PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN CRYPTOCURRENCY PADA MEDIA SOSIAL TWITTER

Rizky Adie Riyanto
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
if18.rizkyriyanto@mhs.ubpkarawang.ac.id

Yana Cahyana
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
yana.cahyana@ubpkarawang.ac.id

Rahmat
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
rahmat@ubpkarawang.ac.id

***Abstract* —** Dalam penelitian ini, sentimen masyarakat mengenai *cryptocurrency* yang disampaikan melalui *tweet* pada media sosial twitter akan di analisis menggunakan *machine learning*. Terdapat beberapa tahapan dalam melakukan analisis sentimen yaitu tahap pengumpulan data, *preprocessing, labeling*, pembobotan kata, klasifikasi dan evaluasi. Proses pengumpulan data dilakukan dengan metode *scraping* menggunakan *library* python snscrape dan menghasilkan sebanyak 10000 data *tweet* bahasa Indonesia yang berkaitan yang berkaitan dengan *cryptocurrency*. Proses *preprocessing* untuk pembersihan data terdiri dari 6 tahapan yaitu *remove duplicate*, *case folding, cleansing*, normalisasi kata, tokenizing, *stopword removal* dan *lemmatization*. Pada proses *labeling* data akan dilakukan secara otomatis menggunakan program python dengan *library* textblob, hasil *labeling* data terbagi menjadi 3 kelas yaitu sentimen positif, negatif dan netral. Pada tahap klasifikasi, algoritma naive bayes akan dikombinasikan dengan fitur pembobotan kata TF-IDF dalam satu kelas *pipeline* python. Data *tweet* yang digunakan dalam tahap klasifikasi sebanyak 5000 data yang terdiri dari 2500 data dengan label positif dan 2500 data dengan label negatif yang kemudian dibagi menjadi 2 yaitu sebanyak 4000 data untuk data latih dan 1000 data untuk data uji. Pengujian dan evaluasi yang dilakukan dengan metode *confusion matrix* menggunakan algoritma naive bayes menghasilkan nilai akurasi sebesar 89%.

***Kata kunci* —** *Cryptocurrency*, Klasifikasi, Analisis sentimen, Naive bayes.

I. PENDAHULUAN

Dalam beberapa tahun terakhir, *cryptocurrency* telah berkembang sangat pesat sehingga menjadi cepat tersebar ke seluruh dunia. *Cryptocurrency* awalnya dibuat sebagai mata uang digital berbasis *blockchain* yang digunakan untuk transaksi perdangangan karena *cryptocurrency* dianggap dapat mengurangi waktu dan biaya untuk melakukan transaksi internasional [1]. *Cryptocurrency* dimulai pada tahun 2009 dengan diciptakannya mata uang digital pertama yaitu bitcoin oleh Satoshi Nakamoto, bitcoin merupakan *cryptocurrency* pertama dan mata uang digital dengan kapitalisasi pasar terbesar hingga saat ini. *Cryptocurrency* menjadi semakin populer baik sebagai media pertukaran maupun peluang untuk berinvestasi. *Volume* perdagangan *cryptocurrency* meningkat secara signifikan yang disebabkan oleh popularitasnya yang terus naik sejak diluncurkan pada tahun 2009 [2].

Opini tentang *cryptocurrency* ramai dibahas pada media sosial twitter. Analisis sentimen sering digunakan untuk melakukan analisis pada komentar atau opini di media sosial dan kemudian mengubahnya menjadi sebuah hal yang lebih bermakna, salah satunya dalam bentuk tren positif dan negatif [3]. Semakin banyak pengguna twitter menyampaikan opini pada melalui akun mereka, maka semakin banyak opini yang akan terbentuk mengenai *cryptocurrency* itu sendiri. Namun, menentukan opini tersebut merupakan sentimen positif atau negatif akan sangat sulit jika data yang digunakan sangat banyak [4].

