

PERANCANGAN APLIKASI KLASIFIKASI SENTIMEN BERBASIS WEB TERHADAP MATA UANG KRIPTO

Fahmi Ilmawan Sulaiman¹, Wing Wahyu Winarno², Mei Parwanto Kurniawan³

^{1,2,3} Program Magister Teknik Informatika Universitas Amikom Yogyakarta

e-mail: [1fahmi.1315@students.amikom.ac.id](mailto:fahmi.1315@students.amikom.ac.id), [2wing@stieykpn.ac.id](mailto:wing@stieykpn.ac.id),
[3meikurniawan@amikom.ac.id](mailto:meikurniawan@amikom.ac.id)

Abstract

Sentiment analysis is a collection of opinions that studies opinions about something like an issue, product, event or topic. In deciding to choose cryptocurrency, the majority of Indonesian people tend to only read the latest reviews about cryptocurrencies and even a few will lead to wrong conclusions. To analyze cryptocurrency sentiment easily and quickly, a sentiment analysis system with a convenient and easy-to-use interface is required. The application to analyze this sentiment is web-based which is made using the PHP (Hypertext Preprocessor) framework by applying the MVC (Model View Controller) concept. This application can import data in CSV format which is then processed into an application using the naive Bayes algorithm in the sentiment classification process which is then evaluated by the confusion matrix method. The reviews are taken from social media twitter and it is hoped that this application will be able to provide classification information accurately, precisely and reliably.

Keywords—*Sentiment Analysis, Cryptocurrency, Naïve Bayes*

PENDAHULUAN

Perkembangan teknologi informasi yang kian meningkat seiring berjalannya waktu berdampak pada pertukaran informasi dan komunikasi yang semakin mudah. Informasi yang hanya berupa teks kebanyakan didorong oleh peningkatan pengguna media sosial twitter sehingga muncul kebutuhan penyajian data yang memudahkan pengguna mendapatkan informasi yang banyak dan akurat. Media sosial *twitter* adalah salah satu media komunikasi yang populer saat ini [1].

Pertumbuhan *twitter* terus meningkat setiap waktu, sehingga hal tersebut dimanfaatkan para pengguna *twitter* untuk menyampaikan informasi mengenai mata uang kripto atau *crypto currency*. Semakin banyak informasi dan pendapat yang disampaikan pengguna melalui akun masing-masing maka informasi dan pendapat tersebut membentuk opini masyarakat terhadap mata uang kripto itu sendiri. Namun untuk mengetahui positif, negatif ataupun netral sangatlah sulit untuk di analisa secara manual terlebih lagi jika data yang diproses sangat banyak.

Pengelompokan opini atau analisis sentimen adalah salah satu metode untuk mengetahui sentimen seseorang atau sekelompok orang terhadap produk, layanan, isu atau golongan tertentu dari berbagai platform media sosial. Untuk mengklasifikasi sentimen dapat dilakukan dengan mengumpulkan data melalui twitter dengan topik tertentu, salah satunya dapat diperoleh melalui cuitan atau tweet yang terdapat dalam twitter [2].

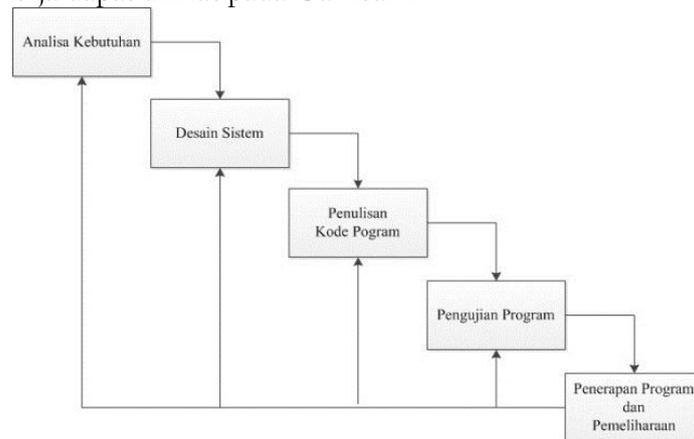
Dalam salah satu penelitian yang membahas analisis sentimen evaluasi untuk meningkatkan kualitas belajar dan standarisasi akademik. Pengumpulan data dilakukan menggunakan kuesioner yang diisikan kemudian menganalisa sentimen untuk mengetahui saran tersebut bernilai positif, negatif, atau netral. Data komentar ini kemudian diklasifikasi menggunakan metode *Naïve Bayes Classifier*. Dilakukan pengujian terhadap 4 semester menghasilkan akurasi 80,1%, *Precision* 80,3%, *Recall* 80,3%, dan *F1-Score* 80% [3].

Dengan memperoleh gambaran umum persepsi suatu topik, masyarakat sangat terbantu dengan adanya analisis sentimen yang dapat dengan mudanya mengelompokkan jenis opini menjadi kategori positif, negatif, ataupun netral. Dengan aplikasi ini diharapkan dapat membantu masyarakat dalam menganalisa ulasan dalam jumlah besar dan dapat memberikan informasi mengenai tren dari suatu mata uang kripto secara tepat dengan menggunakan metode yang tepat.

