

ANALISIS SENTIMEN TERHADAP *CRYPTOCURRENCY*
BERDASARKAN *TWEET* DAN *RETWEET* DENGAN
MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN *CHI-*
SQUARE

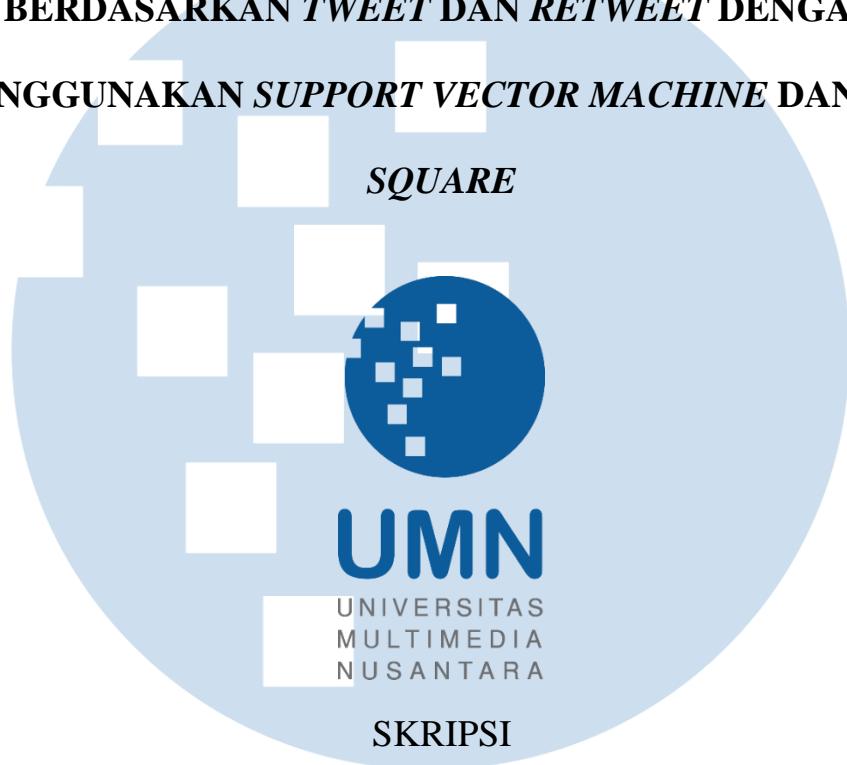


SKRIPSI

Isabella Donita Hasan
00000028068

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2022

**ANALISIS SENTIMEN TERHADAP CRYPTOCURRENCY
BERDASARKAN TWEET DAN RETWEET DENGAN
MENGGUNAKAN SUPPORT VECTOR MACHINE DAN CHI-**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Sistem Komputer (S.Kom)

Isabella Donita Hasan

00000028068

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

N U S A N T A R A

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Isabella Donita Hasan

Nomor Induk Mahasiswa : 00000028068

Program studi : Sistem Informasi

Skripsi dengan judul:

“Analisis Sentimen terhadap *Cryptocurrency* Berdasarkan *Tweet* dan *Retweet* dengan Menggunakan *Support Vector Machine* dan *Chi-Square*”

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS** untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 14 Juni 2022



Isabella Donita Hasan

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS SENTIMEN TERHADAP *CRYPTOCURRENCY* BERDASARKAN
TWEET DAN RETWEET DENGAN MENGGUNAKAN *SUPPORT VECTOR
MACHINE* DAN *CHI-SQUARE*

Oleh

Nama : Isabella Donita Hasan
NIM : 00000028068
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Selasa, 28 Juni 2022
Pukul 08.00 s.d 10.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang

Wella S.Kom., M.MSI.

305119101

Pembimbing I

8-7-2022

Ir. Raymond Sumardi Oetama, M.C.I.S

328046803/023897

Penguji

07/07/2022

Ririn Ikana Desanti, S.Kom., M.Kom.

