

Sentiment Analysis of Cryptocurrency Exchange Application on Twitter Using Naïve Bayes Classifier Method

Analisis Sentimen Aplikasi Bursa Perdagangan Aset Kripto di Twitter dengan Menggunakan Metode Naïve Bayes Classifier

Andhika Octa Indarso^{1*}, Helena Nurramdhani Irmanda², Ria Astriratma³

^{1,2,3} Sistem Informasi, Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

^{1*}andyocta@upnvj.ac.id, ²helenairmanda@upnvj.ac.id, ³astriratma@upnvj.ac.id

Informasi Artikel

Received: December 2022

Revised: January 2023

Accepted: January 2023

Published: February 2023

Abstract

Purpose: The growth and development of the digital currency industry also presents a variety of applications for conducting transactions using these currencies, including utilizing cryptocurrency exchanges to make investments. In Indonesia, there are two applications that fall into the category of the largest cryptocurrency exchange and are recognized by Bappebti (Commodity Futures Trading Regulatory Agency), namely TokoCrypto and Indodax. Both applications are analyzed based on the sentiments of their users on Twitter.

Design/methodology/approach: In this study the data collected is data originating from social media Twitter and has the keywords "indodax" or "#indodax" and "tokocrypto" or "#tokocrypto". The data used is between January 2021 – January 2022. The data collected from Twitter is processed using the Naïve Bayes Classifier algorithm.

Findings/result: From the results of the analysis, it was found that the Indodax application has a higher positive sentiment percentage value of 9% compared to TokoCrypto.

Originality/value/state of the art: The use of the Naïve Bayes algorithm in this study supports sentiment analysis of cryptocurrency exchange application users to consider which application has better positive sentiment for investing in digital currency or cryptocurrency.

Keywords: naïve bayes, twitter, sentiment analysis

Kata kunci: naïve bayes, twitter, analisis sentimen

Abstrak

Tujuan: Tumbuh dan berkembangnya industri mata uang digital turut menghadirkan beragam aplikasi untuk melakukan transaksi menggunakan mata uang tersebut termasuk memanfaatkan cryptocurrency exchange untuk melakukan investasi. Di Indonesia, ada dua aplikasi yang masuk ke dalam kategori cryptocurrency exchange terbesar

dan diakui oleh Bappebti (Badan Pengawas Perdagangan Berjangka Komoditi), yakni TokoCrypto dan Indodax. Kedua aplikasi tersebut dianalisa berdasarkan sentimen dari para penggunanya yang ada di Twitter.

Perancangan/metode/pendekatan: Dalam penelitian ini data yang dikumpulkan ialah data yang berasal dari sosial media Twitter dan memiliki kata kunci “indodax” atau “#indodax” dan “tokocrypto” atau “#tokocrypto”. Data yang digunakan berada pada rentang waktu antara Januari 2021 – Januari 2022. Data yang berhasil dikumpulkan dari Twitter diolah menggunakan algoritma Naïve Bayes Classifier.

Hasil: Dari hasil analisa, didapatkan bahwa aplikasi Indodax memiliki nilai persentase sentimen positif lebih tinggi 9% dibandingkan TokoCrypto.

Keaslian/ state of the art: Penggunaan algoritma Naïve Bayes di dalam penelitian ini mendukung analisis sentimen pengguna aplikasi cryptocurrency exchange untuk memilih aplikasi mana yang memiliki sentimen positif lebih baik dalam investasi mata uang digital atau cryptocurrency.

1. Pendahuluan

Sejarah kehadiran *cryptocurrency* atau mata uang digital dimulai pada tahun 2008 yang dirancang oleh seseorang yang menggunakan nama samaran, Satoshi Nakamoto melalui terobosan teknologi peer-to-peer transaksi elektronik [1] yang menghilangkan peran bank sentral ataupun administrator di dalamnya saat melakukan transaksi dan mata uang tersebut diberi nama Bitcoin. Bitcoin merupakan satu dari sekian ribu jenis mata uang digital yang saat ini beredar di pasar *cryptocurrency* [15]. Mata uang digital yang dibangun oleh Satoshi Nakamoto ini menarik minat para investor baik dari kalangan kelas menengah maupun atas. Berdasarkan data yang dirujuk dari coinmarketcap.com hingga Desember 2021 terdapat 7.812 mata uang digital yang diperdagangkan secara global. Dari sekian banyak mata uang digital yang diperdagangkan, terdapat 5 (lima) mata uang digital terpopuler dengan volume perdagangan yang tinggi, seperti Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), Tether (USDT), dan Cardano (ADA).

Di beberapa negara, mulai dari Amerika Serikat, Kanada, Australia, Inggris hingga negara-negara Uni Eropa, seperti Finlandia, Jerman, Bulgaria, Belgia, dan Siprus melegalkan mata uang digital tersebut sebagai komoditi untuk diperdagangkan akan tetapi bukan sebagai alat pembayaran yang sah. [2] Begitu pula di Indonesia, mata uang digital atau disebut juga sebagai aset kripto belum dapat digunakan sebagai alat pembayaran yang sah menurut Bappebti (Badan Pengawas Perdagangan Berjangka Komoditi) melainkan dapat dimanfaatkan dalam bentuk investasi sebagai komoditi [3].

Aset kripto atau Crypto Asset menurut Bappebti merupakan istilah yang digunakan untuk menyebutkan komoditi tidak berwujud yang berbentuk digital aset, menggunakan kriptografi, jaringan *peer-to-peer*, dan buku besar yang terdistribusi, untuk mengatur penciptaan unit baru, memverifikasi transaksi, dan mengamankan transaksi tanpa campur tangan pihak lain. Di

Indonesia, tahun 2022, aset kripto yang sudah terdaftar dan legal untuk diperdagangkan mencapai 383 aset kripto [4].