METODE PENELITIAN

Alur Penelitian

Berikut adalah gambaran umum yang menunjukkan alur dari penelitian yang dilakukan. Alur kerja dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1, penelitian ini menggunakan metode Waterfall karena model pengembangan aplikasi bersifat linear dari perencanaan hingga pemeliharaan di tahap akhir. Metode Waterfall menjabarkan pendekatan yang cukup sistematis dan berurutan untuk pengembangan perangkat lunak.

a. Analisis Kebutuhan

Tujuan analisis ini adalah memahami dengan sebaik-baiknya kebutuhan dari sebuah sistem baru dan mengembangkan sebuah sistem yang dapat menampung kebutuhan itu atau memutuskan bahwa sebenarnya sistem baru tidak diperlukan untuk dibuat. Untuk mengetahui kebutuhan pengguna, maka dilakukan analisis kebutuhan dengan membagikan kuesioner dengan google form yang disebarakan melalui grup di aplikasi Telegram.

b. Desain Sistem

Tahapan ini bertujuan untuk memberikan gambaran lengkap tentang alur sistem yang dibangun dan bagaimana tampilan dari sistem yang dibangun. Sehingga dapat membantu secara detail terkait kebutuhan hardware dan sistem yang dibutuhkan, juga mendefinisikan arsitektur sistem yang dibuat secara keseluruhan.

c. Penulisan Koding

Pada bagian ini dilakukan pembuatan dari sebuah sistem berdasarkan rancangan dan desain yang telah dibuat untuk tahap analisis dan desain. Aplikasi dibangun sesuai dengan analisis kebutuhan untuk membuat form input dan output dengan aplikasi berbasis web yang menggunakan bahasa pemrograman PHP.

d. Pengujian Program

Pengujian sistem dilakukan untuk mengetahui apakah implementasi sistem sudah sesuai dengan perancangan dan analisis kebutuhan. Selain itu, untuk mengetahui apakah

ada error atau bug pada sistem. Di dalam proses pengujian terdapat metode yang digunakan yaitu pengujian validasi (*black box*) dan pengujian akurasi algoritma naïve bayes menggunakan *confusion matrix*.

e. Implementasi dan Pemeliharaan

Pada tahap ini, aplikasi yang sudah jadi kemudian dijalankan atau dioperasikan oleh penggunanya. Terdapat 3 hal dalam tahap pemeliharaan yaitu perbaikan kesalahan, perbaikan implementasi unit sistem, peningkatan jasa sistem sesuai kebutuhan baru.

Metode Pengumpulan Data

Dalam penelitian ini menggunakan beberapa metode pengumpulan data untuk mencari data dan informasi yang berkaitan dengan penelitian ini, berikut merupakan metode penelitian yang digunakan:

a. Studi Kepustakaan

Dalam metode studi kepustakaan dilakukan dengan mengumpulkan jurnal, paper, buku, maupun situs internet yang menjadi sumber pustaka yang berkaitan dengan penelitian. Selain itu, metode studi literatur bertujuan untuk mempelajari baik teori ataupun konsep yang berkaitan dengan penelitian ini.

Tabel 1. Matriks literatur review

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Perbandingan
1	Studi Perbandingan Algoritma – Algoritma Stemming untuk Dokumen Teks Bahasa Indonesia	Azhar, Siti, Luh, dan Okfalisa, 2019	Melakukan perbandingan algoritma vega dengan algoritma nazief dan ardani dengan tolak ukur seperti lama proses stemming, tingkat akurasi, overstemming, dan understemming.	Algoritma Nazief dan Adriani merupakan yang terbaik, hal ini dikarenakan terdapat penambahan aturan untuk reduplikasi, penambahan aturan untuk awalan dan akhiran dalam meningkatkan presisi dari setiap kata yang distemming.	Pada penelitian sebelumnya dibuat dalam program sederhana untuk memproses dokumen teks inputan, sedangkan pada penelitian yang akan dibuat, aplikasi akan dibangun menggunakan bahasa pemrograman PHP dengan mengadaptasi konsep MVC
2	Analisis Sentimen Pada	Muthhia, e-Journal	Mengintegrasikan metode pemilihan fitur	Akurasi <i>Naïve Bayes</i> sebelum dan sesudah menggunakan	Pada penelitian ini hanya mengelompok

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Perbandingan
	Review Restoran dengan Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes	Nusa Mandiri, 2017	<i>Genetic Algorithm</i> dalam menganalisa sentimen padareview restoran dengan teks bahasa indonesia menggunakan Algoritma Naïve Bayes.	an penggabungan fitur mengalami peningkatan sebesar 4%, yaitu sebesar 86.50% sebelum dan 90.50% setelah menggunakan penggabungan metode <i>Genetic Algorithm</i> .	n review pengguna menjadi opini positif atau negatif, sedangkan pada penelitian yang akan dibangun terdapat 3 pengelompokan opini yaitu positif, negatif, dan netral.

b. Pengumpulan Data Twitter

Data dari tweet dengan dikumpulkan menggunakan salah satu topik tentang cryptocurrency misalnya seperti kata kunci *Cryptocurrency*, yang dikumpulkan dengan memanfaatkan search API Twitter yang telah disediakan oleh Twitter. Metode *crawling* memungkinkan untuk memanggil data pada seluruh twitt dengan rentang waktu tertentu yang berkaitan dengan *Cryptocurrency* sebagai sumber data. Selanjutnya data yang telah terkumpul kemudian dilakukan proses pengolahan dan menghasilkan dataset yang siap diklasifikasi. Gambaran fungsi *crawltesting* yang berisi kode token yang didapat dari API twitter pada program PHP dapat dilihat pada Gambar 2.

```

11 // menentukan keyword yang akan di cari
12 $keyword = 'Dogecoin';
13
14 // ganti dengan API twitter anda
15 $key = 'SzoT7vexriXbp4nx600jbhnGn';
16 $secret_key = 'XfIhJinubF7Wbt5gYPXxfJw0dY8Xs16NkBEnrYRzfVpWGktQI';
17 $token = '1039077944222609409-XC8rNwZbNSKUye1dGR9NoFYAGtQarW';
18 $secret_token = 'fKsv39pIzMKAcRvxrQAqUh1QKaYZVwsKpSYntrgHaV8kI';

```

Gambar 2. Fungsi Koneksi Token

Dalam pengambilan data uji langkah pertama harus masuk menggunakan akun twitter untuk mendapatkan token. Cara mendapatkan token masuk pada website <https://developer.twitter.com/en/apps>, kemudian mengisi kolom pendaftaran sesuai ketentuan oleh pihak twitter. Token yang telah didapat dari twitter ini dimasukkan kedalam fungsi *crawltesting* dalam bahasa program PHP untuk koneksi API twitter sehingga user mendapatkan hak akses secara realtime dalam pencarian tweet.

