

PREDIKSI SENTIMEN INVESTOR PASAR MODAL DI JEJARING SOSIAL MENGUNAKAN *TEXT MINING*

Aestikani Mahani¹, Hendro Margono²

1,2) Universitas Airlangga

ABSTRACT

The decline in optimism for capital market investors is one of the financial impacts on the business world that arose from the SARS-COVID19 pandemic. This event was reflected in a decrease in trading volume followed by a sharp drop in the JCI on the Indonesia Stock Exchange starting March 2020. Thus, a slowdown in the economic recovery resulting from the pandemic is reflected in investor sentiment in the capital market. On the one hand, the rapid development of the internet in Indonesia has triggered the investor's activities in the information searching prior buy and sell securities, mostly use online platforms, which contribute to influencing investor preferences and sentiment. This study conducted a qualitative examination of the features/terms of stock investment in the capital market and collected them in a compact dictionary (lexicon). Therefore, lexicon-based investor opinion extraction was extracted from Twitter, followed by the text sentiment analysis, and forming a classification model based on Naive Bayes and Decision Tree. This research output shows that the polarity of capital market investor sentiment is optimistic with the sentiment features that often appear, namely "cuan", "bearish", "serok", "copet", "untung", "cut loss", and "nyangkut." Meanwhile, the Decision Tree classification model provides better performance.

Keywords : investor, lexicon, social network, stock exchange, text mining

Correspondence to : aestikani.mahani-2019@feb.unair.ac.id

ABSTRAK

Penurunan optimisme investor pasar modal adalah salah satu dampak keuangan pada dunia usaha yang timbul akibat pandemi SARS-COVID19. Hal ini tercermin dari turunnya volume perdagangan yang diikuti penurunan tajam IHSG di Bursa Efek Indonesia mulai Maret 2020. Sehingga kekhawatiran atas perlambatan pemulihan ekonomi sebagai dampak pandemi, tercermin dari sentimen investor di pasar modal. Di satu sisi, perkembangan internet di Indonesia yang pesat, memicu kecenderungan aktivitas investor dalam pencarian informasi sebelum membeli dan menjual surat berharga secara *online*, turut berkontribusi dalam mempengaruhi preferensi dan sentimen investor. Penelitian ini menggali ekspektasi investor yang tercermin pada sentimen investasi, dimana pasar modal sebagai salah satu barometer penting perekonomian suatu negara. Kajian ini mengeksplorasi fitur/terms investasi saham yang kerap muncul di pasar modal dan mengumpulkannya dalam kamus leksikon. Kemudian, dilakukan ekstraksi opini investor berbasis leksikon yang digali dari jejaring sosial Twitter, dilanjutkan dengan tahap *text mining* yaitu menganalisis sentimen, dan membentuk model klasifikasi berbasis Naive Bayes dan *Decision Tree*.

Keluaran penelitian ini menunjukkan bahwa polaritas sentimen investor pasar modal adalah positif dengan fitur sentimen yang sering muncul yaitu “cuan”, “*bearish*”, “serok”, “copet”, “untung”, dan “*cut loss*”. Sedangkan model klasifikasi *Decision Tree* memberikan performansi akurasi yang lebih baik.

Kata Kunci : Analisis sentimen; Investor; Leksikon; *Text mining*; Twitter

Riwayat Artikel:

Received : 22 Januari 2021

Revised : 15 Februari 2021

Accepted : 25 Juni 2021

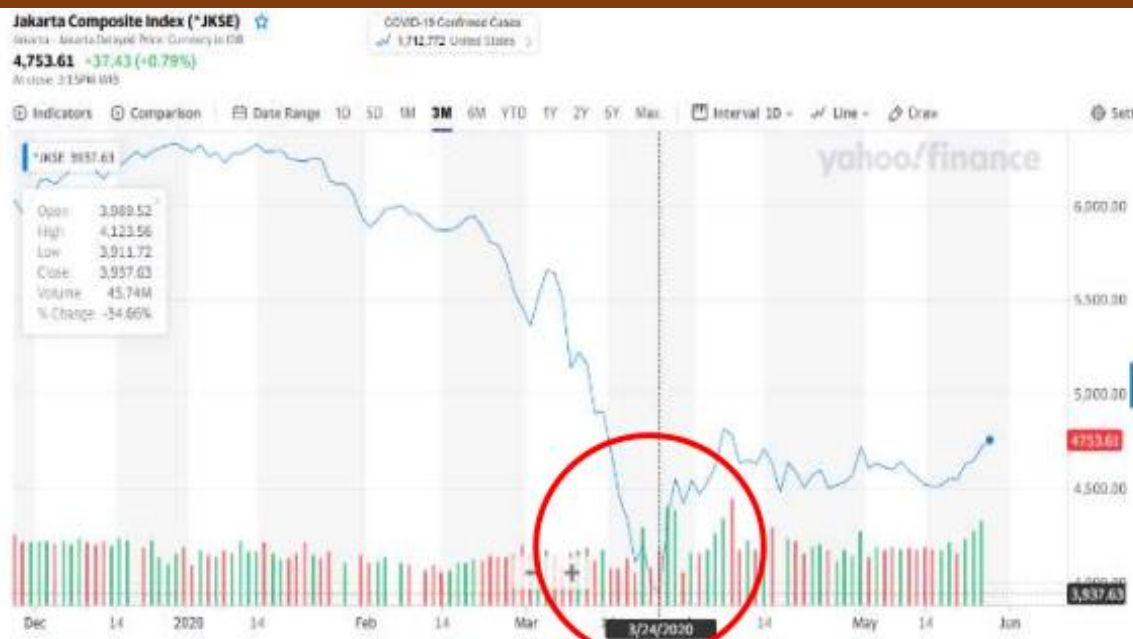
PENDAHULUAN

Turunnya optimisme investor pasar modal adalah salah satu dampak keuangan pada dunia usaha yang timbul akibat pandemi CoVID-19. Hal ini tercermin dari turunnya volume perdagangan yang diikuti turunnya nilai IHSG di Bursa Efek Indonesia medio Maret 2020.

Namun tidak hanya itu saja, kekhawatiran atas perlambatan pemulihan ekonomi sebagai dampak pandemi, akan tercermin dari persepsi dan opini investor atas pasar modal. Opini investor pada akhirnya akan membentuk gambaran harapan terhadap *recovery* roda perekonomian, dimana pasar modal adalah salah satu barometernya. Pola perilaku dan sentimen investor ini secara eksplisit maupun implisit dapat ditangkap dari *posting* di media sosial, mesin pencari, maupun forum investasi baik yang berbasis platform sekuritas atau *cloud based group messaging application*.

Studi kualitatif dengan pendekatan *grounded theory* dilakukan oleh (Hridoy, 2015) untuk mengeksplorasi bagaimana 26 *online investor* mencari dan menggunakan informasi, serta peranan jejaring sosial dalam

konteks yang lebih besar. Hasil kajian ini menunjukkan bahwa selain berpartisipasi pada komunitas *virtual* publik, mereka juga berpartisipasi pada jejaring sosial *private* yang memberi benefit berupa informasi dan pengetahuan baru. Kemudian *text mining* (ekstraksi dan klasifikasi) untuk menganalisis sentimen investor bursa efek China atas *return* saham dalam Index CSI 300 menunjukkan hasil bahwa terdapat faktor pengaruh akumulasi sentimen pada *return* harga saham (Shi, 2019). Sedangkan efek penularan rumor dan meningkatnya risiko investor dalam keterkaitannya dengan harga saham disimulasikan menggunakan SCIR model dalam jaringan bi-layer berpasangan sehingga diketahui dampaknya baik secara global maupun lokal (Dong, 2019).



Gambar 1.

IHSG Bursa Efek Indonesia Terendah di Angka 3937,63 pada Tanggal 24 Maret 2020

Keuangan Keperilakuan

Menurut Statman (2019) bahwa keuangan berperilaku adalah sebuah pendekatan alternatif dari keuangan baku dengan beberapa perbedaan. Pertama, dalam teori keuangan standar manusia dianggap rasional, sedangkan dalam keuangan berperilaku manusia dianggap “normal”, dalam arti manusia tidak sepenuhnya rasional. Manusia memiliki emosi dan manusia tidak dapat merencanakan sepenuhnya kapan menggunakan rasio, kapan menggunakan emosi dan kapan menggunakan keduanya secara bersama-sama. Kedua, dalam teori keuangan standar diasumsikan bahwa pasar modal bersifat efisien (*efficient market theory*). Sebaliknya dalam keuangan berperilaku pasar modal diasumsikan tidak efisien, meskipun diakui tidak mudah untuk ditaklukkan dan diadap. Harga saham emiten dimungkinkan menyimpang dari nilai fundamentalnya disebabkan oleh berbagai hal, termasuk berbagai variabel psikologisnya. Ketiga, menurut teori keuangan baku, investor diasumsikan akan membentuk bauran portofolio berdasarkan kriteria *mean-variance* yang dikemukakan oleh Markowitz. Sebaliknya dalam teori keuangan

keperilaku investor membentuk portofolio berdasarkan aturan yang disebut teori portofolio berperilaku (*behavioural portfolio theory*). Teori portofolio investasi menyatakan bahwa tingkat pengembalian yang diharapkan (*expected return*) dan risiko mencerminkan dua kondisi yang berbanding lurus. Hal ini menunjukkan bahwa semakin tinggi tingkat pengembalian yang diharapkan oleh seorang investor, maka semakin besar pula tingkat risiko yang akan dihadapinya. Karena adanya faktor risiko yang melekat dalam setiap unit *expected return* yang diharapkan, maka seorang investor perlu melakukan analisis secara komprehensif sebelum mengambil suatu keputusan investasi. Keempat, dalam teori keuangan standar, hasil yang diharapkan diukur dengan menggunakan model harga aset atau CAPM (*Capital Assets Pricing Model*), dalam hal ini risiko diukur dengan beta, dan risiko merupakan satu-satunya faktor penentu. Pada teori keuangan berperilaku, hasil yang diharapkan diukur dengan menggunakan *Behavioral Assets Pricing Model*, dalam hal ini hasil yang diharapkan merupakan fungsi dari berbagai variabel berperilaku investor.