Kendati mata uang digital atau aset kripto belum menjadi instrumen investasi yang populer di Indonesia, namun berdasarkan data yang dikeluarkan oleh Bappebti terjadi peningkatan jumlah pedagang aset kripto dari tahun 2021 yang hanya ada 11 (sebelas) pedagang, meningkat pada Agustus tahun 2022 tercatat sudah ada 25 (dua puluh lima) perusahaan atau pedagang aset kripto yang memiliki ijin untuk melakukan perdagangan aset kripto, seperti PT Aset Digital Berkat (Tokocrypto), PT Indodax Nasional Indonesia (Indodax), PT Gudang Kripto Indonesia (Gudang Kripto), PT Pintu Kemana Saja (Pintu), PT Tiga Inti Utama (TRIV), dan masih banyak lagi lainnya.

Cryptocurrency Exchange atau bursa perdagangan aset kripto merupakan sarana elektronik untuk melakukan transaksi perdagangan aset kripto oleh penjual atau pembeli aset kripto melalui *website* ataupun aplikasi pada perangkat *mobile*. Aplikasi Tokocrypto maupun Indodax merupakan aplikasi yang populer digunakan sebagai media pertukaran atau perdagangan aset kripto. Kedua aplikasi tersebut tentu saja memiliki tempat di hati para investor dengan kelebihan dan kekurangan masing-masing aplikasi.

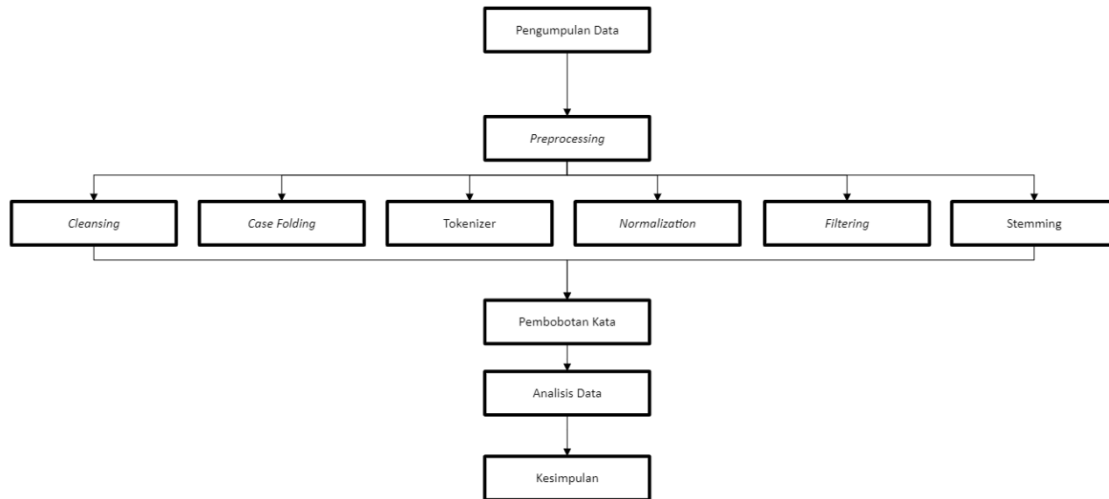
Kepuasan pelanggan hal yang penting untuk dilakukan dalam mengukur tingkat keberhasilan suatu pelayanan baik dalam bentuk produk atau jasa. Tingkat kepuasan pelanggan menjadi tolak ukur bagi perusahaan untuk mengevaluasi kembali terhadap layanan, jasa atau produk yang diberikan ke pelanggan [5]. Bila pelanggan merasa puas, cenderung kembali lagi menggunakan layanan, jasa atau produk yang sama dan mengajak individu lainnya untuk menggunakan hal-hal tersebut. Bahkan mereka mempromosikan hal tersebut via opini di media sosial. Begitu pula sebaliknya bila pelanggan tidak puas, maka mereka memberikan opini negatif di media sosial.

Opinion mining sering kali digunakan untuk menilai suatu produk, individu, isu-isu tertentu, peristiwa, maupun layanan. Pada dasarnya *opinion mining* merupakan bidang keilmuan untuk menganalisis berupa sentimen, pendapat, penilaian, emosi maupun sikap seseorang. Adapun opini yang muncul di tengah-tengah masyarakat bisa dalam bentuk sentimen negatif ataupun positif. Sedangkan analisis sentimen merupakan metode untuk menggali opini dari sebuah teks yang memanfaatkan metode komputasi berupa *machine learning* maupun statistik. Adapun hal-hal yang dianalisis di dalam teks berupa kategori sentimen, subjektivitas, sikap, karakter, emosi, dan lain-lain [6].

Awalnya Twitter merupakan aplikasi sosial media atau *micro blogging* yang memiliki keunikan dalam mengirim dan membaca pesan hingga 140 karakter. Pada tahun 2017, Twitter telah menambah jumlah karakter yang dapat dikirimkan menjadi 280 karakter yang disebut sebagai cuitan atau *tweets*. Jack Dorsey adalah orang dibalik pembuat Twitter yang diluncurkan pertama kali pada tahun 2006 [11]. Para pengguna bebas untuk mengutarakan pendapatnya pada aplikasi Twitter, sehingga opini terhadap suatu produk atau jasa bisa langsung dibaca oleh pengguna aplikasi ini. Hal ini menjadi media yang tepat di dalam pengumpulan opini baik positif maupun negatif khususnya terhadap aplikasi perdagangan aset kripto.

2. Metode/Perancangan

Penelitian ini merupakan jenis penelitian kuantitatif dengan studi kasus mengenai analisis sentimen pengguna aplikasi Indodax dan Tokocrypto. Adapun tahapan-tahapan di dalam penelitian sesuai dengan **Gambar 1**.