Berdasarkan hasil studi pendahuluan, beberapa penelitian tentang analisis sentimen telah dilakukan sebelumnya. Handayani dan Sulistiyawati (2021) menggunakan algoritma Naive Bayes dalam penelitian mereka untuk analisis sentimen tanggapan masyarakat terhadap berita harian Covid-19 di Twitter. Hasil penelitian ini menunjukkan sentimen positif sebesar 11%, sentimen negatif 85%, dan sentimen netral 4%, dengan akurasi 78% [5]. Selain itu, Deviyanto dan Wahyudi (2018) melakukan penelitian menggunakan algoritma K-Nearest Neighbor untuk menganalisis sentimen publik pada pemilihan gubernur Jakarta. Hasil penelitian itu hanya mendapat 67% pada akurasi [6]. Kemudian Salam, Zeniarja & Khasanah (2018) menggunakan algoritma K-Nearest Neighbor untuk analisis sentimen akun ekspedisi J&T. Tingkat akurasi tertinggi dari hasil penelitian ini adalah 79,21% [7]. Fikri dkk (2021) melakukan analisis sentimen terkait Universitas Muhammadiyah Malang dengan membandingkan algoritma Naive Bayes dan Support Vector Machine. Hasil penelitian ini menunjukkan bahwa hasil akurasi algoritma Naive Bayes lebih baik dari algoritma SVM dengan selisih sebesar 3,45% [8].

Penelitian yang telah dilakukan memiliki beberapa kekurangan atau kelemahan, seperti perbandingan label data yang digunakan tidak seimbang dan berisi data dengan label netral sehingga akurasi yang diperoleh kurang baik, terlebih lagi tahap *preprocessing* yang tidak lengkap dapat mempengaruhi akurasi. nilai. Analisis sentimen dengan menggunakan algoritma K-Nearest Neighbor juga tidak memberikan hasil akurasi yang memuaskan.

Berdasarkan uraian masalah diatas, penelitian ini bertujuan untuk menyempurnakan metode penelitian sebelumnya dengan menambahkan tahapan *preprocessing*, penyeimbangan data yang digunakan untuk klasifikasi, dan menghilangkan data netral. Proses yang ditingkatkan diharapkan menghasilkan akurasi yang lebih tinggi dalam klasifikasi *cryptocurrency* oleh pengguna Twitter.

II. TINJAUAN PUSTAKA

A. Analisis Sentimen

Analisis sentimen adalah sebuah proses dalam text mining yang digunakan untuk mengklasifikasikan opini dalam sebuah teks untuk mengetahui apakah opini tersebut merupakan sentimen positif atau sentimen negatif [9]. Analisis sentimen merupakan metode kombinasi dari *data mining* dan *text mining*, atau metode yang digunakan untuk mengetahui perbedaan pendapat yang diungkapkan oleh publik tentang suatu produk, layanan, atau institusi melalui berbagai media [10]. Berdasarkan beberapa sumber tersebut, analisis sentimen dapat dikatakan sebagai penelitian yang menganalisis opini tentang entitas yang diproses dari teks dengan tujuan untuk mengelompokkan teks menjadi opini positif dan negatif. Salah satu analisis sentimen yang sering dilakukan pada penelitian adalah analisis sentimen terhadap suatu hal pada berbagai media sosial termasuk twitter.

B. *Cryptocurrency*

Cryptocurrency merupakan mata uang digital yang dibangun menggunakan teknologi bernama *blockchain*. Teknologi *blockchain* menghubungkan semua data yang ada dalam lingkungan pengguna *cryptocurrency* [11]. Aset *cryptocurrency* dirancang dengan kode enkripsi sehingga dapat disimpan pada sebuah perangkat komputer dan dapat di *transfer* kepada sesama pengguna sehingga *cryptocurrency* sangat memungkinkan untuk digunakan sebagai metode pembayaran dalam suatu transaksi perdangangan. Ada banyak jenis *cryptocurrency* di seluruh dunia saat ini, tetapi hanya sedikit yang dikenal oleh komunitas global seperti Bitcoin, Ripple, Ethereum, Monero dan Litecoin. Tentu yang paling atas adalah Bitcoin dengan nilainya terus meningkat pesat dan memiliki banyak pengguna.

C. Twitter

Twitter merupakan *platform* media sosial yang digunakan untuk mengirim sebuah pesan berbasis teks. Lebih dari 500 juta *tweet* pengguna dikirim dalam setiap hari pada awal tahun 2013. Karena popularitasnya yang sangat tinggi, saat ini twitter sering digunakan untuk berbagai keperluan seperti media untuk berkomunikasi, melakukan pembelajaran online, sarana protes hingga untuk tujuan politik [12]. Twitter memiliki istilah "*tweet*", yang berarti bahwa pengguna twitter dapat memberikan berita, suara dan pendapat pengguna twitter lainnya, terutama pada topik dan hal-hal yang menjadi bahasan utama. Saat ini, masyarakat umum banyak menggunakan twitter untuk menyampaikan keluhan mereka, mulai dari keluhan tentang kehidupan pribadi hingga keluhan tentang layanan yang diberikan oleh industri.

