

Sentimen Analisis Thread Investasi Pada Twitter Menggunakan Naïve Bayes

Muhammad Rafih S Munir¹⁾, Hartatik²⁾, Andri Syafrianto^{3,a)}

^{1,2)} Fakultas Ilmu Komputer

Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Depok, Condongcatur, Sleman, Yogyakarta, Indonesia

³⁾ Prodi Informatika

STMIK EL RAHMA Yogyakarta, Karangajen, Yogyakarta

Author Emails

^{a)} Corresponding author: andrisyafrianto@gmail.com; andrisyafrianto@stmikelrahma.ac.id

Abstract. Social media is one of the market shares that allows people to consume news or promotions from an investment service because the source data is abundant and is updated every second on the internet. Social media is vulnerable to containing news or information on an investment platform and other forms of investment that are illegal or called fraudulent investments, so to overcome this, an Investment Thread Analyst Sentiment on Twitter is needed. The method used to evaluate investment threads to obtain positive and negative feedback is naive Bayes. The research results received a precision score of 0.69, recall of 0.70, f1-score (f-measure) of 0.74, and accuracy of 0.7.

Keywords :

Investasi, Naive Bayes, Sentiment Analyst

Abstraksi. Media sosial merupakan salah satu pangsa pasar yang memungkinkan masyarakat mengkonsumsi berita atau promosi dari suatu layanan investasi, karena data sumber tersebut sangat melimpah dan update tiap detiknya di internet. Media sosial rentan berisi berita atau informasi suatu platform investasi dan bentuk investasi yang ilegal atau disebut investasi bodong, sehingga untuk mengatasi hal tersebut diperlukan Sentimen Analisis Thread Investasi Pada Twitter. Metode yang digunakan untuk mengevaluasi thread investasi guna memperoleh feedback positif dan negative adalah naïve bayes. Hasil penelitian yang dilakukan memperoleh nilai score precision sebesar 0.69, recall sebesar 0.70, f1-score (f-measure) sebesar 0.74 dan accuracy sebesar 0.7.

Kata Kunci : Investasi, Naive Bayes, Sentimen Analisis

PENDAHULUAN

Sentimen analisis adalah proses untuk mendapatkan berbagai sumber data dari internet melalui berbagai platform media sosial menggunakan analisis teks. Tujuannya adalah untuk mengumpulkan pendapat dari pengguna di platform sosial media. Sentimen analisis merupakan bagian dari *text mining*, data kumpulan sentimen yang akan dianalisis adalah data berupa teks yang dapat diambil dari kolom komentar [1]. Setiap hari internet dibanjiri miliaran data dari berbagai sumber. Analisis sentimen bertindak sebagai alat yang dapat menghubungkan data apa pun. Hal ini memungkinkan bisnis untuk secara efisien mendapatkan masukan inti dari pengguna atau konsumen.

Twitter merupakan salah satu media sosial yang sangat populer dikalangan pengguna internet, hal ini dikarenakan kesederhanaan dan kemudahan dalam penggunaannya, serta pengguna dapat dengan bebas mengeluarkan pendapat atau opini mereka [2]. *Twitter* di luncurkan pertama kali pada tahun 2006, awal kemunculan

Twitter masyarakat merespon cukup baik karena fasilitas yang ada pada *Twitter* terbilang cukup segar pada waktu itu yakni fitur *following* yang memungkinkan pengguna dapat menentukan pertemanan di jejaring tersebut.

Naive Bayes adalah klasifikasi dengan metode *probabilistic* dan statistik diusulkan oleh ilmuwan Inggris Thomas Bayes, *Naive Bayes* menghitung untuk setiap kelas keputusan probabilitas yang memiliki syarat bahwa kelas keputusan benar, sesuai vektor dari informasi objek. Algoritma *Naive Bayes* memperkirakan peluang masa depan sesuai *experience* seseorang pada masa lalu hal tersebut dikenal dengan *Teorema Bayes*. Ciri-ciri *Naive Bayes Classifier* ini adalah praduga yang amat kuat (naïf) akan inden pada tiap-tiap situasi kondisi [1]. Investasi adalah informasi tentang bagaimana menggunakan sebagian dana atau sumber daya yang Anda miliki untuk mendapatkan keuntungan di masa depan [3]. Umumnya masyarakat melakukan investasi untuk mendapatkan keuntungan yang cukup dan sesuai dengan yang diharapkan, di era yang serba digital ini investasi dapat dilakukan dimana saja dan kapan saja melalui beberapa aplikasi investasi yang sudah resmi atau terdaftar Otoritas Jasa Keuangan (OJK). Bentuk dari investasi pun beragam mulai dari saham, obligasi, pasar uang dan yang sedang berkembang pesat saat ini adalah *Cryptocurrency* seperti *bitcoin*, *Ethereum*, *binance* dan masih banyak lagi.