Gambar 1. Tahapan Alur Penelitian

Tahapan awal dari alur penelitian ini adalah Pengumpulan Data. Pengumpulan data yang diambil secara langsung dengan hashtag pencarian dengan kata kunci TokoCrypto dan Indodax. Agar dapat melakukan *scraping tweets* menggunakan *Application Program Interface (API)* khusus yang disediakan oleh Twitter [7]. Setelah mendapatkan akses *API key* maka dapat dicari datanya dengan menggunakan *software* untuk *text mining*. Penelitian ini menggunakan *software* RapidMiner Studio untuk melakukan *scraping tweets*, serta melakukan pengolahan data menggunakan algoritma Naïve Bayes Classifier. Dalam penelitian, data dibagi menjadi 2 (dua) kategori komentar yaitu komentar positif, dan negatif yang diberikan oleh pengguna Twitter dalam bahasa Indonesia.

Tahap kedua adalah *Preprocessing*. Pada tahap ini data teks yang sudah ada diubah menjadi analisis sentimen. Dengan menggunakan metode Naïve Bayes Classifier akan dilakukan menggunakan 3 data, 2 data latih dan 1 data uji. Dari 2 data latih tersebut akan menunjukkan klasifikasi positif dan negatif. Ada beberapa tahapan yang akan dilakukan dalam *Preprocessing*, seperti *Cleansing* dimana data akan dibersihkan dari hal-hal atau komponen yang tidak memiliki hubungan pada informasi yang ada di dokumen. Kemudian *Case Folding* dimana data menjadi seragam dengan menggunakan huruf kecil (*lower case*). *Tokenizer* digunakan untuk memisahkan teks menjadi beberapa token. *Normalization* digunakan untuk mengubah kata yang non baku menjadi kata baku. Selain itu ada tahap *Filtering* untuk membuang kata yang tidak penting yang berada pada tahapan tokenizing. Terakhir adalah *Stemming* mengembalikan kata yang memiliki kata sambung menjadi kata dasar.

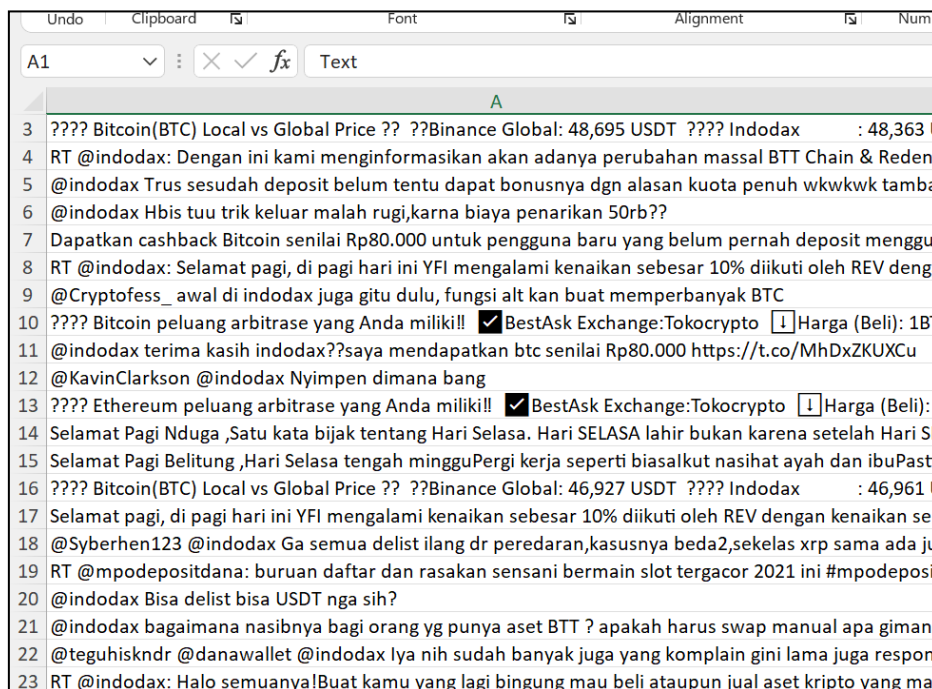
Tahap berikutnya adalah melakukan Pembobotan Kata. Pada tahap ini dilakukan proses mengubah kata menjadi angka atau vektor sedangkan TF (*Term Frequency*) melihat frekuensi kata yang muncul di dalam dokumen pencarian.

Pada tahap Analisis Data menggunakan data latih dan data uji untuk melihat opini-opini sentimen pengguna aplikasi *cryptocurrency exchange*.

Tahapan terakhir adalah Kesimpulan. Pada tahap ini dilakukan pendokumentasian keseluruhan hasil penelitian dan penarikan kesimpulan serta saran yang dapat diberikan untuk penelitian selanjutnya.

2.1. Pengumpulan Data

Dalam penelitian ini data yang dikumpulkan ialah data yang berasal dari sosial media Twitter dan memiliki kata kunci “indodax” atau “#indodax” dan “tokocrypto” atau “#tokocrypto”. Data yang digunakan berada pada rentang waktu antara Januari 2021 – Januari 2022. Jumlah data yang berhasil dikumpulkan dengan menggunakan teknik scrapping dengan kata kunci “indodax” sebanyak 1200 tweets dan kata kunci “tokocrypto” sebanyak 1100 tweets. Dari data yang berhasil dikumpulkan berikutnya akan dilakukan adalah melakukan pemilahan data mana yang merupakan cuitan iklan dan data bukan iklan yang berasal dari tweet pengguna kedua aplikasi *crypto exchange* tersebut. Tweet yang memiliki tujuan komersial tidak akan diikutsertakan dalam penelitian ini seperti yang terlihat pada **Gambar 2**.