Dalam konsep keuangan berperilaku, dikenal *limited arbitrage* dimana saham pada

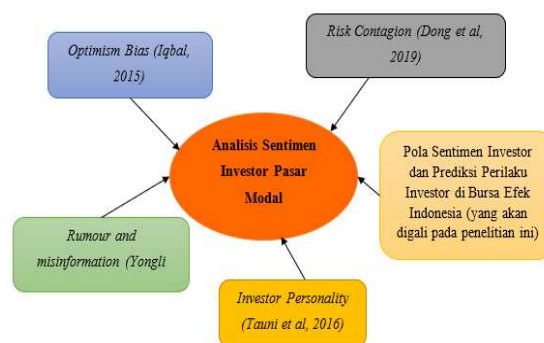
dasarnya mengandung risiko. Arbitrase menjadi terbatas karena dalam dunia investasi yang nyata, jika terjadi *downtrend* harga tidak akan kembali ke tingkat yang sesuai dengan nilai fundamentalnya seketika itu juga. Fakta akan *limited arbitrage* ini membuat asumsi pasar efisien menjadi sulit untuk dipenuhi dan keuangan berperilaku menjadi semakin mampu menjelaskan realitas yang ada di lapangan. Adanya sentimen investor di pasar modal mendukung teori *limited arbitrage*. Secara umum sentimen diartikan sebagai suasana hati (*mood*) yang sedang dialami seseorang. Sentimen inilah yang akan membentuk keyakinan, prediksi dan akhirnya permintaan para investor terhadap surat berharga. Dalam bahasan utama keuangan perilaku dan *neurofinance* berfokus pada bagaimana emosi masuk ke dalam persepsi risiko trader dan investor. Emosi memainkan peran kunci dalam mendorong pengambilan risiko. Emosi yang paling sering diidentifikasi sebagai kontribusi terhadap kerapuhan finansial dan kegembiraan irasional termasuk keserakahan, harapan dan ketakutan (Shefrin, 2002). Risiko merupakan sebagai bagian dari perasaan dan berpendapat bahwa risiko yang dirasakan *trader* dan investor selama transaksi keuangan mereka dapat dicirikan sebagai perasaan takut - dengan efek umpan balik dan mengintensifkan tanggapan ketakutan dan memicu kepanikan (Slovic & Kerjan, 2010). Emosi lain akan mempengaruhi keputusan keuangan secara lebih luas, kegembiraan irasional yang dialami *trader*, yang biasa terlihat di pasar *bullish*, mungkin mencerminkan interaksi antara harapan dan keserakahan (Shiller, 2003).

Dari penelitian tersebut, Akerlof (1970) mengidentifikasi bahwa ketika pembeli tidak memiliki informasi terkait spesifikasi produk dan hanya memiliki persepsi umum mengenai produk tersebut, maka pembeli (*buyer*) cenderung menilai semua produk di harga sama, meskipun produk *high-end* maupun *low-end*, sehingga merugikan penjual produk berkualitas tinggi.

Spence (1974) juga menemukan bahwa *cost of signal* pada berita buruk akan lebih tinggi dari pada berita baik dan perusahaan

dengan isu buruk mengirimkan sinyal yang tidak kredibel. Hal tersebut memotivasi manajer untuk mengungkapkan informasi pribadi untuk mengurangi asimetri informasi dengan harapan dapat mengirimkan sinyal yang baik (*good news*) tentang kinerja perusahaan ke pasar.

Jika pengumuman informasi tersebut dianggap sebagai sinyal baik, maka investor akan tertarik untuk melakukan perdagangan saham, dengan demikian pasar akan bereaksi yang tercermin melalui perubahan dalam volume perdagangan saham maupun kenaikan harga saham (Suwardjono, 2010). Berikut adalah model pengaruh sentimen investor yang akan digali dalam penelitian ini adalah:



Gambar. 2
Model Pengaruh Sentimen Investor

Sumber : Data diolah Peneliti, 2021

Jejaring Sosial

Media sosial atau disebut juga dengan jejaring sosial adalah sekumpulan aplikasi web, berdasarkan teknologi *framework web 2.0* dan budaya, yang memungkinkan sejumlah orang terkoneksi serta berkolaborasi dengan orang maupun komunitas lain dengan menciptakan, menyusun, membuat, dan membagikan konten digital. Secara kolektif, penggunaan media sosial telah mengalihkan lokus kontrol untuk komunikasi publik dari organisasi besar menjadi organisasi yang dibagikan dengan pengguna individu. Tidak hanya organisasi saja yang menyiarkan satu pesan ke khalayak massa, sejumlah besar percakapan mungkin terjadi di antara

sejumlah orang dan organisasi. (Turban, 2018).