Media sosial merupakan salah satu pangsa pasar yang memungkinkan masyarakat mengonsumsi berita atau promosi dari suatu layanan investasi, karena data sumber tersebut sangat melimpah dan *update* tiap detiknya di internet dan hal tersebut menjadikan media sosial rentan berisi berita atau informasi suatu platform investasi dan bentuk investasi yang ilegal atau disebut investasi bodong, sehingga untuk mengatasi hal tersebut diperlukan Sentimen Analisis *Thread* Investasi Pada *Twitter* Menggunakan *Naive Bayes* hasil yang diperoleh berupa komentar, kritik dan saran baik positif ataupun negative pada suatu bentuk, model produk atau aplikasi investasi yang beredar dan di bahas pada *Twitter*. Diharapkan mampu menjadi gambaran pada masyarakat mengenai bentuk, model produk, atau aplikasi investasi yang aman dan terpercaya.

Berdasarkan hal tersebut maka penelitian ini akan melakukan klasifikasi opini di media sosial (*Twitter*) sehingga dapat dilakukan evaluasi terhadap *thread* investasi untuk memperoleh *feedback* positif dan negative dengan algoritma yang digunakan adalah *naive bayes* dan tidak membandingkan atau menggunakan algoritma lain.

TINJAUAN PUSTAKA

Penelitian terdahulu merupakan penelitian yang telah dilakukan sebelumnya baik cara analisis sistem, metode yang digunakan dan hasil yang diperoleh. Penelitian yang dilakukan [1] menunjukkan banyak yang memberikan “sentimen negatif” terhadap layanan Telkomsel. Hasil pengolahan 151 data latih diketahui bahwa hasil klasifikasi sentimen yang diperoleh adalah 51 tweet sentimen negatif, 51 tweet sentimen netral, dan 49 *tweet* sentimen positif, netral sebagai opini *non-sentimental*. Tingkat akurasi penentuan kategori 70,21% dan micro average 70,20% sedangkan penentuan sentimen memiliki tingkat presisi 70,11% dan recall 70,33%

Peneliti [2] melakukan penelitian mengenai analisis sessional aplikasi ruang guru di *twitter* menyimpulkan bahwa penggunaan seleksi fitur PSO pada algoritma klasifikasi dapat meningkatkan kinerja dan akurasi. Hasil pengujian keseluruhan algoritma SVM (PSO) memiliki nilai akurasi dan performa tertinggi jika dibandingkan dengan NB, SVM, K-NN, NB (PSO), SVM (PSO) dan K-NN (PSO). Sedangkan algoritma NB, tanpa PSO atau dengan PSO, mengambil nilai UAC dengan klasifikasi gagal. Dalam penelitian ini, sumber data yang digunakan hanya dari satu sumber media sosial yaitu *Twitter*. Hal ini dapat menjadi acuan dalam penelitian selanjutnya dengan menggunakan sumber data yang berbeda. Sedangkan peneliti [4] memperoleh hasil penelitian berupa proses klasifikasi model dengan algoritma *k-nearest neighbour*, menggunakan model *shared* data dengan rasio 60:40 dengan data *set review user* di aplikasi *seed and barracks*, nilai akurasi, akurasi dan *recall* Diproduksi adalah 85,14%, 91, 91%, dan 76,44% untuk benih sedangkan untuk barak adalah 81,70%, 87,15%, 75,73%.

Berbeda halnya dengan peneliti [5] bahwa *mind sharing* menggunakan *Library Python text blob* dapat dilakukan dengan hasil yang positif. Dalam penelitian ini, semua 3433 data *tweet* dibagikan menggunakan pustaka *Python text blob* dan hasil yang diperoleh ketika 41,3% *tweet positif*, 44,9% netral secara politik, dan 13,7% *negatif*. Dalam penelitian ini, pada efek visual penggunaan *word cloud*, interpretasi dapat diasumsikan bahwa pikiran pengguna *Twitter* sedang melihat harga di pasar yang menunjukkan pembelian tanpa kata-kata positif atau negatif saat ini. Peneliti [6] menghasilkan penelitian bahwa emosi negatif ternyata tidak ada hubungannya dengan harga

saham, dengan nilai emosi positif 16%, emosi negatif 78% dan emosi netral 6%. Dan kata yang paling banyak digunakan pengguna adalah Indonesia dengan total 306 kata