Gambar 2. Contoh data *scrapping* yang berasal dari Twitter dengan kata kunci “indodax”

2.2. Preprocessing

Setelah data dipilah, maka terkumpul data baru yang bisa digunakan untuk melakukan penelitian ini. Data dengan kata “indodax” sebanyak 950 *tweets* dibagi menjadi 80% data latih dan 20% data uji sehingga didapatkan sebanyak 760 data latih dan 190 data uji. Data latih dibagi menjadi 380 *tweets* untuk kata “indodax” yang bermakna positif serta 380 *tweets* yang memiliki kata negatif.

Begitupula dengan kata “tokocrypto” menggunakan data dari 950 *tweets* yang sudah dipilah untuk menghindari bentuk komersil atau iklan. Hasil *tweets* tersebut akan dibagi menjadi 80% data latih dan 20% data uji yang menghasilkan sebanyak 760 *tweets* sebagai data latih dan sisanya 190 *tweets* untuk data uji. Data latih tersebut dibagi menjadi 380 *tweets* kata positif dan 380 kata negatif. Langkah berikutnya yang dilakukan ialah melakukan *text processing* yang merupakan bagian penting dalam melakukan analisis sentimen, yakni dengan melalui serangkaian proses, seperti *cleansing*, *case folding*, *tokenizer*, normalisasi, *filtering*, dan *stemming*.

2.2.1. Cleansing

Proses yang ada di tahapan ini ialah menghilangkan hal-hal atau komponen yang tidak ada hubungannya dengan data, seperti URL, hashtag (#), tanda titik (.), tanda koma (,), tanda titik dua (:), serta penggunaan tanda baca lainnya yang dapat mengganggu hasil analisis pada data. Sebagai contoh kalimat yang akan melalui tahapan ini:

Tweet	Proses	Output
Apliksi Kripto di Indodax, Tmpilan Antrmukanya Gmpng, dan Koinnya Bnyk	<i>Cleansing</i>	Apliksi Kripto di Indodax Tmpilan Antrmukanya Gmpng dan Koinnya Bnyk

2.2.2. Case Folding

Tahapan berikutnya adalah melakukan *case folding* dimana kalimat yang sudah melalui proses *cleansing* akan dilakukan penyeragaman kalimat dengan menggunakan huruf kecil (*lowercase*).

Tweet	Proses	Output
Apliksi Kripto di Indodax Tmpilan Antrmukanya Gmpng dan Koinnya Bnyk	<i>Case Folding</i>	apliksi kripto di indodax tmpilan antrmukanya gmpng dan koinnya bnyk

2.2.3. Tokenizer

Pada tahapan ini data teks yang sudah ada akan dipisahkan menjadi beberapa token ataupun berupa potongan huruf, kata, atau kalimat, sebelum dianalisis lebih lanjut. Bagi entitas yang disebut sebagai token ialah kata, angka, simbol, tanda baca, dan sebagainya.

Tweet	Proses	Output
apliksi kripto di indodax tmpilan antrmukanya gmpng dan koinnya bnyk	<i>Tokenizer</i>	apliksi kripto di indodax tmpilan antrmukanya

		gmpng dan koinnya bnyk
--	--	------------------------------

2.2.4. Normalisasi

Tahapan keempat atau Normalisasi ini dilakukan perubahan pada kata yang tidak baku atau tidak sesuai ejaan yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) menjadi sebuah kata yang baku atau sesuai ejaan KBBI.

Tweet	Proses	Output
aplikasi kripto di indodax tampilan antrmukanya gmpng dan koinnya bnyk	Normalisasi	aplikasi kripto di indodax tampilan antarmukanya gampang dan koinnya banyak

2.2.5. Filtering

Selanjutnya adalah tahapan *filtering* dimana kata-kata yang berjumlah banyak namun tidak memiliki arti (*stopwords*) atau kumpulan kata yang banyak digunakan dalam berbagai bahasa dan berdasarkan jenis kata yang akan di *stopwords*, mulai dari kata seru, kata depan, kata penghubung, kata ganti, dan kata keterangan.

Tweet	Proses	Output
aplikasi kripto di indodax tampilan antrmukanya gmpng dan koinnya bnyk	<i>Filtering</i>	aplikasi kripto indodax tampilan antarmukanya gampang koinnya banyak

2.2.6. Stemming

Proses terakhir dalam *preprocessing* yang dilakukan ialah *stemming* dimana kata dasar menurut kaidah bahasa Indonesia yang benar.

Tweet	Proses	Output
aplikasi	Stemming	aplikasi
kripto		kripto
indodax		indodax
tampilan		tampilan
antarmukanya		antarmuka
gampang		gampang
koinnya		koin
banyak		banyak

2.3. Pembobotan Kata

Suatu kata dalam dokumen teks [8]. Salah satu metode yang populer untuk pembobotan kata adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). *Term Frequency-Inverse Document Frequency* merupakan metode pembobotan yang menggabungkan dua konsep yaitu *Term Frequency* dan *Document Frequency*. *Term Frequency* adalah konsep pembobotan dengan mencari seberapa sering (frekuensi) kemunculan suatu term dalam satu dokumen. Karena setiap dokumen memiliki panjang yang berbeda-beda, maka bisa saja sebuah kata muncul lebih banyak pada dokumen yang panjang dibandingkan dengan dokumen yang pendek. Dengan demikian, *term frequency* sering dibagi dengan panjang dokumen (jumlah kata dalam dokumen).