Data dan Text Mining

Data mining adalah proses penggunaan berbagai alat analisis data untuk menemukan data dalam semua bentuk dan jenisnya yang dapat digunakan untuk membuat prediksi yang valid (Han, 2012). *Data mining* mampu menjawab pertanyaan tentang masa lalu (apa yang telah terjadi) dengan mengekstrak dan menganalisis data lama. Untuk memanfaatkannya dan mengambil pelajaran dari hasil sebelumnya dalam pengambilan keputusan, masa kini (apa yang terjadi) dengan mengekstraksi data baru saat ini tersedia dan membandingkannya dengan data lama dan menganalisisnya, dan masa depan (apa yang mungkin terjadi) seperti meramalkan suatu kejadian yang akan datang sebagai akibat adanya data yang terus diulang dengan menganalisis data yang tersedia tersebut. Analisis dan penggalian informasi dari data dan identifikasi hubungan 'tersembunyi' dalam kumpulan data besar. Hal ini akan memungkinkan informasi yang sebelumnya tersembunyi dan tidak terekspos sekarang menjadi terekspos. Sehingga memberikan lebih banyak dukungan dalam proses pengambilan keputusan (Nemati H.R, 2005).

Sentiment analysis

Analisis sentimen sering juga disebut dengan *opinion mining* atau *content analysis* adalah proses pengelompokan teks/serangkaian kata yang mengandung nada emosi yang diterapkan untuk memperoleh pemahaman tentang sikap, opini, pendapat, dan emosi yang diungkapkan dalam percakapan *online*. Umumnya, sentimen yang terekam dikategorikan menjadi 3 yaitu sentimen positif, sentimen negatif, dan netral. Untuk mengumpulkan data, peneliti harus mengumpulkan konten yang berisi (kumpulan) kata kunci atau fitur yang ditentukan secara manual. Ini disebut target. Analisis mungkin ingin menambahkan konteks

ke analisis yang dilakukan pada kata kunci tertentu dengan menambahkan fitur/kata kunci yang berlawanan atau spesifik. (Makhabel, 2017).

Sehingga model sederhana yang dapat digunakan untuk memahami *text sentiment analysis* ini adalah:

$Sentiment = \{datasource, source, target, sentiment, polarity\}$

Sumber : (Makhabel, 2017)

Tujuan dari analisis opini yaitu untuk mengklasifikasikan atau mengukur polaritas dokumen dalam kumpulan besar data atau dokumen. Skema warna, seperti dicontohkan pada Gambar 2, dapat dipilih berdasarkan spektrum ukuran polaritas, dan warna yang berbeda, seperti merah dan hijau, dapat digunakan untuk memisahkan sentimen. Angka ini memberikan awan kata (*wordcloud*) terpisah untuk sentimen positif dan negatif, dengan positif dalam warna hijau dan negatif dengan warna merah. Jika sentimen netral juga disertakan, itu bisa diplot dalam warna kuning untuk mengikuti tema lampu lalu lintas (Anandarajan, 2019).

Menampilkan *wordcloud* secara berdampingan memungkinkan pembaca melihat gambaran besar dari analisis. penggunaan *word cloud* dapat dibuat secara umum atau disesuaikan sesuai pilihan peneliti. Misalnya, kita dapat membuat *word cloud* untuk ulasan *online*. Beberapa contoh penggunaan analisis sentimen secara luas yaitu pada strategi pemasaran, pengukuran RoI dari kampanye pemasaran, pengembangan produk, dan perbaikan layanan pelanggan.

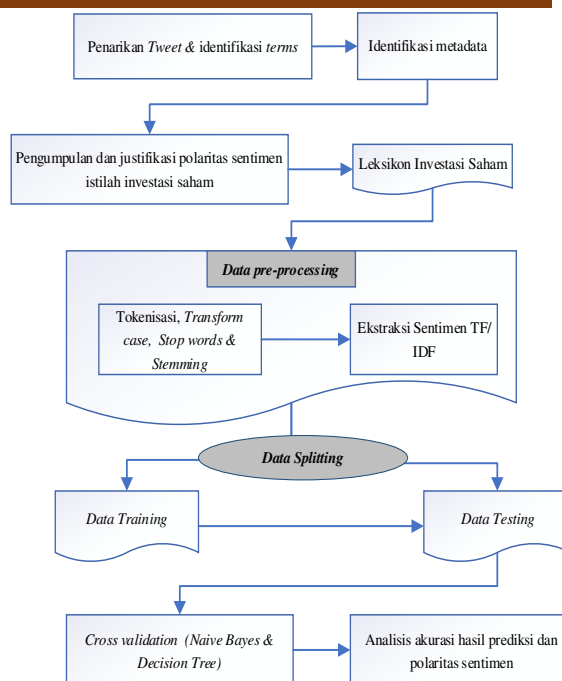
Contoh *Word Cloud* dari Polaritas Opini Positif dan Negatif