#	TWEET	TANGGAL
1	RT @K_SugengWidodo: Gratis 2,7 Doge setiap hari untuk para pejuang regeh dengan cara menonton iklan - jawab pertanyaan - claim Minimal wit...	2021-05-22
2	RT @K_SugengWidodo: Gratis 2,7 Doge setiap hari untuk para pejuang regeh dengan cara menonton iklan - jawab pertanyaan - claim Minimal wit...	2021-05-22
3	RT @BackToTheDoge: @doge_memes sweet setup man #DogeCoin #Doge #dogearmyMEMEdiv https://t.co/mvLEF9Tnb	2021-05-22
4	@AromaFinance Bisa yok @elonmusk @SpaceX @dogecoin #bsc #arfi #BNB #binance	2021-05-22
5	RT @BackToTheDoge: @doge_memes sweet setup man #DogeCoin #Doge #dogearmyMEMEdiv https://t.co/mvLEF9Tnb	2021-05-22
6	#dogecoin #DogeFather #DogeCoinToTheMoon #DogeCoinArmy #HODL DKyua1tqQTQQ3aPWAKTpJzxtL3x3jFbu https://t.co/dNtluonO26	2021-05-22
7	@DogeCoinMemes @dogecoin Hi Bro	2021-05-22
8	Parents in 2010: Agar tere dost kuyei mein koodenge toh tu bhi koodega kya?! Parents in 2021: Agar tere dost #DogeCoin kharedenge toh tu bhi kharedega kya?! #DogeArmy #dogetotheemoon #Cryptoindia #DogeCoinArmy	2021-05-22
9	RT @BackToTheDoge: @doge_memes sweet setup man #DogeCoin #Doge #dogearmyMEMEdiv https://t.co/mvLEF9Tnb	2021-05-22
10	RT @BackToTheDoge: @doge_memes sweet setup man #DogeCoin #Doge #dogearmyMEMEdiv https://t.co/mvLEF9Tnb	2021-05-22

Menampilkan 1 sampai 10 dari 100 data

Sebelumnya 1 2 3 4 5 ... 10 Selanjutnya

PROSES CRAWLING TWEET

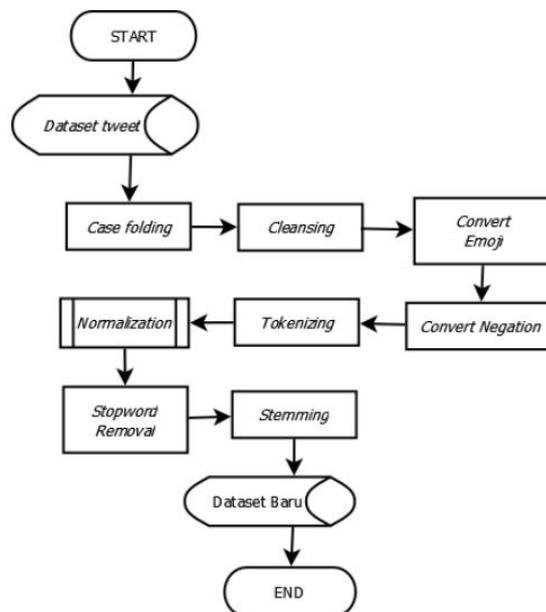
Gambar 3. Hasil Crawling Data

Metode Analisis Data

Metode Analisis data digunakan untuk mengolah data yang telah didapatkan dan dikumpulkan, yaitu:

a. *Preprocessing*

Preprocessing merupakan tahapan untuk mempersiapkan proses penyeleksian data agar data lebih terstruktur dan berkualitas serta dapat diolah dengan cepat dan tepat. Preprocessing sangat penting dalam pembuatan analisis sentimen, terutama untuk sentimen pada media sosial yang sebagian besar berisi kata – kata atau kalimat yang tidak formal dan tidak terstruktur serta memiliki noise yang besar [4]. Diagram alur tahapan-tahapan proses preprocessing dapat dilihat pada Gambar 4.

Gambar 4. *Preprocessing*

Pada tahap *Case Folding* huruf kapital pada semua dokumen tweet diubah menjadi huruf kecil kemudian tahapan *Cleansing* merupakan tahap pembersihan kata yang tidak penting selanjutnya *Convert Emoji* merupakan tahap yang dilakukan untuk mengubah simbol menjadi kata setelah itu untuk mengkonversi kata – kata negasi yang terdapat pada

suatu tweet terdapat pada tahap *Convert Negation*, Lalu terdapat tahap tokenizing yang berfungsi untuk memotong kata berdasarkan tiap kata dan menyusunnya menjadi kata tunggal, Agar semakin baik ditambahkan tahap *Normalization* sebagai pengubah kata yang tidak sesuai dengan EYD, Tahap ini terbagi menjadi empat bagian, yaitu konversi kata typo, konversi kata singkatan, konversi kata slang, konversi kata baku, dan konversi kata bahasa inggris, Selain konversi kata terdapat tahapan stopword removal untuk menghapus kata yang dianggap tidak mempunyai makna dengan topik dokumen, Terakhir terdapat tahap Stemming yang merupakan suatu proses untuk mengubah kata dalam suatu dokumen menjadi kata dasarnya.

b. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) digunakan merupakan teknik pembobotan kata yang mempunyai ide dasar untuk memberikan bobot pada setiap kata atau dokumen, dimana perhitungan bobot term dalam sebuah dokumen dilakukan dengan mengalikan nilai Term Frequency (TF) dengan Inverse Document Frequency (IDF), term dapat berbentuk kata, frasa atau jenis sintatik lainnya [5].