Sentimen Analisis

Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral. Saat ini, peneliti secara luas menggunakan analisis sentimen sebagai salah satu cabang penelitian dalam ilmu komputer. Jejaring sosial, seperti *Twitter*, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi publik. Analisis sentimen juga dapat disamakan dengan *opinion mining*, karena berfokus pada pendapat yang menyatakan positif atau negatif. Dalam analisis sentimen, penambangan data dilakukan untuk menganalisis, memproses, dan mengekstrak data tekstual dalam suatu entitas, seperti layanan, produk, individu, fenomena atau topik tertentu. Proses analisis dapat mencakup teks ulasan, forum, *tweet*, atau blog, dengan data *pre-processing* mencakup proses *tokenization*, *stop word*, penghapusan, *stemming*, identifikasi sentimen, dan klasifikasi sentimen [7].

Analisis sentimen dapat mengklasifikasikan opini-opini tersebut diantaranya menggunakan algoritma data mining seperti Klasifikasi *Naive Bayes*. Analisis sentimen adalah studi komputasi berkaitan dengan opini masyarakat mengomentari suatu kejadian disampaikan dalam bentuk teks. Problem mendasar pada *opinion mining* adalah mengekstrak opini berdasarkan klasifikasi. Besarnya manfaat bagi pengetahuan dan industri, membuat analisis sentimen berkembang pesat [8].

Investasi

Investasi dapat diartikan sebagai komitmen untuk menggunakan sebagian dana atau sumberdaya untuk mendapatkan keuntungan di masa depan investasi adalah suatu informasi mengenai bagaimana cara menggunakan sebagian dana atau sumberdaya yang dimiliki untuk mendapatkan keuntungan di masa depan [3].

Investasi dapat dibagi menjadi dua yaitu [13]:

1. Investasi Asing

Menurut UU Penanaman Modal No. 25 Tahun 2007 penanaman modal asing adalah kegiatan menanam modal untuk melakukan usaha di wilayah negara Republik Indonesia yang dilakukan oleh penanam modal asing, baik yang menggunakan modal asing sepenuhnya maupun yang berpatungan dengan penanam modal dalam negeri.

2. Investasi Dalam Negeri

Menurut UU Penanaman Modal No. 25 Tahun 2007, penanaman modal dalam negeri adalah kegiatan menanam modal untuk melakukan usaha di wilayah negara Republik Indonesia yang dilakukan oleh penanam modal dalam negeri dengan menggunakan modal dalam negeri.

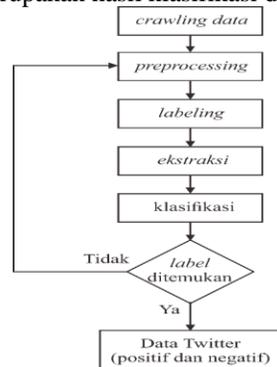
Investasi adalah kegiatan menanam modal pada satu jenis aktiva atau serta dalam kurun waktu yang lama dengan harapan di masa depan akan mendapat keuntungan yang besar. Penanaman modal tersebut dilakukan oleh individu yang mempunyai kelebihan dana atau dana dingin yang tidak mempengaruhi kondisi finansial kebutuhan pokok [14].

METODE PENELITIAN

Alur penelitian yang dilakukan disajikan pada Gambar 1. Berdasarkan Gambar 1, maka:

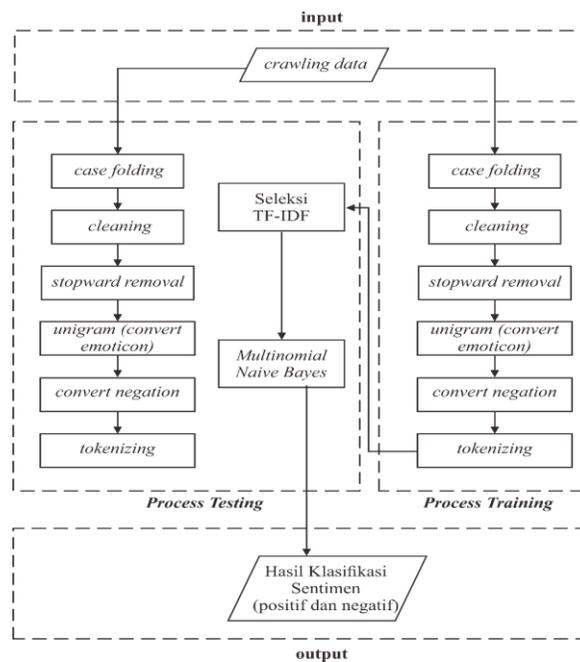
1. *Crawling Data*, data yang diperoleh dari *Twitter* yang telah dikumpulkan dan diolah, data tersebut opini mengenai investasi baik bentuk, jenis, maupun aplikasi investasi yang sedang banyak di perbincangkan di *Twitter*. Data tersebut diambil dari bulan Februari 2021-Ferburai 2022.
2. Data yang telah di peroleh kemudian di filtering menggunakan *Python*, data di beri label sentiment *tweet* positif dan negatif.
3. Tahap selanjutnya *preprocessing* tujuannya untuk *cleaning data*, dilakukan dengan beberapa tahapan yaitu, *case folding*, *filtering*, *tokenization*, *slang replacement* serta *stop word removal*.
4. Kemudian ekstraksi, tahap ini meliputi *Unigram (convert emoticon)*, *Negation*, *TF* dan *TD-IDF* (pembobotan).
5. Klasifikasi untuk mengelompokkan *tweet* sesuai label atau kelas yang ditentukan.

6. Label ditemukan, data atau tweet ini merupakan hasil klasifikasi data baru *Naïve Bayes*.



Gambar 1. Alur Penelitian

Adapun perancangan sistem yang akan dibuat terdapat pada Gambar 2.



Gambar 2. Desain perancangan sistem

HASIL DAN PEMBAHASAN

Aplikasi yang dibuat ini termasuk dalam *opinion mining*. *Opinion mining* digunakan untuk mengetahui suatu teks termasuk positif atau negatif. Proses yang digunakan untuk mencarinya dari tread twitter mengenai investasi. Setelah fitur tersebut pengelompokan dipilih menggunakan *Multinomial Naïve Bayes*. Pada model *multinomial* fitur vektor dokumen mengambil frekuensi dari kata-kata tidak hanya ada atau tidaknya kata tersebut. Sehingga setiap kata merupakan fitur penting. *Naïve Bayes* merupakan *supervised learning* sehingga membutuhkan data training. Pada proses *training* pelabelan dilakukan manual oleh peneliti.

Pengambilan data twitter.

Twitter menyediakan API untuk mengambil datanya. Untuk pengambilan data melalui API harus melakukan registrasi aplikasi pada halaman yang telah disediakan twitter. Twitter menerapkan beberapa batasan dalam melakukan konsumsi pada API-nya. Salah satunya batasannya adalah hanya diberikan tweet sampai batas 7 hari sebelum pengambilan data. Sehingga untuk tweet yang telah berumur 7 hari dari pengambilan data tidak bisa dilakukan. Batasan tersebut berlaku untuk melakukan pencarian dengan kata kunci tertentu. Untuk mengatasi hal itu teknik pengambilan data secara manual dengan mengambil *thread* langsung di twitter dan dikelola menggunakan excel. Data yang diperoleh adalah 188 dengan kata kunci “Investasi” data yang diambil dari rentang waktu Maret 2021 – Januari 2022. Kemudian dilakukan pembagian klasifikasi data negatif dan positif pada *thread* tweet dengan kata kunci “Investasi” sehingga data langsung dibagi menjadi 2 *sheet* yakni menjadi Positif dan Negatif.

Ekstraksi Fitur

Ekstraksi fitur ini termasuk diantaranya adalah melakukan proses pembersihan data tweet, pembersihan dari *stop word*, proses pencarian kata dasar, dan lain-lain. Hal ini dilakukan untuk mengurangi gangguan (*noise*) yang bisa mengurangi keakuratan. Beberapa hal yang dilakukan dalam penelitian ini adalah pengurangan pembersihan tweet dari hashtag, pungtuasi (kata gaul), *URL*, dan lain-lain.