Sedangkan *Document Frequency* adalah jumlah dokumen dimana suatu term muncul. Semakin kecil frekuensi kemunculannya, semakin kecil nilai bobotnya. Saat menghitung *term frequency*, semua kata di dalamnya dianggap penting. Namun, ada kata-kata yang sebenarnya kurang penting dan tidak perlu diperhitungkan seperti “di-”, “ke-”, “dan”, dll. Oleh karena itu, kata-kata yang kurang penting ini perlu dikurangi bobotnya dan menambah bobotnya pada kata-kata penting lainnya. Ini adalah ide dasar mengapa *stopword* diperlukan. Oleh karena itu diperlukan perhitungan TFIDF, agar skor dapat diperoleh dengan menggunakan persamaan (1):

$$tf - idf_{t,d} = tf_{t,d} * idf_t \tag{1}$$

2.4. Naïve Bayes

Naïve Bayes merupakan *machine learning* yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan Bayes [9][10]. Penggunaan teorema Bayes pada algoritma Naïve Bayes adalah dengan menggabungkan *prior probability* dan *conditional probability* dalam suatu rumus yang dapat digunakan untuk menghitung probabilitas dari setiap kemungkinan klasifikasi.

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \tag{2}$$

Setelah proses klasifikasi analisis sentimen selesai dilakukan, diperlukan satu langkah lagi untuk menentukan kualitas proses yang telah dilakukan yaitu mengevaluasi hasil. Pada tahap ini kinerja dari perhitungan yang telah dilakukan akan diuji dengan parameter akurasi. Akurasi (A) adalah jumlah dokumen yang diklasifikasikan dengan benar, baik *True Positive* maupun *True Negative*. Menghitung nilai akurasi dapat menggunakan persamaan (3):

$$A = \frac{(TP + TN)}{(TP + FP + TN + FN)} \times 100\% \quad (3)$$

Variabel seperti TP, TN, FP, dan FN berasal dari *confusion matrix*. TN adalah singkatan dari *True Negative*, data negatif tergolong negatif. TP adalah singkatan dari *True Positive*, data positif tergolong positif. FN adalah singkatan dari *False Negative*, data positif tergolong negatif. FP adalah singkatan dari *False Positive*, data negatif tergolong positif. **Tabel 1** merupakan penjelasan dari *confusion matrix* [12].

Tabel 1. Penjelasan *Confusion Matrix*

	Prediction Yes	Prediction No
True Yes	TP	FN
True No	FP	TN

3. Hasil dan Pembahasan

Setelah pada bagian *preprocessing* selesai, maka selanjutnya adalah menganalisis data yang sudah ada dan dilakukan pengujian dengan menggunakan model validasi K-Fold Cross. Cross validation merupakan sebuah metode tambahan dari teknik *data mining* yang bertujuan untuk memperoleh hasil akurasi yang maksimal. Metode ini sering juga disebut dengan k-fold cross validation dimana percobaan sebanyak k kali untuk satu model dengan parameter yang sama. Pada pengujian terhadap data yang ada, akan menggunakan 10 lipatan atau 10-fold cross validation dimana nantinya data akan dibagi menjadi 10 bagian seperti yang terlihat pada **Tabel 2**.

Tabel 2. Nilai K-Fold menggunakan 10-Fold Cross

Nilai K-Fold	Dataset									
1	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
2	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
3	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
4	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
5	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
6	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
7	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
8	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
9	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
10	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

3.1. Pengujian Data Indodax

Pada pengujian yang pertama, menggunakan data milik Indodax. Data yang akan digunakan sebanyak 950 *tweet* termasuk data uji yang nantinya akan dibagi menjadi 10 bagian atau k=10 sehingga data yang diperoleh adalah 950 data dibagi menjadi 10 lipatan dengan masing-masing

isinya adalah 95 data. Selain itu akan ditentukan mana yang termasuk data training dan mana yang termasuk data testing dengan perbandingan 80-20, didapatkan 760 data sebagai data latih dan 190 sisanya sebagai data uji. Pengujian menggunakan data yang sudah dipartisi akan diulang sebanyak 10 kali ($k=10$) dengan posisi data tes berbeda disetiap iterasinya. Misalkan iterasi pertama data tes pada posisi awal, iterasi kedua data tes di posisi kedua begitu seterusnya. Didapatkan hasilnya seperti yang ditampilkan pada **Tabel 3** berikut.

Tabel 3. Hasil K-Fold Cross dengan menggunakan data Indodax

K-Fold Cross	Accuracy	Precision	Recall
2	0,637	0,809	0,590
3	0,700	0,768	0,755
4	0,733	0,765	0,815
5	0,720	0,711	0,839
6	0,783	0,775	0,880
7	0,770	0,743	0,912
8	0,795	0,771	0,914
9	0,782	0,740	0,944
10	0,750	0,713	0,930

Dari tabel tersebut, terlihat hasil dari pengujian terhadap data tes terlihat tingkat akurasi tertinggi pada fold ke-8 sebesar 79%, kemudian *fold* ke-6 sebesar 78%, dan *fold* ke-10 dengan tingkat akurasi sebesar 75%. Sedangkan tingkat akurasi terendah sebesar 63% terjadi pada *fold* ke-2.