0313058001

Pembimbing II

Aldo Lionel Saonard, S.Si.Kom., M.Komm

0308099401/L00999

Ketua Sistem Informasi

07/07/2022

Ririn Ikana Desanti, S.Kom., M.Kom.

0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas academica Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Isabella Donita Hasan
NIM : 00000028068
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Nonekslusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul.

“ANALISIS SENTIMEN TERHADAP CRYPTOCURRENCY BERDASARKAN TWEET DAN RETWEET DENGAN MENGGUNAKAN SUPPORT VECTOR MACHINE DAN CHI-SQUARE”

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Nonekslusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 14 Juni 2022

Yang menyatakan,

Isabella Donita Hasan

UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa atas berkat yang telah diberikan kepada penulis laporan skripsi yang berjudul "*Analisis Sentimen Terhadap Cryptocurrency Berdasarkan Tweet Dan Retweet Dengan Menggunakan Support Vector Machine Dan Chi-Square*" dapat selesai dengan baik dan tepat waktu. Laporan skripsi ini disusun guna memenuhi salah satu prasyarat kelulusan Program Strata I Sistem Informasi di Universitas Multimedia Nusantara. Setelah melewati proses penyusunan laporan ini dengan segala hambatan dan rintangan, penulis menyadari dalam penyusunan skripsi ini tidak akan selesai tanpa bantuan dari berbagai pihak. Oleh karena itu penulis ingin mengucapkan terima kasih sebesar – besarnya kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Bapak Ir. Raymond Sunardi Oetama, M.C.I.S selaku Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesaikan tugas akhir ini.
5. Bapak Aldo Lionel Saonard, S.Si.Kom., M.Kom. selaku Pembimbing kedua yang telah memberikan bimbingan, arahan, dan motivasi atas terselesaikan tugas akhir ini.
6. Keluarga dan teman saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Laporan skripsi ini telah diusahakan untuk dibuat sebaik mungkin, namun penulis juga dapat menyadari bahwa skripsi ini masih ada kekurangan dalam penulisannya. Maka dari itu diharapkan kritik dan saran yang membangun dari berbagai pihak. Semoga skripsi ini dapat memberikan informasi yang bermanfaat dan inspirasi yang bermanfaat bagi pembaca.

Tangerang, 14 Juni 2022

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Isabella Donita Hasan

ANALISIS SENTIMEN TERHADAP CRYPTOCURRENCY

BERDASARKAN TWEET DAN RETWEET DENGAN

MENGGUNAKAN SUPPORT VECTOR MACHINE DAN CHI-

SQUARE

Isabella Donita Hasan

ABSTRAK

Tingginya minat masyarakat terhadap *cryptocurrency* semakin bertambah seiring berjalananya waktu. *Cryptocurrency* menjadi perhatian dunia karena sifatnya yang unik dan tidak dapat diprediksi. Pada zaman sekarang *cryptocurrency* dapat digunakan untuk berbagai macam hal, salah satunya adalah untuk investasi. Dalam berinvestasi *cryptocurrency* dibutuhkan berbagai analisa untuk mempertimbangkan arus *trade*.

Salah satu cara untuk menganalisisnya adalah dengan analisis sentimen masyarakat yang terdiri dari sentimen positif maupun negatif. Penelitian dilakukan dengan menganalisis sentimen masyarakat dengan menggunakan algoritma *machine learning*, yaitu *Support Vector Machine*. Selain dengan menggunakan *Support Vector Machine*, digunakan juga *Chi-square* dalam *feature selection* agar dapat membantu dalam pengurangan *noise* di sebuah kalimat.