Cleaning Data

Low case untuk menggantikan huruf kapital menjadi huruf kecil, *Removing Whitespace* untuk menghapus spasi antar kata, *Removing Number* untuk menghapus karakter angka, *Removing Punctuation* untuk menghapus bahasa gaul dan *emoticon*. *Cleaning data* bertujuan untuk meningkatkan keakuratan data saat pengujian. Dari hasil *cleaning* data diperoleh data *clean sentiment* positif dan data *clean sentiment* negative. Berikut hasil dari *cleaning* data yang telah dilakukan

```
[ 'maunambahinthebestinvestmentisinyourselbelajarduluapa itu investasipebenerpahamanngertibenahimindsetkrnyangnamayainvestasiitulongtermjangansampeturundikitlgstakutdancetpediambillihatdalamskalatahunabukanbulan', 'akudulumahasiswa skrgdudakherjarajinnabungdibibitkbulandiliatdoangnaikturunnyaterumasihbingungsebenarnyainginmanaharusdipelajarigayabiardapatuntunglebihlagidrin itapimayalahadatabunggangapernaturunygdrastisbgtsemogabeginiterus', 'notagreemawakunyaexplorinvestyangberisikojaduluyangkib ipenasarancryptosahampcrowdfundingforexatauapapunyangkamualebihbaguslagiinvestkeilmumasi mudajahagagalmasihbanyakntarmakin tuaburumakinkonservatifpilihannya', 'buatanaksekolahmahpalingbenerinvestasi buatasadepanadalahbelajarpunyauangrebumendingpakebu atbelibukubagusdaripadasehariantengingrafikahyangtunggasamperebumenngwaktunypakebuatkejarprestasidanilmumasadepancera hsahamk', 'ajaiblebihkeahamsihemglmccokbuatpemulabisabikinsenanjantungdikittpikalaujantungnyaudhbiasanenambelhocobabelajarb elajarduluyoutubenyandisenjayaadaplaylistngajarinsahmidisanacocokbuatpemula', 'kaloakulebihpreferkeinvestasiamulabelisa hamygadadilqatauidxdiajar kananalise teknikal dan fundamental yg sederhana sudah banyak mentorsaham yg bisadi jadikan referensiu ntuk investas ipeulamlulaidarimodalkecilajasupayacont', 'namanyajugainvestasiyangpentingstayinlongtermstanding', 'kalogyapreferkesahamsetadad isahamudahlaisejakakhirsempekarantahaumaspekanygpentingtambahuaterusan demangretur nyalumayandgdprtrnrbukand ipokayaknyagbakalholdselamngksampebisahidupdrdividen', 'ngaturkeuanganemgusahbgtsgumurujuudamulaisitopuatseengber lebih setidaknyamaugkmauharusdiinvestasiinyobaininvestkerdupdulujalansebulankokkurangantepindahkerdpttrusmikirkoknanggangakhirnyab eraniinvestkesaham', 'makasihbanyaktelahberbagiceritajadimakinsemangatsybuatmanagekeuangankarenapernahjugakemarinkerjabula ntaphasilnyaentahkemana semaukwkwalhamdulillahsekarangadakerjaantetaptujuanbuat savingjugaudhadatinggadidisiplnngjalaninyaa jah', 'investasinetworkingsihmagichangetakusamegaperlujuantapikliendatengteruskeakukarenarekomendasimulutkemulut', 'tapikala uinvestdiprodukemaserivatifikokkarenalowrisk', 'korbanfomoinvestasikriptoyangmudahangampakaiuangngingindanlongtermcukupa ruhiditwalleterjasepertibiasabukatahunatauhkemudi ankala utradingharus jagoteknikal analisisadharusbisakendaliknemosisertakuat mental', 'karenatradngadlahperpindahanuangdaritanganoranyangkurangsabar ketanganoranyanglebihsabarnakebetulandiakurangsabar mungkinjadigacupindahuangpindahalamjugapadahalanda idiauaanda idiamasihhidupmarketsekaranglagibull', 'mungkinmaikriptonyadipl apportakresimokanyabunuhdirixchangeresmidanexchangeterbersardiduniapunadadanamanorang sabarpastipuan', 'klrugikriptogabanyak bangetkrngangkinturunterustpanaikitujudidibalutkriptoygbslosstrsdbnkintresberakhirbnhdiri', 'akibatdaritradngggakdisertaid enganttechnicalmentalyememadai', 'tergantungklo di main leveragevailangkloinvestasidisipotamanajabijaklahsaatmainkriptomasadepanada dikriptobtcsaatuuangfiatgakadilainya', 'sayatimreksadanaobligasi', 'reksadanpendapatantetapsihbangsecararesikyorelatifsamata pireturnlebihdangpendapatantetap', 'sayasihrdpasaruangkarenangamungkinturupastinaikteruscumayaitulambatpergerakannyatapisa stihitung nabungtapinilainyabertambah', 'cryptocryptoclubikindicoinmanufacturingchef', 'sudahpastitothemoontothecroot', 'mauna mbahinthebestinvestmentisinyourselbelajarduluapa itu investasipebenerpahamanngertibenahimindsetkrnyangnamayainvestasiitulon gtermjangansampeturundikitlgstakutdancetpediambillihatdalamskalatahunabukanbulan', 'akudulumahasiswa skrgdudakherjarajinnabung gdibibitkbulandiliatdoangnaikturunnyaterumasihbingungsebenarnyainginmanaharusdipelajarigayabiardapatuntunglebihlagidrin itapima yanlahadatabunggangapernaturunygdrastisbgtsemogabeginiterus', 'buatanaksekolahmahpalingbenerinvestasi buatasadepanadalahbelajarp
```