Sumber : (Anandarajan, 2019)

Beberapa metode yang digunakan untuk menganalisis sentimen yaitu metode *rule-based* menggunakan SentiwordNet dan metode *Machine Learning*. Algoritma klasifikasi akan melakukan klasifikasi atas *learning data* dengan cara menganalisis *training data*. Pendekatan ini terdiri dari beberapa metode yang banyak dipakai oleh peneliti yaitu : *Support Vector Machine* (SVM), Naive Bayes, dan *Decision Tree*. Secara sederhana, algoritma Naive Bayes mengasumsikan bahwa keberadaan fitur tertentu dalam kelas tidak terkait dengan keberadaan fitur lainnya. Algoritma *Decision Tree* berfungsi seperti diagram alir, memisahkan titik data menjadi dua kategori serupa sekaligus dari "batang pohon" ke "cabang", ke "daun", di mana kategorinya menjadi lebih mirip. Menciptakan kategori dalam kategori, memungkinkan klasifikasi organik dengan pengawasan manusia terbatas.

Pengumpulan Data

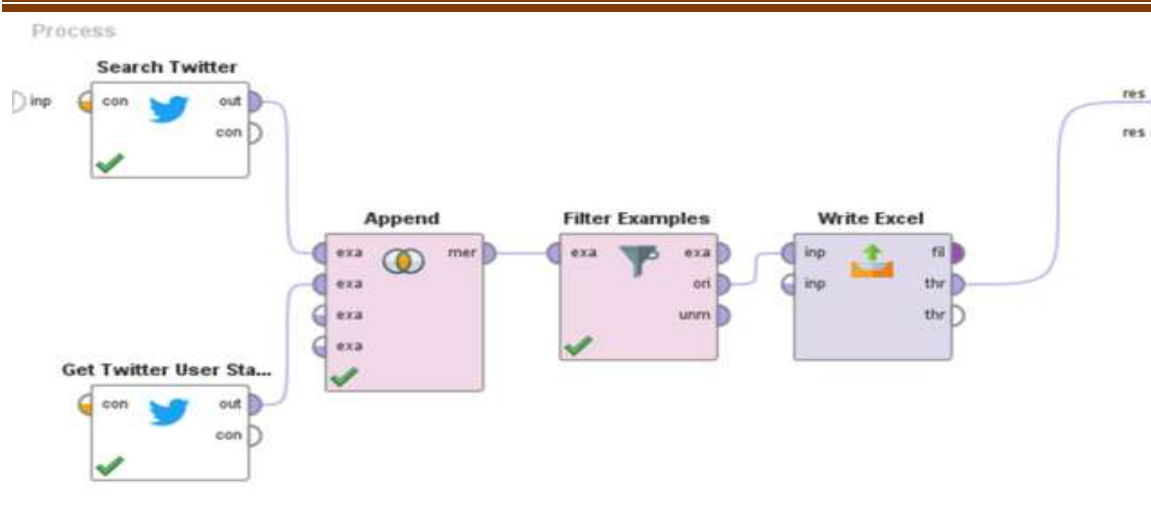
Berikut adalah diagram alur proses *Sentiment analysis* yang akan dibahas dalam penelitian ini dimulai dari data *scraping* di jejaring sosial Twitter periode Maret-Desember 2020, tahap menghimpun fitur sentimen dalam sebuah leksikon, ekstraksi sentimen berbasis leksikon, hingga klasifikasi sentimen :



Tahapan Identifikasi Sentimen, Ekstraksi, dan Klasifikasi Sentimen

Sumber : Data diolah Peneliti, 2021

Pada tahap persiapan data dilakukan *scraping* data *tweet* dari jejaring sosial Twitter menggunakan RapidMiner API. Kemudian *raw data* disimpan dalam file excel serta mengidentifikasi atribut dan label metadata. Secara paralel, dilakukan sampling justifikasi polaritas sentimen secara kualitatif, dengan mengambil sampel 32 *tweet* untuk dinilai oleh 100 responden investor dan *trader* aktif. Kemudian mengidentifikasi *terms*/fitur investasi saham sebagai bahan pembuatan leksikon. Hasil keluaran ini berupa leksikon dan nilai polaritas kemudian disimpan dalam format .CSV. Berikut ini adalah diagram operasi pengambilan *raw data* :