Langkah pertama adalah mencari nilai TF, dengan menggunakan formula TF murni (raw TF), dimana nilai TF diambil dari jumlah kemunculan suatu term dalam suatu dokumen. Contohnya, jika muncul empat (4) kali maka kata tersebut bernilai empat (4). Langkah selanjutnya adalah mencari nilai IDF untuk menunjukkan hubungan ketersediaan sebuah term dalam seluruh dokumen, dengan menggunakan rumus pada persamaan (1) sebagai berikut:

$$idf(t, D) = \log \left(\frac{N}{n_j} \right) \quad (1)$$

Proses selanjutnya setelah mendapatkan nilai TF dan IDF dari dokumen d. Kemudian dilakukan perhitungan bobot term t dalam sebuah dokumen dengan mengalikan nilai Term Frequency dengan Inverse Document Frequency pada persamaan sebagai berikut:

$$a_{ij} = tf_{ij} \times idf_j \quad (2)$$

$$a_{ij} = tf_{ij} \times \log \left(\frac{N}{n_j} \right) \quad (3)$$

Ketika N sama dengan n_j , maka a_{ij} menjadi nol, ini sering muncul dalam dataset yang kecil, jadi perlu diterapkan beberapa teknik perataan untuk meningkatkan rumus pada persamaan 3, maka dari persamaan tersebut dapat ditulis sebagai berikut [6]:

$$a_{ij} = \log (tf_{ij} + 1) \times \log \left(\frac{N+1}{n_j} \right) \quad (4)$$

Keterangan:

t : *term* (kata)

d : dokmen atau kumpulan teks

a_{ij} : bobot *term* t_j terhadap dokumen d_i

tf_{ij} : jumlah kemunculan *term* t_j dalam dokumen d_i

N : jumlah total dokumen dalam dataset

n_j : jumlah dokumen yang mengandung *term* t_j

Proses pembobotan TF-IDF digunakan untuk mencari bobot tiap term/kata dalam sebuah dokumen yang diharapkan dapat menghasilkan daftar term/kata yang penting. Dengan cara menghitung bobot term sesuai rumus (4), proses pembobotan TF-IDF secara sederhana dapat dicontohkan menggunakan tweet data latih pada tabel 3.1, dari tabel tersebut dilakukanlah proses preprosesing dan hasilnya dapat dilihat pada tabel 3.15 sebagai berikut.

Tabel 2. Koleksi Data Latih

#	Tweet	Kategori
1	Suka memang luar biasa ayo Gassss terbanglah setinggi tingginya #dogecoin	Positif
2	Dogecoin Drop, Bitcoin & Ethereum Naik, Pesannya: Hati-hati!	Negatif
3	Doge coin naik lagi, Mana indodax nya udah di uninstall	Positif

Proses selanjutnya untuk menghitung bobot TF-IDF term pada tabel 2, dengan menggunakan persamaan rumus (4). Berikut contoh cara mendapatkan nilai bobot TF-IDF pada kata suka dan naik.

$$\text{suka} = \log(1 + 1) \times \log\left(\frac{3+1}{1}\right) = 0.90308998698$$

$$\text{naik} = \log(2 + 1) \times \log\left(\frac{3+1}{2}\right) = 0.77815125038$$

Setelah melakukan perhitungan, Berikut adalah hasil pembobotan menggunakan TF-IDF dapat dilihat pada tabel 3.

Tabel 3. Contoh hasil TF-IDF pada dataset term

#	Term	Log (TF + 1)	LOG IDF(N+1/N _j)	TF-IDF
1	Suka	0.30102999566	0.60205999132	0.90308998698
2	naik	0.47712125472	0.30102999566	0.77815125038

Dengan bobot TF-IDF tersebut setiap kata akan diranking dan hasilnya akan didapatkan bobot yang terbaik. Untuk keperluan klasifikasi data uji nantinya akan diambil bobot dengan TF-IDF yang tinggi.

c. *Naïve Bayes Classifier*

Naïve Bayes Classifier, merupakan salah satu metode yang digunakan dalam data mining yang didasarkan pada teori keputusan Bayes. Naive Bayes Classifier memiliki kemampuan klasifikasi seperti metode decision tree, neural network, k-nearestneighbourhood classifier, classification (IF-THEN) rule. Kelebihan dari Naïve Bayes Classifier adalah sederhana tetapi memiliki akurasi yang tinggi dan dapat menangani permasalahan banyak kelas [7].

$$P(V_j) = \left(\frac{docs_j}{training} \right) \quad (5)$$

$$P(W_k | V_j) = \left(\frac{n_k + 1}{n + |Kosakata|} \right) \quad (6)$$

Keterangan :

$docs_j$: Jumlah dokumen pada kategori j

$training$: Jumlah dokumen yang digunakan dalam proses data latih

W_k : Kata (*word*) k dalam semua dokumen yang diberi label sebagai j

V_j : Semua Kata (kosakata) dalam kelas j

n_k : Jumlah munculnya kata W_k dalam kelas V_j

n : Jumlah kosakata yang ada pada kategori V_j

Kosakata: Total kata unik dalam dokumen atau data latih

Pada proses klasifikasi Naïve Bayes Classifier setiap tweet akan dipresentasikan dalam pasangan atribut “X1, X2, X3, . . . Xn” yang mempunyai suatu makna jika X1 untuk kata pertama, X2 adalah kata kedua, dan seterusnya [8]. untuk himpunan kelas tweet dipresentasikan dengan V. Pada saat melakukan tahapan proses klasifikasi dokumen, Naïve Bayes Classifier akan menghasilkan suatu kategori/kelas dengan probabilitas tertinggi (VMAP) dari:

$$V_{MAP} = \underset{V_j \in V}{argmax} \prod_{i=1}^n P(x_k | V_j) P(V_j) \quad (2)$$