Gambar 3. Hasil Cleaning Data

Tokenizing

Setelah *cleaning data*, maka hasil masing-masing data positif dan negatif di lakukan *tokenizing* untuk memisahkan kata. Setelah import data selanjutnya dilakukan tokenizing pada sentimen positif dan negative. Berikut hasil dari *Tokenizing* yang dilakukan

	kalimat	kalimat_token_negatif
0	reksadana manajer investasinya malah bikin kon...	[reksadana, manajer, investasinya, malah, biki...
1	bru mo ikut crypto tp trnyata itu riba gnti y...	[bru, mo, ikut, crypto, tp, trnyata, itu, riba...
2	jgn lupa manage resiko hidup jgn sampe hasil i...	[jgn, lupa, manage, resiko, hidup, jgn, sampe,...
3	yqng aneh jelas gak pernah tertarik too good t...	[yqng, aneh, jelas, gak, pernah, tertarik, too...
4	ga percaya sama pemerintah ujar masyarakat yg ...	[ga, percaya, sama, pemerintah, ujar, masyarak...

Gambar 4. Hasil Tokenizing

Filtering

Filtering merupakan penyaringan pada kalimat dengan menghapus kata-kata yang tidak diperlukan menggunakan *Stop words* modul dari NLTK. Berikut hasil dari *filtering* menggunakan modul *stop words*.

	kalimat	kalimat_token_negatif
0	reksadana manajer investasinya malah bikin kon...	[reksadana, manajer, investasinya, bikin, kons...
1	bru mo ikut crypto tp trnyata itu riba gnti y...	[bru, mo, crypto, trnyata, riba, gnti, pdahal,...
2	jgn lupa manage resiko hidup jgn sampe hasil i...	[jgn, lupa, manage, resiko, hidup, jgn, sampe,...
3	yqng aneh jelas gak pernah tertarik too good t...	[aneh, gak, tertarik, too, true, ngejar, iming...
4	ga percaya sama pemerintah ujar masyarakat yg ...	[ga, percaya, pemerintah, masyarakat, gampang,...
...
87	investor nangis darah si ini	[investor, nangis, darah, si]
88	ga pake aturan batas arb	[ga, pake, aturan, batas, arb]
89	mknnya ketika tmn gue rame maen kripto maen jud...	[mknya, tmn, gue, rame, maen, kripto, maen, ju...
90	gilee padahal udah jelasjelas ke tipu masih aj...	[gilee, udah, jelasjelas, tipu, aja, denial, n...
91	jika anda kehilangan uang di pasar saham hany...	[kehilangan, uang, pasar, saham, penyebabnya, ...

92 rows × 2 columns

Gambar 5. Hasil Filtering

Stemming

Stemming untuk mengubah kata yang memiliki makna yang sama ke bentuk dasar. Berikut hasil dari *stemming* yang telah dilakukan.

```
reksadana : reksadana
manajer : manaj
investasinya : investasinya
malah : malah
bikin : bikin
konsorsium : konsorsium
gorenggoreng : gorenggoreng
saham : saham
bru : bru
mo : mo
ikut : ikut
crypto : crypto
tp : tp
trnyata : trnyata
itu : itu
riba : riba
```

Gambar 6. Hasil Stemming

Pembobotan Kata

Pembobotan kata yang digunakan menggunakan TFIDF (*term frequency Inverse Document Frequency*). Pembobotan kata untuk mencari penting atau tidaknya kata tersebut dalam dokumen tersebut. Pembobotan

dilakukan dengan menghitung banyaknya kata yang muncul dalam dokumen dan dibandingkan dengan total kata dari seluruh dokumen. Berikut hasil TFIDF yang diperoleh