Dalam penelitian ini *dataset* dibagi menjadi dua, yaitu *dataset* dengan *retweet* dan tanpa *retweet*. Melalui penelitian ini dihasilkan rata – rata akurasi sebesar 86,87%. Selain itu, dapat diketahui juga hasil analisis sentimen dari *dataset* tanpa *retweet* menghasilkan nilai rata – rata *precision*, *recall*, akurasi sebesar 79,25% dan *f1-score* sebesar 77,5%. Sedangkan *dataset* dengan *retweet* menghasilkan nilai *precision* sebesar 94,75% dan nilai *recall*, *f1-score* dan akurasi sebesar 94,5%.

Kata kunci: *Chi-square*, *Cryptocurrency*, *Support Vector Machine*

UNIVERSITAS
MULTIMEDIA
NUSANTARA

ANALISIS SENTIMEN TERHADAP CRYPTOCURRENCY

BERDASARKAN TWEET DAN RETWEET DENGAN

MENGGUNAKAN SUPPORT VECTOR MACHINE DAN CHI-SQUARE

Isabella Donita Hasan

ABSTRACT (English)

The high public interest in cryptocurrencies is increasing over time. Cryptocurrencies are gaining worldwide attention because of their unique and unpredictable nature. Nowadays, cryptocurrencies can be used for various things, one of which is for investment. In investing in cryptocurrencies, various analyzes are needed to consider trade flows.

One way to analyze it is by analyzing public sentiment which consists of positive and negative sentiments. The research was conducted by analyzing public sentiment using a machine learning algorithm, namely the Support Vector Machine. In addition to using the Support Vector Machine, Chi-square is also used in feature selection to help reduce noise in a sentence.

In this study, the dataset is divided into the dataset with retweets and without retweets. Through this research, the average accuracy is 86.87%. In addition, it can also be seen that the results of the sentiment analysis from the dataset without retweets yielded an average value of precision, recall, accuracy of 79.25%, and f1-score of 77.5%. While the dataset with retweets produces a precision value of 94.75% and a recall value, f1-score, and an accuracy of 94.5%.

Keywords: Chi-square, Cryptocurrency, Support Vector Machine

UNIVERSITAS
MULTIMEDIA
NUSANTARA

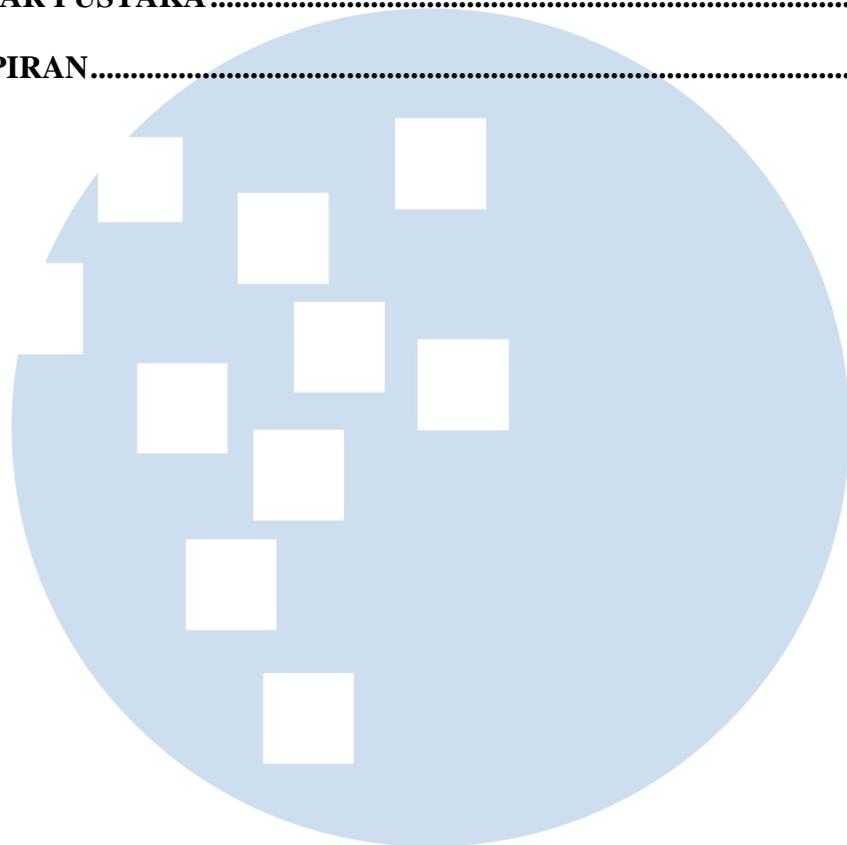
DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT (English).....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS	xvi
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat Penelitian.....	4
1.4.1 Tujuan Penelitian	4
1.4.2 Manfaat Penelitian	5
1.5 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI	7
2.1 <i>Cryptocurrency</i>	7