Keterangan :

VMAP : semua kategori yang diujikan

V_j : kategori tweet

$P(X_i | V_j)$: probabilitas X_i pada kategori V_j

$P(V_j)$: probabilitas dari V_j

$$\begin{aligned} & (tweet1 | V_{positif}) \\ &= P(X_{suka} | V_{positif}) \times P(X_{memang} | V_{positif}) \times P(X_{luar} | V_{positif}) \times P(X_{biasa} | V_{positif}) \\ & \times P(X_{ayo} | V_{positif}) \times P(X_{Gasss} | V_{positif}) \times P(X_{terbanglah} | V_{positif}) \times \\ & P(X_{setinggi} | V_{positif}) \times P(X_{tingginya} | V_{positif}) \times P(X_{dogecoin} | V_{positif}) \\ &= 0.05128 \times 1 \times 1 \times 0.02564 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \\ &= 0.0013609712 \end{aligned}$$

$$(tweet1 | V_{negatif})$$

$$\begin{aligned}
&= P(X_{\text{suka}} | V_{\text{negatif}}) \times P(X_{\text{memang}} | V_{\text{negatif}}) \times P(X_{\text{luar}} | V_{\text{negatif}}) \times P(X_{\text{biasa}} | V_{\text{negatif}}) \times \\
&P(X_{\text{ayo}} | V_{\text{negatif}}) \times P(X_{\text{Gassss}} | V_{\text{negatif}}) \times P(X_{\text{terbanglah}} | V_{\text{negatif}}) \times P(X_{\text{setinggi}} | V_{\text{negatif}}) \\
&\times P(X_{\text{tingginya}} | V_{\text{negatif}}) \times P(X_{\text{dogecoin}} | V_{\text{negatif}}) \\
&= 0.08571 \times 1 \times 1 \times 0.02857 \times 1 \times 0.05128 \times 1 \times 1 \times 1 \times 1 \\
&= 0.00012557111
\end{aligned}$$

$$\begin{aligned}
&(tweet1 | V_{\text{netral}}) \\
&= P(X_{\text{suka}} | V_{\text{netral}}) \times P(X_{\text{memang}} | V_{\text{netral}}) \times P(X_{\text{luar}} | V_{\text{netral}}) \times P(X_{\text{biasa}} | V_{\text{netral}}) \times \\
&P(X_{\text{ayo}} | V_{\text{netral}}) \times P(X_{\text{Gassss}} | V_{\text{netral}}) \times P(X_{\text{terbanglah}} | V_{\text{netral}}) \times P(X_{\text{setinggi}} | V_{\text{netral}}) \times \\
&P(X_{\text{tingginya}} | V_{\text{netral}}) \times P(X_{\text{dogecoin}} | V_{\text{netral}}) \\
&= 0.08571 \times 1 \times 1 \times 0.06122 \times 1 \times 0.02564 \times 1 \times 1 \times 1 \times 1 \\
&= 0.00013453734
\end{aligned}$$

Setelah menghitung probabilitas dari setiap data uji, diperoleh hasil probabilitas bahwa hasil dari data uji pertama yaitu tweet1 didapatkan nilai probabilitas positif lebih besar dari nilai probabilitas negatif dan netral sehingga dapat disimpulkan bahwa tweet1 termasuk kategori sentimen positif.

d. *Confusion Matrix*

Confusion Matrix digunakan untuk melakukan proses evaluasi performa sistem yang menguji hasil akhir klasifikasi dengan mengukur nilai performa dari suatu sistem yang telah dibuat atau dikembangkan. Parameter pengujian yang digunakan untuk evaluasi hasil dari klasifikasi yaitu akurasi yang perhitungannya diperoleh dari tabel *confusion matrix* [9].

Tabel 2. *Confusion Matrix*

Confusion Matrix		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	True Positive (TP)	False Positive (FP)	False Positive (FP)
	Negatif	False Negative (FNeg)	True Negative (TNeg)	False Negative (FNeg)
	Netral	False Neutral (FNeut)	False Neutral (FNeut)	True Neutral (TNeut)

Untuk menghitung akurasi prediksi kelas positif, negatif, dan netral menggunakan rumus persamaan (8) sebagai berikut:

$$\text{Akurasi} = \frac{TP + TNeg + TNeut}{TP + TNeg + TNeut + FP + FNeg + FNeut} \times 100\% \quad (3)$$

Data Flow Diagram (DFD)

a. Diagram Konteks

Diagram konteks merupakan diagram yang terdiri dari sebuah proses dan memvisualisasikan ruang lingkup dari suatu sistem. Diagram ini merupakan level tertinggi

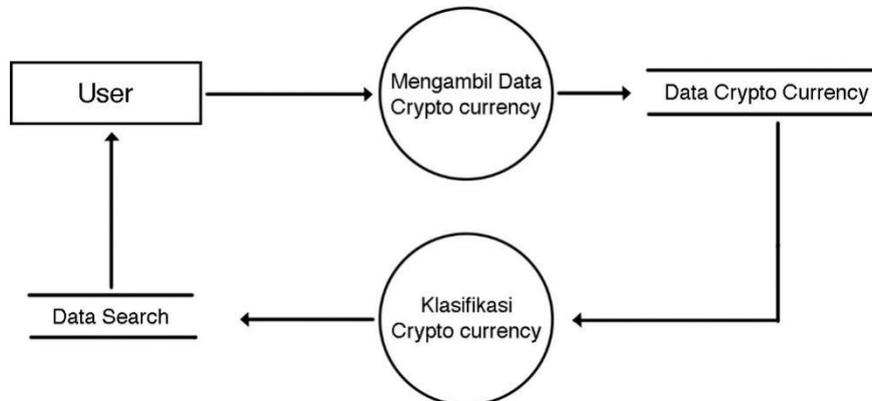
dari DFD yang menggambarkan seluruh input ke sistem atau output dari sistem yang memberikan gambaran tentang keseluruhan sistem. Diagram. Diagram konteks pada aplikasi Analisis Sentimen Cryptocurrency dapat dilihat pada Gambar 5.