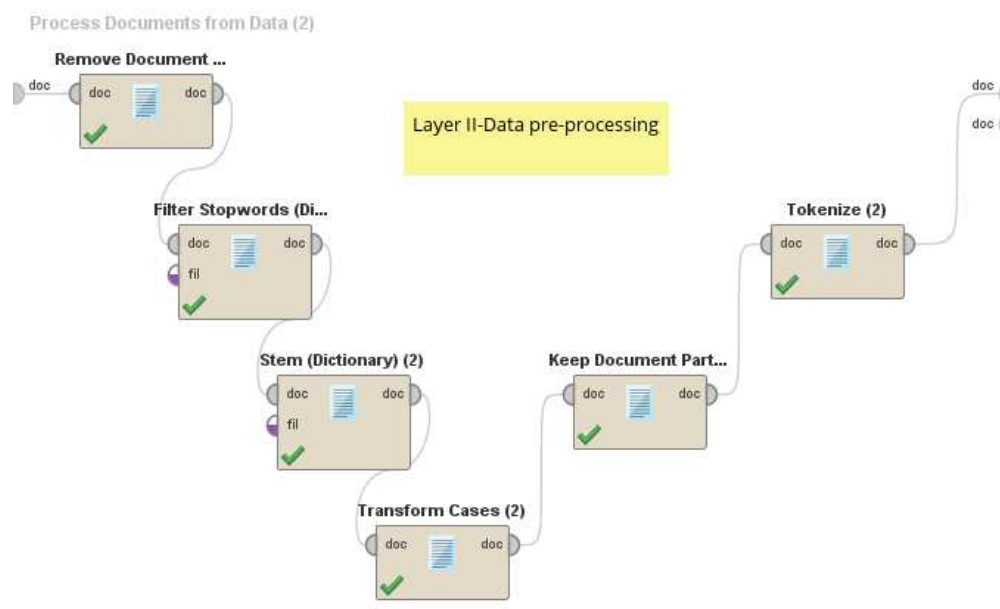


Gambar. 5
 Alur Operasi *Scraping* Raw Data berupa *Tweet*

***Pre-processing*, Transformasi, dan Ekstraksi**

Kemudian pada data yang sudah ditransformasi diberi pelabelan untuk membuat data *training* dengan memberi label klasifikasi berisi nilai dan kelas sentimen “positif”, “netral” dan “negatif”. Hasil keluaran dari proses ini disimpan format Excel. Kemudian dilakukan *data-preprocessing* meliputi tokenisasi (pemisahan kata per kata dalam 1 *tweet*), *transform case* (menyeragamkan huruf besar dan kecil), eliminasi *stopwords* (kata umum yang tidak bermakna seperti kata penghubung, kata keterangan dan tanda baca), filterisasi

stemwords (asal mula kata untuk kata yang berimbuhan), dan filterisasi sentimen terms yang terhimpun dalam leksikon. Proses ini dilakukan baik pada *data training* maupun pada *data testing*.



Gambar. 6

Tahap Data Pra-processing pada RapidMiner

Sumber : Data diolah Peneliti, 2021

Berikut ini contoh hasil pemrosesan teks dari tahap *data pre-processing* :

Tweet Sample	Lhaaa mumpung harganya bagus ini serok deh !
Transform case	lha mumpung harganya bagus ini serok deh !
Tokenisasi	lha mumpung harganya bagus ini serok deh
Eliminasi Stop words	mumpung harganya bagus serok
Stem Dictionary	harga bagus serok

Gambar. 7

Contoh Hasil Data Pre-Processing

Sumber : Data diolah Peneliti, 2021

Klasifikasi dan Analisis Sentimen

Setelah data dibersihkan dalam *pre-processing*, langkah selanjutnya adalah ekstraksi sentimen dan melakukan proses klasifikasi atas *testing data* menggunakan training data set. Untuk itu dialokasikan sebanyak 32 *training data*. Sentimen diekstraksi menggunakan leksikon yang diunggah ke dalam proses/operasi “*Extract Sentiment*”. Secara paralel, dilakukan proses