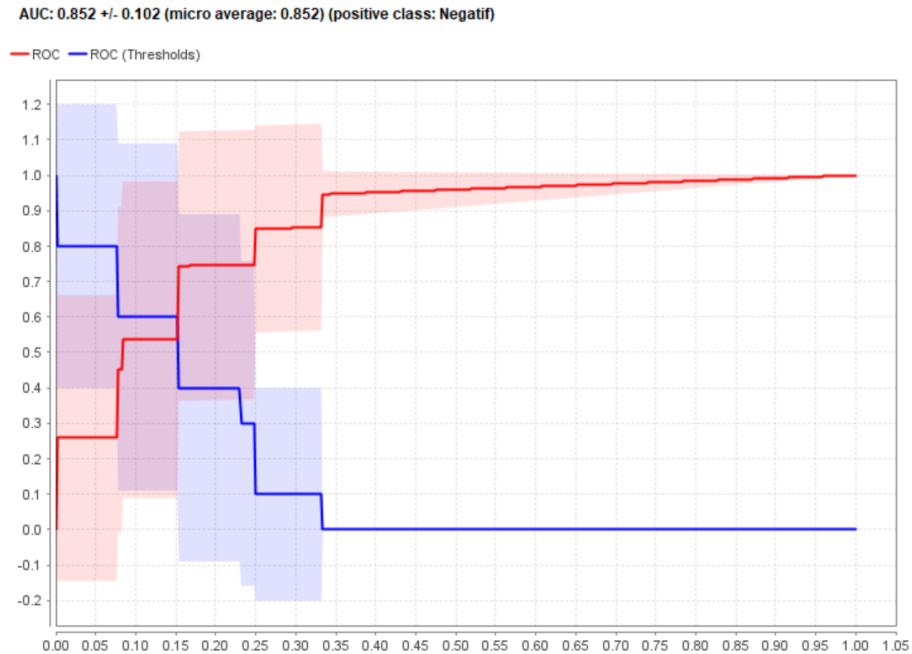
Sedangkan pada tingkat *Precision*, terlihat tingkat akurasi tertinggi terjadi pada *fold* ke-2 sebesar 80%, kemudian terjadi pada *fold* ke-6 sebesar 77%, dan *fold* ke-3 sebesar 76%. Tingkat *precision* terendah terjadi pada *fold* ke-5 sebesar 71,1%.

Recall tertinggi terjadi pada *fold* ke-9 sebesar 94%, kemudian *fold* ke-10 sebesar 93%, dan *fold* ke-8 sebesar 91%. Tingkat *recall* terendah terjadi pada *fold* ke-2 sebesar 59%.

Setelah melakukan pengujian terhadap dataset yang ada, maka langkah selanjutnya adalah menguji secara cross validation terhadap dataset menggunakan algoritma Naïve Bayes dengan menggunakan Rapidminer dengan data latih sebanyak 760 tweet dan sisanya sebagai data uji. Didapat hasil pengujian untuk accuracy, precision, dan recall mendapatkan nilai sebagai berikut: Accuracy: 85.46% ; Precision: 89.60%; Recall: 81.99%

Hasil nilai akurasi yang dihasilkan dari dataset tersebut mencapai 85,46% yang menunjukkan data tersebut memiliki nilai yang baik dan memiliki kecenderungan positif yang lebih besar daripada negatif. Hasil nilai presisi yang dihasilkan data tersebut mencapai 89.60% yang terbilang sangat akurat dalam menunjukkan data yang positif. Sedangkan hasil Recall menunjukkan hasilnya yang baik, yakni 81.99%, dimana hasil tersebut menunjukkan data rasio prediksi benar positif yang tinggi terhadap data keseluruhan yang benar positif.

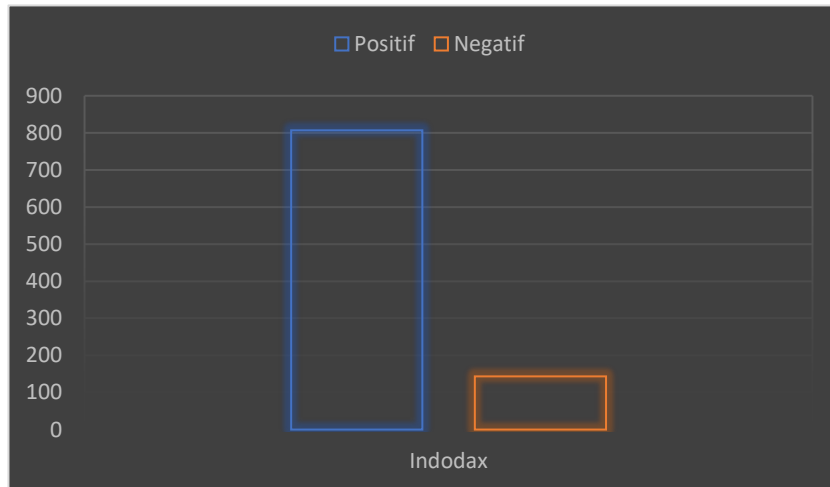
Nilai ROC dari dataset Indodax menghasilkan nilai seperti berikut ini.



Gambar 3. Nilai ROC dataset Indodax

Nilai AUC yang dihasilkan adalah 0.852 yang mendekati nilai 1.0. Artinya dataset ini memiliki nilai uji diagnostik yang baik karena nilai $AUC > 0.5$ [16].

Setelah data diolah, terlihat bahwa jumlah Tweet yang menunjukkan sentimen positif terhadap aplikasi Indodax yang ada di Twitter terdapat 807 cuitan, sedangkan 143 menunjukkan sentimen negatif terhadap Indodax.



Gambar 4. Jumlah cuitan positif dan negatif terhadap Indodax

3.2. Pengujian Data Tokocrypto

Di pengujian yang kedua, menggunakan data milik Tokocrypto. Data yang akan digunakan jumlahnya sama dengan sebelumnya, yakni sebanyak 950 tweet termasuk data uji yang nantinya akan dibagi menjadi 10 bagian atau $k=10$ sehingga data yang diperoleh adalah 950 data dibagi

menjadi 10 lipatan dengan masing-masing isinya adalah 95 data. Didapatkan hasilnya seperti yang ditampilkan pada **Tabel 4** berikut.