D. *Data Mining*

Data mining merupakan adalah cara metode yang relatif mudah sederhana untuk mendapatkan memperoleh sebuah pengetahuan dan pola secara otomatis. *Data mining* dapat didefinisikan sebagai aktivitas mengekstrak informasi yang sebelumnya tidak diketahui dari suatu data. Dapat disimpulkan bahwa *data mining* adalah proses pencarian informasi melalui ekstraksi otomatis menggunakan suatu pola atau metode untuk menemukan informasi baru yang sebelumnya tidak diketahui dan mungkin berguna di masa yang akan datang.

E. Naive Bayes

Naive Bayes adalah algoritma penambangan data yang mudah digunakan dengan waktu pemrosesan yang cepat, implementasi yang mudah, struktur yang cukup sederhana, dan efisiensi yang tinggi. Naive Bayes merupakan salah satu algoritma klasifikasi yang digunakan untuk memprediksi probabilitas kelas. Naive bayes sangat akurat dalam hal implementasi database dengan jumlah yang besar, serupa dengan beberapa algoritma lainnya seperti algoritma neural network dan decision tree. Algoritma naive bayes menggunakan teori probabilitas yang dikembangkan oleh ilmuwan dari Inggris, Thomas Bayes. Prinsip metode dari pendekatan ini adalah untuk memprediksi kemungkinan kejadian di masa depan berdasarkan data yang diperoleh sebelumnya. Naive Bayes adalah probabilitas yang dapat digunakan untuk mendefinisikan sekelompok kelas dalam dokumen teks dan dapat mengolah data dalam jumlah yang besar [13].

Adapun persamaan dari dasar metode pada teorema bayes tersebut sebagai berikut.

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

Keterangan:

X	: Data dengan class yang belum diketahui
H	: Hipotesis data X yang merupakan suatu class spesifik
P(H|X)	: Probabilitas hipotesis H berdasarkan kondisi X
P(H)	: Probabilitas hipotesis H
P(X|H)	: Probabilitas X berdasarkan kondisi pada hipotesis H
P(X)	: Probabilitas X

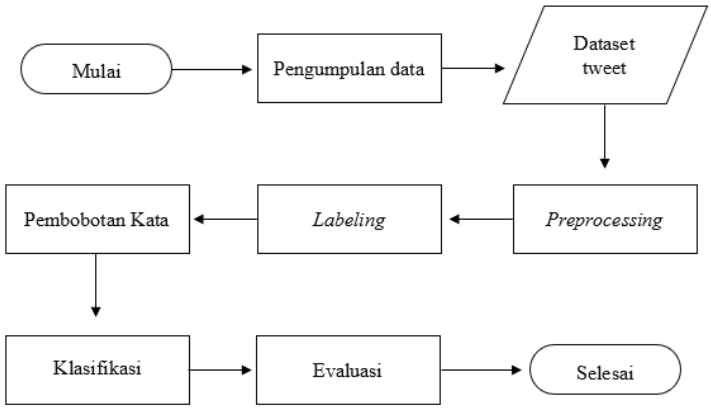
III. METODOLOGI PENELITIAN

A. Objek Penelitian

Objek penelitian yang digunakan pada penelitian ini berupa data *tweet* yang berkaitan dengan *cryptocurrency* berbahasa Indonesia pada media sosial twitter. Data tweet yang digunakan sebanyak 10000 data dengan rentang waktu pengambilan yaitu selama dua tahun ke belakang dimulai pada bulan juni tahun 2020 hingga bulan juni tahun 2022.

B. Prosedur Penelitian

Prosedur penelitian merupakan gambaran umum dari alur penelitian yang dilaksanakan selama pengerjaan tugas akhir.