bs - 0.21088980028152832	(0, 77)	0.21088980028152832
sich - 0.21088980028152832	(0, 476)	0.21088980028152832
rugpull - 0.21088980028152832	(0, 460)	0.21088980028152832
keras - 0.19733576159562857	(0, 275)	0.19733576159562857
dalem - 0.21088980028152832	(0, 114)	0.21088980028152832
candle - 0.21088980028152832	(0, 85)	0.21088980028152832
sihh - 0.19733576159562857	(0, 477)	0.19733576159562857
kepolisian - 0.21088980028152832	(0, 273)	0.21088980028152832
dibuka - 0.1868224324151014	(0, 125)	0.1868224324151014
terrgoda - 0.1541650491438289	(0, 508)	0.1541650491438289
kesimpulannya - 0.21088980028152832	(0, 279)	0.21088980028152832
menghasilkan - 0.1868224324151014	(0, 341)	0.1868224324151014
kemakan - 0.19733576159562857	(0, 265)	0.19733576159562857
masing ² - 0.21088980028152832	(0, 327)	0.21088980028152832
begituan - 0.21088980028152832	(0, 46)	0.21088980028152832
mokat - 0.19733576159562857	(0, 355)	0.19733576159562857
ngejar - 0.1868224324151014	(0, 373)	0.1868224324151014
belasan - 0.17823241701025577	(0, 47)	0.17823241701025577
pusing - 0.3182581448497659	(0, 440)	0.3182581448497659
	(0, 209)	0.26404199010616713
	(0, 43)	0.1541650491438289
	(0, 212)	0.21088980028152832
	(0, 343)	0.21088980028152832
	(1, 452)	0.22039157606347098
	(1, 58)	0.18626279335488874

Gambar 7. Hasil TFIDF Kalimat Positif

bs - 0.21088980028152832	(0, 77)	0.21088980028152832
sich - 0.21088980028152832	(0, 476)	0.21088980028152832
rugpull - 0.21088980028152832	(0, 460)	0.21088980028152832
keras - 0.19733576159562857	(0, 275)	0.19733576159562857
dalem - 0.21088980028152832	(0, 114)	0.21088980028152832
candle - 0.21088980028152832	(0, 85)	0.21088980028152832
sihh - 0.19733576159562857	(0, 477)	0.19733576159562857
kepolisian - 0.21088980028152832	(0, 273)	0.21088980028152832
dibuka - 0.1868224324151014	(0, 125)	0.1868224324151014
terrgoda - 0.1541650491438289	(0, 508)	0.1541650491438289
kesimpulannya - 0.21088980028152832	(0, 279)	0.21088980028152832
menghasilkan - 0.1868224324151014	(0, 341)	0.1868224324151014
kemakan - 0.19733576159562857	(0, 265)	0.19733576159562857
masing ² - 0.21088980028152832	(0, 327)	0.21088980028152832
begituan - 0.21088980028152832	(0, 46)	0.21088980028152832
mokat - 0.19733576159562857	(0, 355)	0.19733576159562857
ngejar - 0.1868224324151014	(0, 373)	0.1868224324151014
belasan - 0.17823241701025577	(0, 47)	0.17823241701025577
pusing - 0.3182581448497659	(0, 440)	0.3182581448497659
	(0, 209)	0.26404199010616713
	(0, 43)	0.1541650491438289
	(0, 212)	0.21088980028152832
	(0, 343)	0.21088980028152832
	(1, 452)	0.22039157606347098
	(1, 58)	0.18626279335488874

Gambar 8. Hasil TFIDF Kalimat Negatif

Hasil Pengujian Multinomial Naïve Bayes

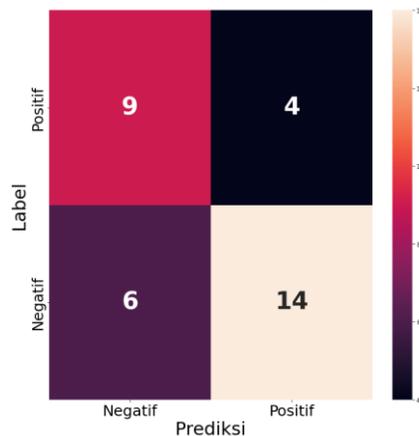
Setelah pembobotan selanjutnya pengujian hasil menggunakan metode *multinomial naïve bayes*. Berikut perintah yang dilakukan.

```

from sklearn.metrics import confusion_matrix, accuracy_score, classification_report, f1_score
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn import metrics
import seaborn as sns

x_train, x_test, y_train, y_test = train_test_split(words, inves['label'],
test_size=0.2, random_state=4)
#MultinomialNB
MNB=MultinomialNB()
MNB.fit(x_train,y_train)
predic = MNB.predict(x_test)
probabilitas = MNB.predict_proba(y_test)
print('Accuracy score: ', round(accuracy_score(y_test, predic), 2))
_, ax = plt.subplots(figsize=(12,12))
ax = sns.heatmap(confusion_matrix(y_test, predic),
annot=True, fmt='d', annot_kws={"size": 40, "weight": "bold"})
labels = ['Negatif', 'Positif']
ax.set_xticklabels(labels, fontsize=25);
ax.set_yticklabels(labels[::-1], fontsize=25);
# columns are the prediction
ax.set_ylabel('Label', fontsize=30);
  
```

Perintah diatas untuk memperoleh *confusion matrix* yang digunakan untuk perhitungan prediksi menggunakan *multinomial naive bayes*. Split data training sebesar 80% dan testing 20%. Berikut hasil dari perintah diatas.



