dan 80:20, model menunjukkan performa yang lebih baik dibandingkan dengan rasio 90:10. Hal tersebut dapat disebabkan oleh berbagai faktor, seperti jumlah data latih yang lebih besar pada rasio 70:30 dan 80:20, yang memungkinkan model untuk mempelajari lebih banyak pola dari data sehingga kemampuan model dalam mengklasifikasikan meningkat. Dan untuk rasio 90:10, jumlah data latih yang lebih sedikit mungkin tidak cukup untuk menangkap variasi yang ada dalam data, sehingga menyebabkan penurunan akurasi. Secara keseluruhan, hasil ini menunjukkan pemilihan rasio yang tepat dalam pembagian data latih dan data uji untuk mendapatkan performa model yang optimal.

Pemilihan rasio yang baik dapat membantu model mempelajari pola dengan lebih baik dan menghasilkan prediksi yang lebih akurat.

Berikut penghitungan model akurasi rasio 90:10 pada persamaan rumus (4.3).

$$Akurasi = \frac{TP + TN}{Total} + \frac{5 + 32}{32 + 5 + 12 + 5} = \frac{37}{54} = 0.685 \quad (4.3)$$

Berikut penghitungan model akurasi rasio 80:20 pada persamaan rumus (4.4).

$$Akurasi = \frac{TP + TN}{Total} + \frac{8 + 72}{72 + 8 + 20 + 8} = \frac{80}{108} = 0.740 \quad (4.4)$$

Berikut penghitungan model akurasi rasio 70:30 pada persamaan rumus (4.5).

$$Akurasi = \frac{TP + TN}{Total} + \frac{11 + 108}{108 + 13 + 30 + 11} = \frac{119}{162} = 0.734 \quad (4.5)$$