Gambar 1 Prosedur Penelitian

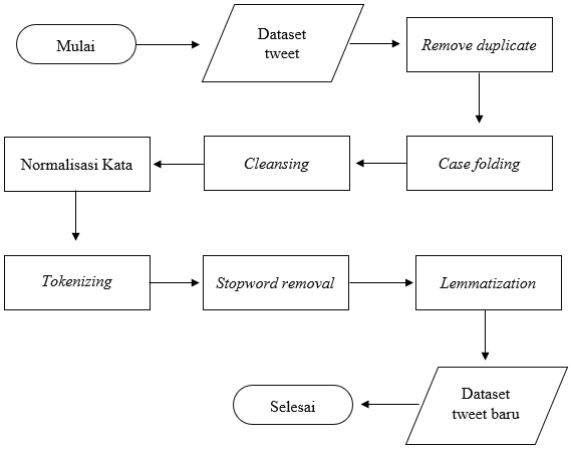
Penelitian ini diawali dengan tahap pengumpulan data menggunakan teknik *scraping* dengan bahasa pemrograman python. Tahap selanjutnya setelah pengambilan data yaitu *preprocessing* yang berguna untuk menyeleksi dan membersihkan data mentah dari beberapa bagian yang dapat mengganggu proses selanjutnya. Setelah data menjadi bersih lalu dilakukan *labeling* ke dalam dua kelas yaitu positif dan negatif menggunakan python dengan *library* textblob. Setelah tahap *labeling* yaitu pembobotan kata menggunakan metode TF-IDF yang dibarengi dengan implementasi algoritma yang akan digunakan yaitu naive bayes. Tahap terakhir yaitu evaluasi yang bertujuan untuk mengetahui hasil dari pengujian yang sudah dilakukan, pada tahap ini menggunakan metode *confusion matrix* yang telah digunakan pada penelitian sebelumnya.

1. Pengumpulan Data

   Pengumpulan data pada penelitian ini dilakukan dengan cara mengumpulkan *tweet* berbahasa Indonesia pada media sosial Twitter yang membahas mengenai *cryptocurrency*. Pengumpulan data dilakukan pada media sosial twitter karena topik yang berkaitan dengan *cryptocurrency* sering kali masuk dalam trending Twitter Indonesia. Data dikumpulkan secara acak baik dari *tweet* pengguna biasa atau dari media *online*. Proses *scraping* dilakukan menggunakan bahasa pemrograman python.

2. *Preprocessing*

   Data *tweet* yang sudah dikumpulkan dari hasil *scraping* tidak dapat langsung digunakan untuk klasifikasi karena masih terdapat banyak gangguan, karena itu memerlukan tahap bernama *preprocessing*. Tujuan dari tahap *preprocessing* yaitu membersihkan data atau *tweet* dari kata-kata yang tidak diperlukan dalam penelitian dan tidak memiliki makna serta mengurangi *volume* kata.



Gambar 2 Tahapan *Preprocessing*

3. Labeling

Langkah selanjutnya setelah *preprocessing* yaitu melakukan *labeling*. Proses ini dilakukan untuk memberi label pada konten atau *tweet* pengguna twitter dan kemudian setiap *tweet* akan diberi indeks untuk membedakan setiap nilai sentimen. Dalam penelitian ini terdapat dua kelas dalam memberi label ke suatu *tweet* yaitu kelas positif dan negatif. Hasil dari pelabelan akan di visualisasikan ke dalam dua bentuk yaitu *bar chart* dan *word cloud*. *Bar chart* digunakan untuk menunjukkan perbandingan pada jumlah sentimen sedangkan *word cloud* akan menampilkan kata-kata umum yang sering muncul pada hasil sentimen positif dan negatif.

4. Pembobotan Kata

Pembobotan kata dilakukan setelah selesai melakukan tahap *labeling*. Pembobotan kata merupakan suatu metode yang digunakan untuk memberikan nilai pada frekuensi kemunculan kata dalam sebuah dokumen yang nantinya akan diproses pada tahap klasifikasi naive bayes. Semakin sering kemunculan suatu kata dalam dokumen maka akan memberikan nilai kesesuaian yang semakin besar. Dalam penelitian ini proses pembobotan kata dilakukan menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF). Setelah proses pembobotan selesai maka dokumen data telah siap untuk dilakukan training dalam tahap klasifikasi.