Gambar 5. Diagram Konteks

b. DFD Level 1

Data Flow Diagram (DFD) merupakan alat pembuatan model yang memungkinkan untuk membentuk sebuah sistem profesional untuk menggambarkan sistem sebagai suatu jaringan proses fungsional yang dihubungkan satu dan yang lainnya dengan alur data, baik secara manual atau terkomputerisasi. Data Flow Diagram untuk aplikasi Analisis Sentimen Cryptocurrency dapat dilihat pada Gambar 6.



Gambar 6. DFD Level 1

DFD Level 1 diatas menjelaskan bahwa aplikasi klasifikasi mata uang kripto ini memiliki 2 data proses yang menjadi inti proses dalam aplikasi ini. Data crypto currency berisi jenis-jenis mata uang kripto yang sebelumnya telah disimpan kedalam database. Lalu dilakukan klasifikasi yang berfungsi untuk melakukan proses pengklasifikasian pada setiap komentar mata uang kripto yang dipilih oleh user, kemudian hasilnya disimpan pada database pada data search atau sentimen result.

HASIL DAN PEMBAHASAN

Analisis Kebutuhan

Dalam menentukan kebutuhan pengguna, dilakukan penyebaran kuesioner kepada pengguna mata uang digital pada grup Telegram dan didapatkan kesimpulan dari jawaban terhadap 100 responden yang terdiri dari 10 butir pertanyaan dan kesesuaian butir soal dengan indikator. Dalam menghitung hasil kuesioner digunakan metode skala likert dan didapatkan hasil kesimpulan mengenai kebutuhan pengguna yaitu:

- Pengguna membutuhkan aplikasi yang dapat mengumpulkan opini masyarakat terhadap mata uang kripto.

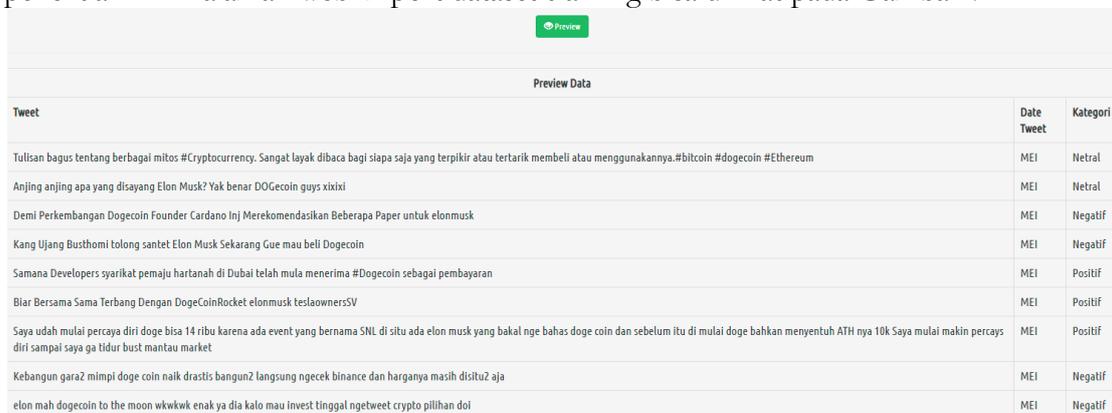
b. Pengguna membutuhkan aplikasi yang dapat mengklafisikasikan opini menjadi positif, negatif, atau netral.

Implementasi

Dalam implementasinya penelitian ini menggunakan platform web browser, dengan bahasa pemrograman PHP dan didesain menggunakan HTML, CSS dan Javascript, serta menggunakan MySQL sebagai sistem manajemen databasenya. Implementasi dimulai dengan tahap pengumpulan data, melakukan sentimen manual, menghitung bobot kata, hingga mendapatkan klasifikasi sentimen secara otomatis.

a. Tampilan *Web Import Dataset Training*

Pada halaman import dataset training menampilkan tabel untuk melihat data training yang telah di impor dengan cara menggunakan fitur impor data untuk melakukan impor data training serta pada browser akan mendownload format dokumen yang beformat csv. Peneliti menggunakan data sebanyak 478 tweet yang digunakan pada penelitian ini. Halaman web import dataset training bisa dilihat pada Gambar 7.