2.2	Analisis Sentimen.....	7
2.3	<i>Data Preprocessing</i>	9
2.4	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	11
2.5	<i>Chi-square</i>	12
2.6	<i>Support Vector Machine (SVM)</i>	12
2.7	<i>Evaluation</i>	13
2.7.1	<i>Accuracy</i>	13
2.7.2	<i>Recall</i>	13
2.7.3	<i>Precision</i>	14
2.7.4	<i>F1-Score</i>	14
2.7.5	<i>Confusion Matrix</i>	14
2.8	Penelitian Terdahulu.....	15
	BAB III METODOLOGI PENELITIAN	19
3.1	Gambaran Umum Objek Penelitian	19
3.2	Metode Penelitian.....	19
3.2.1	<i>Data Extraction</i>	20
3.2.2	<i>Data Labelling</i>	20
3.2.3	<i>Data Preprocessing</i>	20
3.2.4	<i>Feature Extraction</i>	21
3.2.5	<i>Feature Selection</i>	22
3.2.6	<i>Split Data</i>	23
3.2.7	<i>Classification</i>	24
3.2.8	<i>Evaluation</i>	25

3.3	Variabel Penelitian	25
3.3.1	Variabel Independen	25
3.3.2	Variabel Dependen.....	25
3.4	Teknik Pengumpulan Data	26
3.5	Teknik Pengambilan Sampel	27
3.6	Teknik Analisis Data	27
BAB IV HASIL PENELITIAN.....	28	
4.1	<i>Data Extraction</i>	28
4.2	<i>Data Labelling</i>	30
4.3	<i>Data Preprocessing</i>	33
4.3.1	<i>Data Cleaning & Stopword Removal</i>	33
4.3.2	<i>Tokenization</i>	34
4.3.3	<i>Case folding</i>	34
4.3.4	<i>Stemming & Lematization</i>	34
4.4	<i>Feature Extraction</i>	38
4.5	<i>Feature Selection</i>	39
4.6	<i>Split Data</i>	46
4.7	<i>Classification</i>	46
4.8	<i>Evaluation</i>	48
4.9	Hasil Diskusi	61
BAB V SIMPULAN & SARAN	64	
5.1	Simpulan	64
5.2	Saran	64