5. Klasifikasi

Algoritma yang digunakan untuk klasifikasi dalam penelitian ini adalah Naive Bayes. Dalam beberapa penelitian terkait algoritma naive bayes memiliki tingkat akurasi yang tinggi dalam proses klasifikasi. Pada proses ini data yang digunakan merupakan data bersih hasil proses *text processing* hingga pembobotan kata menggunakan TF-IDF. Proses ini dimulai dengan menentukan berapa banyak data yang akan dipelajari. Setelah pelatihan data berhasil, pengujian dijalankan menggunakan data uji untuk menguji keakuratan klasifikasi yang dilakukan.

6. Pengujian dan Evaluasi

Pengujian dan evaluasi model dilakukan untuk mengetahui kinerja model. Evaluasi model dilakukan menggunakan metode *confusion matrix* dengan indikator *true positive*, *true negative*, *false positive* dan *false negative*. Hasil dari *confusion matrix* akan menunjukkan tabel untuk setiap model dengan nilai *accuracy*, *precision*, *recall* dan *F1-score*.

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Pengambilan data pada penelitian ini dilakukan dengan cara *scraping* data dari media sosial Twitter dengan menggunakan *library* snscrape pada pemrograman python, data yang diambil merupakan *tweet* berbahasa Indonesia dengan kata kunci "kripto". *Library* snscrape mudah digunakan karena tidak perlu mengakses Twitter API dengan menggunakan token yang hanya bisa didapat jika mengaktifkan akun developer Twitter. Setelah itu, *dataset tweet* mentah kemudian disimpan dalam bentuk xlsx. *Library* snscrape dapat menjalan beberapa perintah tambahan seperti *start-date* dan *end-date* untuk mengatur rentang waktu pengambilan *tweet*, *filtering links* untuk menghilangkan *tweet* yang berisi *link* dan *filtering replies* untuk menghilangkan *reply* atau *mention* dari sebuah *tweet*.

Tabel 1 Hasil pengumpulan data

No	Kalimat
1	Industri kripto di Indonesia semakin berkembang. Bagi yg ingin terjun ke industri blockchain kenali diri & potensi. Apapun yang kamu pilih selalu berikan yg terbaik dan jangan mudah menyerah #oscardarmawan #KriptoCoin #Blockchain
2	Latah uang kripto ini bahaya... Entar lama2, masing2 agama bikin uang kripto sendiri, lalu persembahan kudu pake kripto. Bisa2 nanti klo ndak lwt kripto dianggap tdk beri persembahan...
3	Kegiatan ini bisa menjadi wadah bagi masyarakat dan investor pemula untuk lebih memahami blockchain dan aset kripto#TokoinvasionSurabaya
4	Crypto itu scam semua Semua adalah manipulasi #CryptocurrencyNews #kripto #penipuan #bukapikiran
…	
10000	Kirain investasi cuma saham, reksadana dan emas aja. Ternyta baru tau ada investais aset kripto. Tempat yang lagi naik daun buat invest kripto nih @tokocrypto #TokocryptoAja #UpgradeYuk

B. Hasil *Preprocessing*

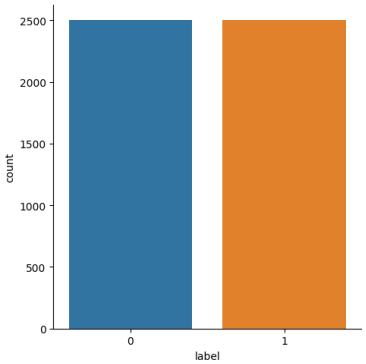
Data yang baru didapat dari hasil scraping akan dibersihkan terlebih dahulu melalui tahapan *preprocessing* agar tidak ada atribut yang mengganggu, terdapat 6 proses pada *preprocessing* ini.