Gambar 9. Hasil Prediksi Multinomial Naïve Bayes

Gambar diatas memperoleh hasil perhitungan secara detail sebagai berikut.

	precision	recall	f1-score	support
0	0.60	0.69	0.64	13
1	0.78	0.70	0.74	20
accuracy			0.70	33
macro avg	0.69	0.70	0.69	33
weighted avg	0.71	0.70	0.70	33
Accuracy score:	0.7			
F1 Score:	0.74			
Probabilitas:	[[6.27489823e-01 3.72510177e-01]]			

Dari hasil *multinomial naïve bayes* memperoleh *score precision* sebesar 0.69, *recall* sebesar 0.70, *f1-score (f-measure)* sebesar 0.74 dan *accuracy* sebesar 0.7. Probabilitas mengambil salah satu sampel kata yakni “investasi” sebesar 6.27489823e-01.

Pembahasan

Penelitian Sentimen Analisis Thread Investasi Pada *Twitter* Menggunakan *Naive Bayes* telah berhasil dilakukan dengan menggunakan bahasa pemrograman python v.3 dengan bantuan tools IDE Jupyter Notebook. Analisis data dimulai dari pengumpulan data thread twitter sebanyak 188 data dengan pembagian sentimen positif sebanyak 96 dan sentimen negatif sebanyak 92.

Data tersebut di proses dengan menggunakan teknik ekstraksi fitur yaitu *cleaning data*, *tokenizing*, *filtering* dan *stemming* data. Setelah ekstraksi fitur selanjutnya pembobotan TFIDF. Setelah pembobotan masuk ke metode *multinomial naïve bayes* dengan pembagian data training 80% dan testing 20%. Pengujian dibantu *tools confusion matrix* dengan perolehan *score precision* sebesar 0.69, *recall* sebesar 0.70, *f1-score (f-measure)* sebesar 0.74 dan *accuracy* sebesar 0.7. Probabilitas mengambil salah satu sampel kata yakni “investasi” sebesar 6.27489823e-01.

KESIMPULAN

Berdasar hasil penelitian yang dilakukan diperoleh nilai *score precision* sebesar 0.69, *recall* sebesar 0.70, *f1-score (f-measure)* sebesar 0.74 dan *accuracy* sebesar 0.7. Probabilitas mengambil salah satu sampel kata yakni “investasi” sebesar 6.27489823e-01. Klasifikasi opinion thread twitter dengan kata kunci “Investasi” memperoleh data sentimen positif sebanyak 96 dan sentimen negatif sebanyak 92. Dengan pembagian data (splitting) training 80% dan testing 20% pengujian menggunakan bantuan *tools confusion matrix*

TINJAUAN PUSTAKA

- [1] N. Haqqizar and T. N. Larasyanti, “Analisis Sentimen Terhadap Layanan Provider Telekomunikasi Telkomsel Di Twitter Dengan Metode Naïve Bayes,” *Pros. TAU SNAR-TEK 2019 Semin. Nas. Rekayasa dan Teknol.*, vol. 10, no. 2, pp. 1–15, 2019.
- [2] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [3] A. WIBOWO, “Pengaruh Pengetahuan Investasi, Kebijakan Modal Minimal Investasi, Dan Pelatihan Pasar Modal Terhadap Minat Investasi (Studi Kasus Mahasiswa Fe Unesa Yang Terdaftar Di Galeri Investasi Fe Unesa),” *J. Ilmu Manaj.*, vol. 7, no. 1, pp. 192–201, 2018.
- [4] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [5] R. Parlita, S. I. Pradika, A. M. Hakim, and K. R. N. M., “Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob,” *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, pp. 33–37, 2020.
- [6] M. G. Pradana, A. C. Nurcahyo, and P. H. Saputro, “Pengaruh Sentimen Di Sosial Media Dengan Harga Saham Perusahaan,” *Eduatic - Sci. J. Informatics Educ.*, vol. 6, no. 2, 2020, doi: 10.21107/edutic.v6i2.6992.