Tabel 4. Hasil K-Fold Cross dengan menggunakan data Tokocrypto

K-Fold Cross	Accuracy	Precision	Recall
2	0,616	0,775	0,641
3	0,630	0,886	0,435
4	0,668	0,908	0,483
5	0,655	0,836	0,591
6	0,674	0,944	0,435
7	0,694	1	0,389
8	0,726	0,884	0,637
9	0,695	0,844	0,642
10	0,707	0,890	0,600

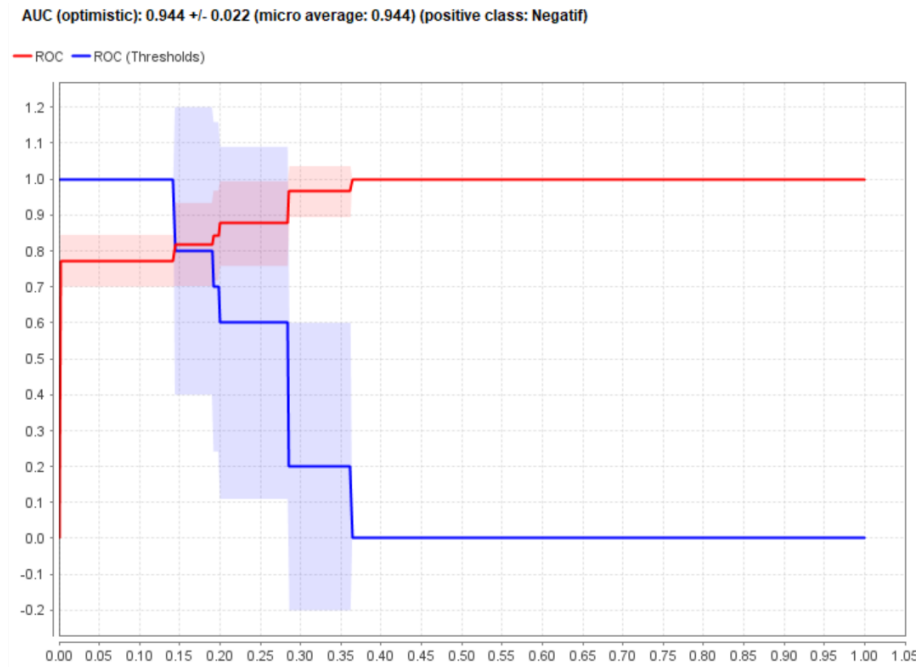
Dari tabel tersebut, terlihat hasil dari pengujian terhadap data tes terlihat tingkat akurasi tertinggi pada fold ke-8 sebesar 72%, kemudian *fold* ke-10 sebesar 70%, dan *fold* ke-9 dengan tingkat akurasi sebesar 69%. Sedangkan tingkat akurasi terendah sebesar 61% terjadi pada *fold* ke-2.

Sedangkan pada tingkat *Precision*, terlihat tingkat akurasi tertinggi terjadi pada *fold* ke-7 sebesar 100%, kemudian terjadi pada *fold* ke-6 sebesar 94%, dan *fold* ke-4 sebesar 90%. Tingkat *precision* terendah terjadi pada *fold* ke-1 sebesar 77%.

Recall tertinggi terjadi pada *fold* ke-9 sebesar 64,2%, kemudian *fold* ke-2 sebesar 64,1%, dan *fold* ke-8 sebesar 63%. Tingkat *recall* terendah terjadi pada *fold* ke-7 sebesar 38%.

Setelah melakukan pengujian terhadap dataset yang ada, maka langkah selanjutnya adalah menguji secara *cross validation* terhadap dataset menggunakan algoritma Naïve Bayes dengan menggunakan Rapidminer dengan data latih sebanyak 760 *tweet* dan sisanya sebagai data uji. Didapat hasil pengujian untuk *accuracy*, *precision*, dan *recall* adalah sebagai berikut: *Accuracy*: 76.29%, *Precision*: 75.40%, *Recall*: 77.29%. Hasil nilai akurasi yang dihasilkan dari dataset tersebut mencapai 76,29% yang menunjukkan data tersebut memiliki nilai yang baik dan memiliki kecenderungan positif yang lebih besar daripada negatif. Hasil nilai presisi yang dihasilkan data tersebut mencapai 75.40% yang terbilang akurat dalam menunjukkan data yang positif. Pada data recall menunjukkan hasil yang baik mencapai 77.29% menunjukkan bahwa prediksi benar positif yang tinggi terhadap data seluruhnya positif.

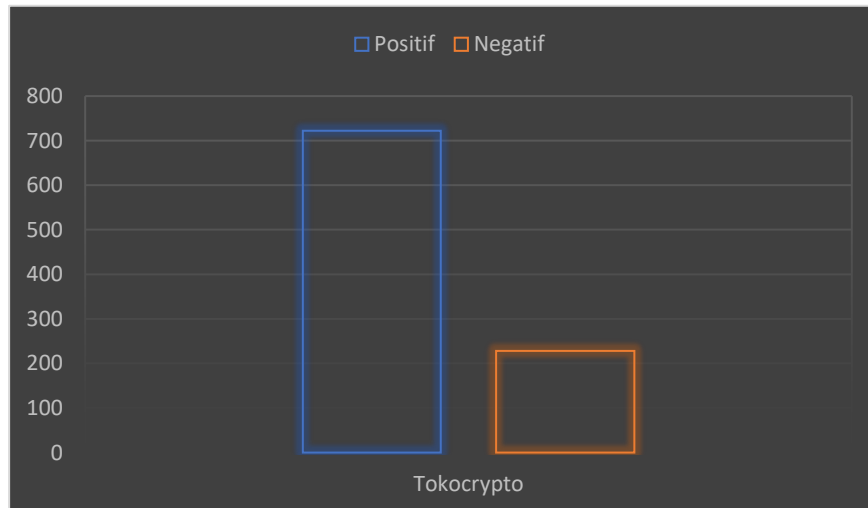
Nilai ROC dari dataset Tokocrypto menghasilkan nilai seperti berikut ini.