Tabel 2 Hasil *preprocessing*

No	Proses	Sebelum	Sesudah
1	*Remove Duplicate*	Kalo aku sih dari awal udah ngedukung adanya pajak kripto, kan cinta Indonesia. Mantep sih emang. #SerbaSerbiPajakKripto Kalo aku sih dari awal udah ngedukung adanya pajak kripto, kan cinta Indonesia. Mantep sih emang. #SerbaSerbiPajakKripto	Kalo aku sih dari awal udah ngedukung adanya pajak kripto, kan cinta Indonesia. Mantep sih emang. #SerbaSerbiPajakKripto
2	*Case Folding*	Kalo aku sih dari awal udah ngedukung adanya pajak kripto, kan cinta Indonesia. Mantep sih emang. #SerbaSerbiPajakKripto	kalo aku sih dari awal udah ngedukung adanya pajak kripto, kan cinta indonesia. mantep sih emang. #serbaserbipajakkripto
3	*Cleansing*	kalo aku sih dari awal udah ngedukung adanya pajak kripto, kan cinta indonesia. mantep sih emang. #serbaserbipajakkripto	kalo aku sih dari awal udah ngedukung adanya pajak kripto kan cinta indonesia mantep sih emang
4	Normalisasi Kata	kalo aku sih dari awal udah ngedukung adanya pajak kripto kan cinta indonesia mantep sih emang	kalau aku sih dari awal sudah ngedukung adanya pajak kripto kan cinta indonesia mantap sih emang
5	*Tokenizing*	sangat dukung dan keren tokocrypto berarti mendukung pemerintah ini dengan menerapkan pajak kripto	[sangat, dukung, dan, keren, tokocrypto, berarti, mendukung, pemerintah, ini, dengan, menerapkan, pajak, kripto]
6	*Stopword Removal*	[sangat, dukung, dan, keren, tokocrypto, berarti, mendukung, pemerintah, ini, dengan, menerapkan, pajak, kripto]	[dukung, keren, tokocrypto, mendukung, pemerintah, menerapkan, pajak, kripto]
7	*Lemmatizing*	[dukung, keren, tokocrypto, mendukung, pemerintah, menerapkan, pajak, kripto]	['dukung', 'keren', 'tokocrypto', 'dukung', 'perintah', 'terap', 'pajak', 'kripto']

C. Hasil *Labeling*

Langkah selanjutnya yaitu *labeling*. Pada penelitian ini *labeling* data dilakukan menggunakan program python dengan bantuan *library* textblob. Pada proses *labeling* setiap *tweet* akan diberikan nilai polaritas antara -1 sampai 1, *tweet* yang sudah mendapatkan nilai polaritas selanjutnya akan disimpan dalam kelas sentimen positif dan negatif. Untuk memudahkan proses selanjutnya maka sentimen positif akan diberi label 0 sedangkan sentimen negatif akan diberi label 1.



Gambar 3 Hasil *labeling*

D. Hasil Pembobotan Kata

Pada proses ini semua kata akan diberikan nilai atau bobot sesuai dengan frekuensi kemunculan kata tersebut di dalam sebuah dokumen atau *dataset*. Pembobotan kata menggunakan metode TF-IDF dengan bantuan program python beserta *library* TfidfVectorizer.

No	kata	tfidf_value
1	kripto	334,3097382
2	pajak	152,1080792
3	aset	129,3952093
4	pakai	110,0220646
5	donasi	103,3683786
6	orang	97,47254973
7	investasi	96,49807811
8	keren	94,77304534
9	uang	86,61795692
10	beli	80,62716725

Gambar 4 Hasil pembobotan kata

E. Klasifikasi

Pada tahapan klasifikasi ini pembobotan kata dan implementasi algoritma akan digabung pada satu *class pipeline* menggunakan bahasa pemrograman python dengan bantuan *library* scikit-learn. Langkah pertama yang dilakukan dalam klasifikasi yaitu menginstall *library* yang dibutuhkan, dilanjutkan dengan mendeklarasi semua *library* yang sudah terinstall. *Library* scikit-learn yang dibutuhkan pada klasifikasi ini yaitu *feature extraction* TfidfVectorizer yang digunakan untuk pembobotan kata, selanjutnya MultinomialNB yang merupakan algoritma dari naive bayes, *library metrics* untuk menampilkan nilai hasil klasifikasi dan *model selection train_test_split* yang digunakan untuk membagi data menjadi data latih dan data uji.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import Pipeline
from sklearn import metrics
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
```

Gambar 5 Kode program *library* klasifikasi

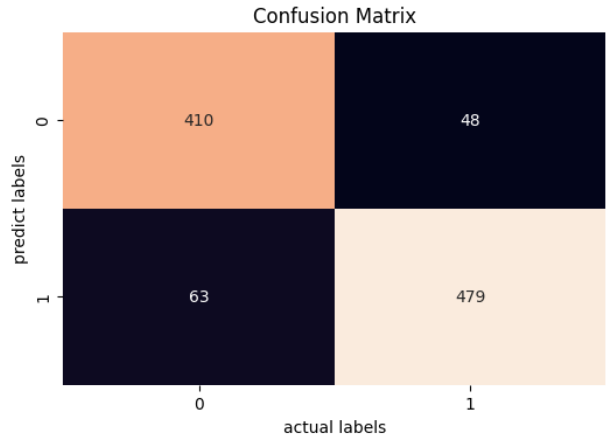
Setelah deklarasi *library* proses selanjutnya yaitu membuat sebuah *class pipeline* yang berisikan pembobotan kata TfidfVectorizer dan algoritma MultinomialNB. Pada proses klasifikasi ini data yang digunakan sebagai test yaitu 0.2 atau sebanyak 20% dengan menambahkan fungsi *random state* agar hasil klasifikasi tidak berubah-ubah.