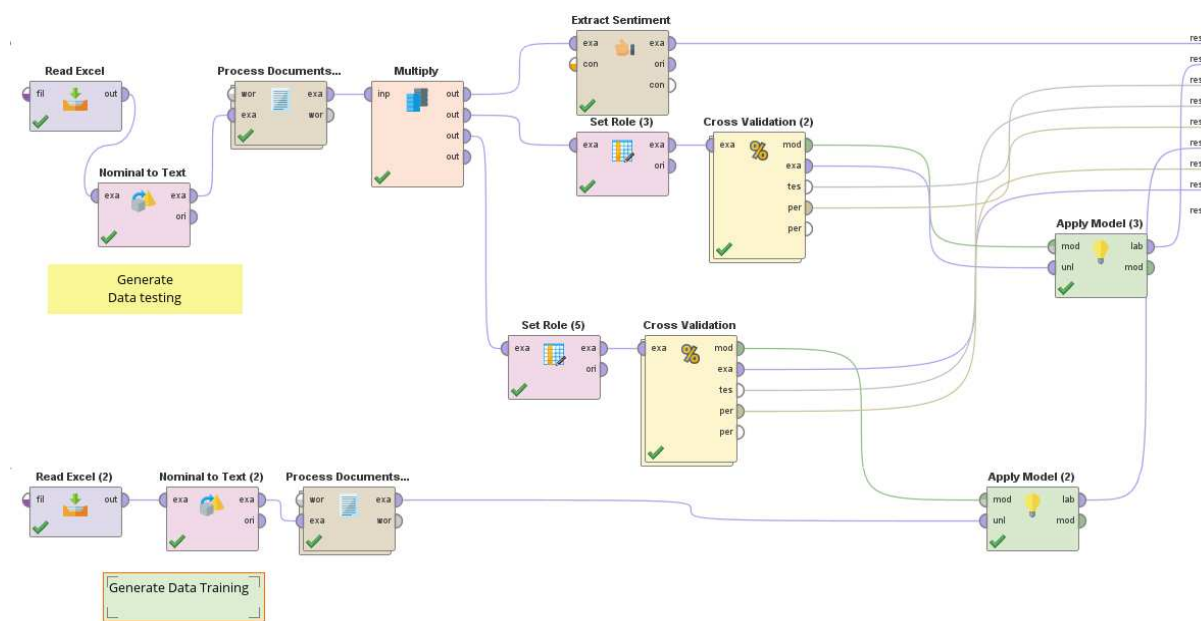
klasifikasi sentimen menggunakan algoritma Naive Bayes dan *Decision Tree*. Perbedaan dari kedua jenis metode klasifikasi ini adalah pada cara memperlakukan *identifier/classifier*. Naive Bayes mengasumsikan bahwa fitur (*terms*) memiliki nilai yang independen pada posisi fitur(*terms*) lain. *Decision Tree* berfungsi seperti diagram alir, memisahkan titik data menjadi dua kategori serupa sekaligus dari "batang pohon"

ke "cabang", ke "daun", di mana kategorinya menjadi lebih mirip. Ini menciptakan kategori dalam kategori, memungkinkan klasifikasi organik dengan pengawasan manusia terbatas. Proses ini melibatkan pengulangan setiap subset yang diperoleh secara rekursif, yang dikenal sebagai partisi rekursif (Han, 2012). Proses rekursi dilakukan jika *subset* pada *node* memiliki nilai yang identik dari variabel target, atau jika tahap pembagian tidak lagi menghasilkan nilai dan menambahkannya ke prediksi.

PEMBAHASAN

Hasil penelitian untuk menjawab rumusan masalah penelitian dibagi menjadi 2 aspek yaitu: ekstraksi sentimen dan klasifikasi *tweet* pada *testing data*.

Secara global, model analisis sentimen investor yang berupa proses utama di RapidMiner adalah sebagai berikut :



Gambar. 8

Model Analisis Sentimen Investor dan Proses Klasifikasi

Sumber : Data diolah Peneliti, 2021

Menghimpun istilah investasi saham ke dalam leksikon dan identifikasi nilai polaritas. Leksikon investasi saham sebagaimana diteliti sebelumnya (Prayogi, 2018) disempurnakan dengan identifikasi polaritas menggunakan metode kualitatif tematik dan penambahan istilah yang belum teridentifikasi.

1. Berikut adalah 35 istilah investasi yang terhimpun dalam leksikon:

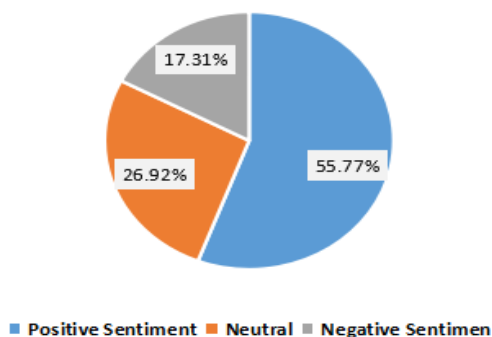
TABEL 1. Fitur Sentimen Investor dan Polaritas dalam Leksikon

Terms identifier	Polarity	Value
bandar	N	0
gorengan	N	0
haka	N	0
haki	N	0
hajar	N	0
copet	N	0
pompom	N	0
scalping	N	0
gocek	N	0
menu	N	0
cuan	+	1
profit	+	1
untung	+	1
bullish	+	1
belanja	+	1
diskon	+	1
guyur	+	1
hijau	+	1
ijo	+	1
lebar	+	1
luber	+	1
maknyos	+	1
mantul	+	1
masuk	+	1
serok	+	1
tiktok	+	1
TP	+	1
nanjak	+	1
panen	+	1
kotos	+	1
bearish	-	-1
cut loss	-	-1
CL	-	-1
nyangkut	-	-1
merah	-	-1