Gambar 5. Nilai ROC dataset Tokocrypto

Nilai AUC yang dihasilkan adalah 0.852 yang mendekati nilai 1.0. Artinya dataset ini memiliki nilai uji diagnostik yang baik karena nilai $AUC > 0.5$.

Setelah data diolah, terlihat bahwa jumlah Tweet yang menunjukkan sentimen positif terhadap aplikasi Tokocrypto yang ada di Twitter terdapat 722 cuitan, sedangkan 228 menunjukkan sentimen negatif terhadap Tokocrypto.



Gambar 6. Jumlah cuitan positif dan negatif terhadap Tokocrypto

4. Kesimpulan dan Saran

Setelah melakukan analisa terhadap 2 (dua) aplikasi *cryptocurrency exchange* yang terdiri dari 950 data cuitan di Twitter terhadap aplikasi Indodax dan Tokocrypto maka dapat disimpulkan sebagai berikut. Berdasarkan data latih dan data uji Indodax maupun Tokocrypto, aplikasi Indodax memiliki banyak sentimen positif yang lebih baik daripada Tokocrypto. Indodax memiliki sentimen positif 85%, sedangkan Tokocrypto memiliki nilai sentimen positif 76%. Dengan demikian, Indodax jauh lebih unggul 9% memiliki sentimen positif daripada Tokocrypto. Bagi yang ingin melanjutkan penelitian ini, disarankan untuk menggunakan aplikasi *cryptocurrency exchange* selain Indodax maupun Tokocrypto. Hal ini memungkinkan untuk dilakukan mengingat ada 25 aplikasi lainnya yang bisa dijadikan bahan penelitian lanjutan.

Daftar Pustaka

- [1] M. Noh, M. S., and Abu Bakar, M. S. (2020). Cryptocurrency as A Main Currency: A Maqasidic Approach. *Al-Uqud: Journal of Islamic Economics*, 4(1), 115–132. <https://doi.org/10.26740/aluqud.v4n1.p115-132>
- [2] O. Dniprov, Y. Chyzhmar, A. Fomenko, V. Shablysty, and O. Sydorov, “Legal Status Of Cryptocurrency as Electronic Money,” *Journal of Legal, Ethical and Regulatory Issues*, vol. 22, no. 2, pp. 1-6, 2019.
- [3] Bappebti. 2019. Peraturan Badan Pengawas Perdagangan Berjangka Komoditi Nomor 5 Tahun 2019 tentang Ketentuan Teknis Penyelenggaraan Pasar Fisik Aset Kripto (Crypto Asset) di Bursa Berjangka.
- [4] Bappebti. 2022. Peraturan Badan Pengawas Perdagangan Berjangka Komoditi Nomor 11 Tahun 2022 tentang Penetapan Daftar Aset Kripto yang diperdagangkan di Pasar Fisik Aset Kripto.
- [5] S. Rizkia, E. B. Setiawan, and D. Puspadari, “Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF,” *e-Proceeding of Engineering*, vol. 6, no. 2, pp. 9683-9693, 2019, ISSN: 2355-9365.
- [6] S. Christina and D. Ronald, “Peran *Opinion Mining* dan *Sentiment Analysis* Untuk Mengidentifikasi Sentimen Publik Dalam Sistem E-Governance,” *J. Teknologi Informasi*, vol. 10, no. 2, pp. 65-69, 2016, doi: <https://doi.org/10.47111/jti.v10i2.1433>.
- [7] B. M. Akbar, A. T. Akbar, and R. Husaini, “Analysis of Sentiments and Emotions about Sinovac Vaccine Using Naïve Bayes,” *Telematika: Jurnal Informatika dan Teknologi Informasi*, vol. 19, no. 2, pp. 185-200, 2022, doi: 10.31515/telematika.v19i2.7601.
- [8] A. Deolika, Kusrini, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *Jurnal Teknologi Informasi*, vol. 3, no. 2, pp 179-184, 2019.
- [9] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335
- [10] Putra, M. Wino Adi, Susanti, Erwin, and Herwin, “Analisis Sentimen Dompok Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier,” *IT Journal Research and Development (ITJRD)*. 5(1). 72-86, 2020, DOI : 10.25299/itjrd.2020.vol5(1).5159.
- [11] T. O’Reilly and Sarah Milstein. 2012. *The Twitter Book 2nd Edition*. California: O’Reilly Media Inc.
- [12] F. Tempola, M. Muhammad, A. Khairan, “Perbandingan Klasifikasi Antara KNN dan Naïve Bayes Pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 5, no. 5, 2018, DOI: 10.25126/jtiik20185983.
- [13] F. Syarief, “Pemanfaatan Media Sosial Dalam Proses Pembentukan Opini Publik (Analisa Wacana Twitter SBY),” *Jurnal Komunikasi*, vol. 8, no. 3, pp. 262-266, 2017.

- [14] N. K. Widyasanti, I. K. G. D. Putra, and N. K. D. Rusjyanthi, “Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia,” *J. Merpati*, vol. 6, no. 2, pp. 119-126, 2018.
- [15] Budi Raharjo. 2022. *Uang Masa Depan: Blockchain, Bitcoin, Cryptocurrencies*. Indonesia: Yayasan Prima Agus Teknik.
- [16] T. T. Maskoen, A. Masthura, and Suwarman, “Nilai Area Under Curve dan Akurasi Neutrophil Gelatinase Associated Lipocalin untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma di Instalasi Gawat Darurat RSUP dr. Hasan Sadikin Bandung,” *J. Anesthesia & Critical Care*, vol. 35, no. 3, pp. 158 – 164, 2017.