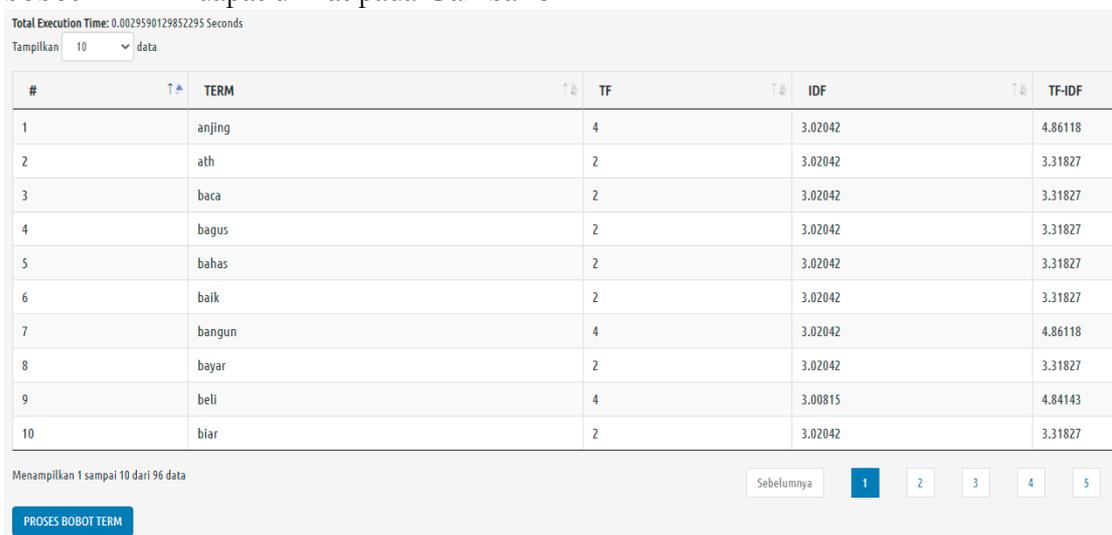


Tweet	Date Tweet	Kategori
Tulisan bagus tentang berbagai mitos #Cryptocurrency. Sangat layak dibaca bagi siapa saja yang terpikir atau tertarik membeli atau menggunakannya.#bitcoin #dogecoin #Ethereum	MEI	Netral
Anjing anjing apa yang disayang Elon Musk? Yak benar DOGecoin guys xixixi	MEI	Netral
Demi Perkembangan Dogecoin Founder Cardano Inj Merekomendasikan Beberapa Paper untuk elonmusk	MEI	Negatif
Kang Ujang Bushthomi tolong santet Elon Musk Sekarang Gue mau beli Dogecoin	MEI	Negatif
Samana Developers syarikat pemaju hartanah di Dubai telah mula menerima #Dogecoin sebagai pembayaran	MEI	Positif
Biar Bersama Sama Terbang Dengan DogeCoinRocket elonmusk teslaownersV	MEI	Positif
Saya udah mulai percaya diri doge bisa 14 ribu karena ada event yang bernama SNL di situ ada elon musk yang bakal nge bahas doge coin dan sebelum itu di mulai doge bahkan menyentuh ATH nya 10k Saya mulai makin percays diri sampai saya ga tidur bust mantau market	MEI	Positif
Kebangun gara2 mimpi doge coin naik drastis bangun2 langsung ngecek binance dan harganya masih disitu2 aja	MEI	Negatif
elon mah dogecoin to the moon wkwkwk enak ya dia kalo mau invest tinggal ngetweet crypto pilihan doi	MEI	Negatif

Gambar 7. Tampilan *Web Import Dataset Training*

b. Tampilan *Web Pembobotan TF-IDF*

Pada halaman ini terdapat tombol untuk memproses bobot term untuk melakukan proses pembobotan TF-IDF dan pada tampilan ini penyajian informasi berupa tabel bobot TF-IDF yang telah melalui proses pembobotan. Halaman ini menampilkan kata dari dataset training yang telah diberi bobot TF-IDF. Tampilan web bobot TF-IDF dapat dilihat pada Gambar 8.



#	TERM	TF	IDF	TF-IDF
1	anjing	4	3.02042	4.86118
2	ath	2	3.02042	3.31827
3	baca	2	3.02042	3.31827
4	bagus	2	3.02042	3.31827
5	bahas	2	3.02042	3.31827
6	baik	2	3.02042	3.31827
7	bangun	4	3.02042	4.86118
8	bayar	2	3.02042	3.31827
9	beli	4	3.00815	4.84143
10	biar	2	3.02042	3.31827

Gambar 8. Tampilan *Web* Pembobotan TF-IDFc. Tampilan *Web* Klasifikasi *Naïve Bayes*

Pada halaman ini terjadi proses klasifikasi naive bayes dengan langkah menekan tombol proses klasifikasi, halaman ini menyajikan informasi hasil klasifikasi sentimen data uji dan juga menampilkan total sentimen yang telah di klasifikasi. Tampilan halaman web klasifikasi sentimen dapat dilihat pada Gambar 9.

TWEET	PREPROSES	PROBABILITAS POSITIF	PROBABILITAS NEGATIF	PROBABILITAS NETRAL	KLASIFIKASI
RT @K_SugengWidodo: Gratis 2,7 Doge setiap hari untuk para pejuang regeh dengan cara menonton iklan - jawab pertanyaan - claim Minimal wit...	gratis doge juang regeh tonton iklan claim minimal wit	1.4367113549686e-17	9.3539522711588e-18	5.9525446812976e-17	neutral
RT @K_SugengWidodo: Gratis 2,7 Doge setiap hari untuk para pejuang regeh dengan cara menonton iklan - jawab pertanyaan - claim Minimal wit...	gratis doge juang regeh tonton iklan claim minimal wit	1.4367113549686e-17	9.3539522711588e-18	5.9525446812976e-17	neutral
RT @BackToTheDoge: @doge_memes sweet setup man #DogeCoin #Doge #dogearmyMEMEDiv https://t.co/mVLEF9Tnb	sweet setup man	0.00012398988338192	0.000095904081632653	0.00028404944606414	neutral
@AromaFinance Bisa yok @elonmusk @SpaceX @dogecoin #bsc #arfi #BNB #binance	yok	0.31778425655977	0.20408163265306	0.4198250728863	neutral
RT @BackToTheDoge: @doge_memes sweet setup man #DogeCoin #Doge #dogearmyMEMEDiv https://t.co/mVLEF9Tnb	sweet setup man	0.00012398988338192	0.000095904081632653	0.00028404944606414	neutral
#dogecoin #DogeFather #DogeCoinToTheMoon #DogeCoinArmy #HODL Dkyua1tqQYQ3aPWAKTpJzTL3xsjfbu https://t.co/dNtluonO26	dkyua tqatqyq apwaktjztl xsjfbu	0.31778425655977	0.20408163265306	0.4198250728863	neutral
@DogeCoinMemes @dogecoin Hi Bro	hi bro	0.000000096754265598251	0.000000090135451122449	0.000000096093927603499	positive

Gambar 9. Tampilan *Web* Klasifikasi *Naïve Bayes***Blackbox Testing**

Pengujian aplikasi ini menggunakan metode *black box* dimana pengujian didasarkan pada fungsionalitas yakni untuk melihat apakah program berhasil dijalankan tanpa kesalahan. Metode ini melakukan pengujian yang berfokus pada spesifikasi fungsional tanpa menguji desain dan program. Pengujian ini dilakukan dengan membuat kasus uji seperti memeriksa atau mencoba semua fungsi dengan memakai perangkat lunak apakah sudah sesuai dengan spesifikasi yang dibutuhkan. Berikut merupakan hasil pengujian dapat dilihat pada Tabel 3.