```
clf = Pipeline(steps=[
    ('features', TfidfVectorizer()),
    ('model', MultinomialNB())
])

X_train, X_test, y_train, y_test = train_test_split(
    df["tweet_bersih"], df["label"], test_size=0.2, random_state=0)

clf.fit(X_train, y_train)
# predicted
y_pred = clf.predict(X_test)
report = metrics.classification_report(y_test, y_pred)
accuracy = metrics.accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
print("Report : \n", report)
print("accuracy : ", accuracy)
print("confusion matrix\n", cm)
```

Gambar 6 Kode program klasifikasi

F. Hasil Pengujian dan Evaluasi

Pengujian klasifikasi pada penelitian ini menggunakan algoritma naive bayes untuk analisis sentimen dengan menggunakan data uji sebanyak 1000 data atau 20% dari keseluruhan data yang digunakan. Untuk hasil dari setiap kelas *confusion matrix* dapat dilihat pada gambar dibawah.



Gambar 7 Hasil *confusion matrix*

Akurasi digunakan untuk mengetahui seberapa akurat model algoritma yang digunakan dalam proses klasifikasi, nilai akurasi pengujian didapat dari jumlah data yang benar dibagi dengan total data, untuk mendapatkan nilai akurasi dapat dihitung dengan cara.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$
$$= \frac{410 + 479}{410 + 48 + 63 + 479} = \frac{889}{1000} = 0.889$$

Selain menggunakan perhitungan manual, nilai akurasi juga dapat dihasilkan dengan menggunakan pemrograman pyhton yang dapat dilihat pada gambar dibawah.

accuracy			0.89
macro avg	0.89	0.89	0.89
weighted avg	0.89	0.89	0.89

Gambar 8 Hasil akurasi

Nilai akurasi yang didapatkan dengan perhitungan manual menghasilkan angka 0.889, sementara untuk hasil dari pemrograman pyhton akan otomatis dibulatkan menjadi 0.89. Dari kedua hasil tersebut maka dapat disimpulkan bahwa nilai akurasi pada penelitian ini yaitu sebesar 89%. Selain itu, kinerja klasifikasi untuk setiap kelasnya dapat diperiksa dengan hasil *precision*, *recall* dan *f1-score*.

Report :

	precision	recall	f1-score	support
0	0.90	0.87	0.88	473
1	0.88	0.91	0.90	527

Gambar 9 Hasil evaluasi

Semua nilai yang dihasilkan oleh pemrograman pyhton akan secara otomatis dibulatkan. Hasil dari evaluasi model menunjukkan nilai *precision* untuk kelas atau label positif yaitu 90% dan kelas negatif 88%. Nilai *Recall* untuk kelas positif yaitu 87% dan kelas negatif 91%. Hasil terakhir yaitu *f1-score* yang merupakan perbandingan antara *precision* dengan *recall* dalam satu kelas, nilai *f1-score* kelas positif yaitu 88% dan kelas negatif 90%

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan beberapa hal sebagai berikut.

1. *Dataset* pada penelitian ini didapat dengan teknik *scraping* dari media sosial Twitter menggunakan pemrograman pyhton dengan *librabry* snscrape. Hasil *scraping* data yaitu sebanyak 10000 data berupa *tweet* yang akan dibersihkan melalui tahapan *preprocessing*.
2. Pemberian label atau sentimen pada data tweet dilakukan dengan bantuan *library* textblob pada pyhton yang menghasilkan 3742 data dengan label positif dan 2097 data dengan label negatif, sementara untuk data netral akan dihilangkan.

3. Pengujian dilakukan menggunakan *confusion matrix* pada pemrograman python dengan model yang berisi TF-IDF dan algoritma naive bayes. Dari hasil pengujian model, dapat disimpulkan bahwa analisis sentimen mengenai *cryptocurrency* menggunakan algoritma naive bayes menghasilkan nilai akurasi sebesar 89%.