DAFTAR PUSTAKA	65
LAMPIRAN.....	72



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

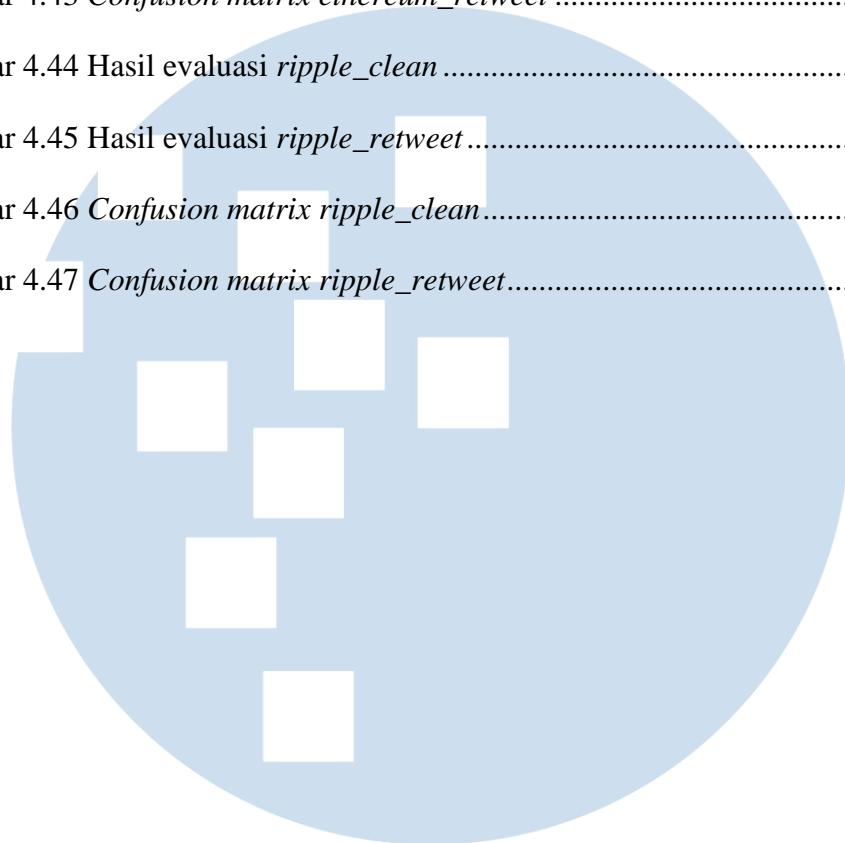
Tabel 3.1 Perbandingan <i>Feature Extraction</i>	21
Tabel 3.2 Perbandingan Metode <i>Feature Selection</i>	22
Tabel 3.3 Perbandingan Algoritma Analisis Sentimen	24
Tabel 3.4 <i>Hastag</i> dan <i>Keywords</i>	26
Tabel 4.1 Hasil <i>Labelling Dataset</i>	32
Tabel 4.2 Nilai k Pada Setiap <i>Dataset</i>	45
Tabel 4.3 <i>Classification Report Cryptocurrency</i>	50
Tabel 4.4 <i>Classification Report Bitcoin</i>	53
Tabel 4.5 <i>Classification Report Ethereum</i>	55
Tabel 4.6 <i>Classification Report Ripple</i>	58
Tabel 4.7 Perbandingan <i>Confusion Matrix</i> Pada Setiap <i>Dataset</i>	59
Tabel 4.8 Perbandingan <i>Precision</i> , <i>Recall</i> , <i>F1-Score</i> , dan <i>Accuracy</i> Pada Setiap <i>dataset</i>	61
Tabel 4.9 Perbandingan <i>Precision</i> , <i>Recall</i> , <i>F1-Score</i> , dan <i>Accuracy</i> pada <i>dataset clean</i>	62
Tabel 4.10 Perbandingan <i>Precision</i> , <i>Recall</i> , <i>F1-Score</i> , dan <i>Accuracy</i> pada <i>dataset retweet</i>	62