Tabel 3. Tabel Pengujian

No	Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Kesimpulan
1	Menampilkan hasil tweet yang berkaitan dengan <i>Cryptocurrency</i> .	Menekan button Proses Crawling Data pada menu Dataset Testing.	Sistem akan menampilkan daftar seluruh tweet terbaru yang berkaitan dengan <i>Cryptocurrency</i> .	Valid

No	Skenario Pengujian	Kasus Pengujian	Hasil Pengujian	Kesimpulan
2	Menambahkan data tweet secara manual dan diproses sebagai data latih.	Menekan button Import Data pada menu Data Training	Sistem akan menampilkan data yang telah di import dan disimpan sebagai data latih.	Valid
3	Menampilkan informasi berupa data tabel dan grafik sebagai informasi kesimpulan klasifikasi sentimen	Menekan button Proses Klasifikasi pada menu Klasifikasi Data Testing	Sistem akan menampilkan data tabel dan grafik informasi kesimpulan klasifikasi sentimen (positif, negatif, netral)	Valid

Confusion Matrix

Pengujian dilakukan dengan memberi berbagai macam batas minimum berdasarkan nilai bobot terms antara lain 0 dan 5. Hasil prediksi klasifikasi, dihitung akurasi secara manual dengan mengimplementasikan persamaan (8). Berikut adalah hasil perhitungan manual confusion matrix berdasarkan masing-masing batas minimum bobot terms.

Tabel 4. *Confusion Matrix* bobot minimum 0

Confusion Matrix		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	10	8	8
	Negatif	4	0	4

	Netral	34	34	22
--	--------	----	----	----

$$Akurasi = \frac{10 + 0 + 22}{10 + 0 + 22 + 8 + 4 + 34} \times 100\% = \mathbf{41.0256\%}$$

Hasil akurasi dari bobot minimum 0 adalah 41.0256 %

Tabel 5. *Confusion Matrix* bobot minimum 5

Confusion Matrix		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	14	4	4
	Negatif	4	0	4
	Netral	36	36	20

$$Akurasi = \frac{14 + 0 + 20}{14 + 0 + 20 + 4 + 4 + 36} \times 100\% = \mathbf{43.9857\%}$$

Hasil akurasi dari bobot minimum 5 adalah 43.5897 %

KESIMPULAN

Kesimpulan dari penelitian ini adalah telah berhasil dilakukan rancang bangun aplikasi analisis sentimen mata uang kripto dengan menggunakan algoritma naïve bayes. Aplikasi ini memiliki menu data training yang di dalamnya terdapat fitur impor data dalam format CSV, dataset training, pembobotan term dan naïve bayes data training, serta menu data testing yang terdapat fitur dataset testing, preprocessing data testing dan klasifikasi data testing. Aplikasi ini berbasis web dengan menggunakan PHP, framework native serta menggunakan database MySQL. Pengujian dilakukan menggunakan blackbox testing dan pengujian akurasi menggunakan confusion matrix berdasarkan data tweet yang dijadikan data testing, metode naïve bayes berhasil mengklasifikasi 57 dari 73 data yang diuji sehingga metode Naïve Bayes berhasil memprediksi tingkat akurasi opini sebesar 78,623%.

SARAN

1. Menambahkan koleksi kamus tipografi (typo), kata slang ataupun kata gaul
2. Pada penelitian selanjutnya peneliti menyarankan agar mengganti metode pembobotan TF-IDF dengan metode TF-IDF-CF sebagai peningkatan akurasi perhitungan.
3. Pada penelitian selanjutnya peneliti menyarankan untuk menambah metode Adaboost dan Information Gain yang berguna untuk meningkatkan akurasi klasifikasi.

DAFTAR PUSTAKA

- [1] E. Retnawiyati, Fatoni, and E. S. Negara, "Analisis Sentimen Pada Data Twitter dengan Menggunakan Text Mining Terhadap Suatu Produk," *Indones. Journal Comput. Cybern. Syst.*, 2015.
- [2] Febriansyah, "Sentimen Analisis Twitter Berbahasa Indonesia," pp. 1–59, 2016.
- [3] S. W. U. Vitandy, A. A. Supianto, and F. A. Bachtiar, "Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency- Inverse Document Frequency dan Naïve Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, 2019.
- [4] A. Syakuro, "PADA MEDIA SOSIAL MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC) DENGAN SELEKSI FITUR INFORMATION GAIN (IG) HALAMAN JUDUL SKRIPSI Oleh : ABDAN SYAKURO," *Anal. sentimen Masy. terhadap e-commerce pada media Sos. menggunakan Metod. naive bayes Classif. dengan Sel. fitur Inf. gain*, 2017.
- [5] I. F. Rozi, E. N. Hamdana, M. Balya, and I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter (Studi Kasus Samsat Kota Malang)," *Inform. Polinema*, 2018.
- [6] M. Liu and J. Yang, "An improvement of TFIDF weighting in text categorization," in *International Conference on Computer Technology and Science (ICCTS)*, 2012.
- [7] S. Fajar Rodiyansyah and E. Winarko, "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, 2013.
- [8] F. AFSHOH, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter," *Inform. Progr. Stud. Komunikasi, Fak. Inform. D A N Surakarta, Univ. Muhammadiyah*, 2017.
- [9] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeding Eng.*, 2015.