B. Saran

Penelitian ini masih memiliki beberapa kekurangan yang perlu diperbaiki, agar penelitian selanjutnya dapat mencapai hasil yang lebih baik maka saran yang diberikan adalah sebagai berikut.

1. Pada tahapan preprocessing data menggunakan *library* lain tersedia pada python seperti sastrawi dengan tambahan membuat kamus berisi kata-kata yang tidak ada pada *library* bawaan.
2. Pada tahapan labeling dilakukan secara manual dengan cara memberi label pada satu persatu data baik secara pribadi, berkelompok atau dengan bantuan pakar.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian tugas akhir milik Rizky Adie Riyanto dengan judul Analisis Sentimen Terhadap Cryptocurrency Pada Twitter Menggunakan Algoritma Naive Bayes yang di bimbing oleh Yana Cahyana dan Rahmat.

DAFTAR PUSTAKA

[1] Nitha, Dewa Ayu Fera & I Ketut Westra. (2020). Investasi Cryptocurrency Berdasarkan Peraturan Bappebti No. 5 Tahun 2019. Jurnal Magister Hukum Udayana (Udayana Master Law Journal), 9(4), 712–722.

[2] Schut, Milan. (2017). Bitcoin analysis from an investor's perspective Insight into market relations and diversification possibilities. Semantic Scholar, 33.

[3] Prasetya, Adam et al. (2021). Sentiment Analisis Terhadap Cryptocurrency Berdasarkan Comment Dan Reply Pada Platform Twitter. Journal of Information Systems and Informatics, 3(2), 268–277.

[4] Sulaiman, Fahmi Ilmawan, Wing Wahyu Winarno & Mei Parwanto Kurniawan. (2021). Perancangan Aplikasi Klasifikasi Sentimen Berbasis Web Terhadap Mata Uang Kripto. FAHMA, 19(3), 27–41.

[5] Handayani, Eni Tri & Ari Sulistiyawati. (2021). Analisis Sentimen Respon Masyarakat Terhadap Kabar Harian Covid-19 Pada Twitter Kementerian Kesehatan Dengan Metode Klasifikasi Naive Bayes. Jurnal Teknologi dan Sistem Informasi (JTSI). 2(1), 32-37.

[6] Deviyanto, Akhmad & M. Didik R. Wahyudi. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. JISKa (Jurnal Informatika Sunan Kalijaga). 3(1), 1-13.

[7] Salam, Abu, Junta Zeniarja, & Rima Septiyan Uswatun Khasanah. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspress Indonesia). Prosiding SINTAK. 2, 480-486.

[8] Fikri, Mujaddid Izzul, Trifebi Shina Sabrila & Yufis Azar. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. Smatika Jurnal, 10(02), 71–76.

[9] Siregar, Amril Mutoi, Anis Fitri Nur Masruriyah & Yogi Firman Alfiansah. (2022). Penerapan Algoritma K-Nearest Neighbors untuk Analisis Sentimen pada Buletin APTIKOM. Scientific Student Journal for Information, Technology and Science. 3(1), 125-132.

[10] Sipayung, Evasaria Magdalena, Herastia Maharani & Ivan Zefanya. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier. Jurnal Sistem Informasi (JSI), 8(1), 959–965.

[11] Bhiantara, Ida Bagus Prayoga. (2018). Teknologi Blockchain Cryptocurrency Di Era Revolusi Digital. Prosiding Seminar Nasional Pendidikan Teknik Informatika, 9, 173–177

[12] Pintoko, Brata Mas & Kemas Muslim Lhaksamana. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. E-Proceeding of Engineering, 5(3), 8121–8130.

[13] Lestari, Agnes Rossi Trisna, Rizal Setya Perdana, & M. Ali Fauzi. (2017). Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Näive Bayes dan Pembobotan Emoji. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 1(12), 1718–1724.