DAFTAR GAMBAR

Gambar 1.1 Media Sosial yang Paling Banyak Digunakan [7]	2
Gambar 2.1 Cara Kerja <i>Support Vector Machine</i>	13
Gambar 3.1 Kerangka Kerja	19
Gambar 4.1 <i>Key & Token</i> Twitter	28
Gambar 4.2 <i>Syntax Data Crawler</i>	29
Gambar 4.3 Hasil <i>Data Extraction</i>	30
Gambar 4.4 <i>Tweet Lengkap</i>	30
Gambar 4.5 Contoh Hasil <i>Labelling</i>	31
Gambar 4.6 Visualisasi perbandingan data positif dan negatif hasil <i>labelling</i>	32
Gambar 4.7 <i>Syntax Cleaning & Stopword Removals</i>	33
Gambar 4.8 <i>Syntax Tokenization</i>	34
Gambar 4.9 <i>Syntax Case Folding</i>	34
Gambar 4.10 <i>Syntax Stemming & Lematization</i>	35
Gambar 4.11 <i>Syntax</i> penghapusan data <i>duplicate</i> pada <i>tweets_lematization</i>	35
Gambar 4.12 Contoh <i>wordcloud</i> positif (kiri) dan negatif (kanan)	36
Gambar 4.13 <i>Syntax</i> Polaritas dan Subjektivitas.....	36
Gambar 4.14 Hasil perhitungan polaritas positif dan negatif	37
Gambar 4.15 Visualisasi perbandingan hasil perhitungan polaritas	38
Gambar 4.16 <i>Syntax Feature Extraction</i>	38
Gambar 4.17 <i>Syntax Feature Selection</i>	39
Gambar 4.18 <i>Chart Feature Selection crypto_clean</i>	40
Gambar 4.19 <i>Chart Feature Selection crypto_retweet</i>	40

Gambar 4.20 <i>Chart Feature Selection bitcoin_clean</i>	41
Gambar 4.21 <i>Chart Feature Selection bitcoin_retweet</i>	42
Gambar 4.22 <i>Chart Feature Selection ethereum_clean</i>	42
Gambar 4.23 <i>Chart Feature Selection ethereum_retweet</i>	43
Gambar 4.24 <i>Chart Feature Selection ripple_clean</i>	44
Gambar 4.25 <i>Chart Feature Selection ripple_retweet</i>	44
Gambar 4.26 Contoh pemilihan <i>Feature Selection</i> dengan k = 300	45
Gambar 4.27 <i>Syntax split data</i>	46
Gambar 4.28 <i>Syntax data Scaling & Transform</i>	47
Gambar 4.29 <i>Syntax Pembangunan model</i>	48
Gambar 4.30 Hasil <i>GridSearchCV</i>	48
Gambar 4.31 <i>Syntax Classification Report</i>	49
Gambar 4.32 Hasil evaluasi <i>crypto_clean</i>	49
Gambar 4.33 Hasil evaluasi <i>crypto_retweet</i>	50
Gambar 4.34 <i>Confusion matrix crypto_clean</i>	51
Gambar 4.35 <i>Confusion matrix crypto_retweet</i>	51
Gambar 4.36 Hasil evaluasi <i>bitcoin_clean</i>	52
Gambar 4.37 Hasil evaluasi <i>bitcoin_retweet</i>	52
Gambar 4.38 <i>Confusion matrix bitcoin_clean</i>	53
Gambar 4.39 <i>Confusion matrix bitcoin_retweet</i>	54
Gambar 4.40 Hasil evaluasi <i>ethereum_clean</i>	54
Gambar 4.41 Hasil evaluasi <i>ethereum_retweet</i>	55
Gambar 4.42 <i>Confusion matrix ethereum_clean</i>	56

Gambar 4.43 <i>Confusion matrix ethereum_retweet</i>	56
Gambar 4.44 Hasil evaluasi <i>ripple_clean</i>	57
Gambar 4.45 Hasil evaluasi <i>ripple_retweet</i>	57
Gambar 4.46 <i>Confusion matrix ripple_clean</i>	58
Gambar 4.47 <i>Confusion matrix ripple_retweet</i>	59



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

(2.1) Rumus TF-IDF	11
(2.2) Rumus TF-IDF	11
(2.3) Rumus <i>Chi-square</i>	12
(2.4) Rumus <i>Accuracy</i>	13
(2.5) Rumus <i>Recall</i>	13
(2.6) Rumus <i>Precision</i>	14
(2.7) Rumus <i>F1-Score</i>	14



LAMPIRAN

Lampiran A <i>Form Bimbingan Skripsi 1</i>	72
Lampiran B <i>Form Bimbingan Skripsi 2</i>	73
Lampiran C Hasil <i>Turnitin</i>	74