Sumber : Diolah Peneliti, 2021

- Ekstraksi sentimen dan klasifikasi kelas pada *testing data*. Berikut ini adalah hasil ekstraksi sentimen :

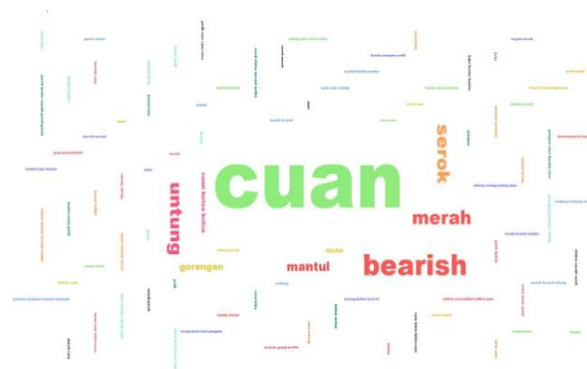
Hasil Klasifikasi Sentimen



Gambar. 9

Hasil Klasifikasi Sentimen pada *Data Testing*

Diketahui sentimen positif adalah dominan dengan proporsi sebesar 55,77%, sedangkan sentimen paling kecil dengan proporsi 16,98%. Frekuensi fitur sentimen ini juga diperkuat pada grafik *wordcloud* sebagai berikut dimana istilah “Cuan”, “*bearish*”, “serok”, “copet”, “untung”, dan “*cut loss*” ; memiliki frekuensi kemunculan paling besar :



Gambar. 10

Wordcloud Ekstraksi Sentimen

Sumber : Data diolah Peneliti, 2021

Frekuensi fitur sentimen yang sering muncul adalah :

TABEL 2. Hasil Ekstraksi Fitur Sentimen yang Sering Muncul

No. Fitur	Fitur Sentimen	Terms Frequent	Fraksi
1	Cuan	27	0.103846
2	Bearish	17	0.065385
3	Serok	12	0.046154
4	Copet	9	0.034615
5	Untung	8	0.030769
6	Cutloss	8	0.030769

Sumber : Data diolah Peneliti, 2021

Berikut ini adalah hasil klasifikasi sentimen menggunakan metode Naive Bayes dan *Decision Tree* :

TABEL 3. Nilai Akurasi Klasifikasi untuk Setiap Kelas Sentimen

Kelas Label	Parameter	Naïve Bayes	Decision Tree
Prediksi Kelas Positif	Akurasi	55.38%	55%
	Presisi	55.60%	59.98%
	Recall	100%	99.13%
Prediksi Kelas Negatif	Akurasi	0%	0.40%
	Presisi	0%	50%
	Recall	0%	2.27%
Prediksi kelas Sentimen Netral	Akurasi	3.80%	6.50%
	Presisi	100%	94.44%
	Recall	1.39%	23.61%
Perbandingan Akurasi Metode		55.77%	61.92%
Perbandingan Presisi Metode		51.87%	68.00%
Perbandingan Recall Metode		33.80%	41.73%

Sumber : Data Diolah Peneliti, 2021

Dapat diketahui bahwa akurasi dari metode *Decision Tree* adalah 61,92% dengan toleransi +/- 2,63%. Tingkat presisi dan *recall* dari metode *Decision Tree* juga lebih tinggi dari metode *Naive Bayes*.

Dari hasil performansi klasifikasi diatas, diketahui bahwa metode *Decision Tree* lebih akurat karena memiliki nilai presisi yang lebih tinggi. Perbedaan hasil ini dapat dipengaruhi oleh *dataset* maupun jumlah *testing data*.

KESIMPULAN

Hasil ekstraksi menunjukkan sentimen positif merupakan sentimen yang banyak muncul dengan istilah “Cuan”, “*bearish*”, , “serok”, “copet”, “untung”, dan “*cut loss*” merupakan istilah yang sering dimuat dalam cuitan jejaring sosial.

Sedangkan dari upaya klasifikasi, metode *Decision Tree* menunjukkan performansi yang lebih baik dibanding *Naive Bayes*. Sehingga *insight* akhir yang diperoleh dari penelitian ini adalah investor pasar modal menunjukkan sentimen positif atas aktivitas investasi mereka di pasar modal walaupun berada dalam kondisi *downtrend/bearish* pada periode pengambilan data yaitu Maret 2020 hingga Desember 2020. Namun sentimen

positif yang ditunjukkan adalah sebagai harapan bahwa kondisi akan perekonomian akan pulih kembali. Aktivitas investasi juga dipandang investor sebagai *opportunity* bahwa akan *diperoleh return of investment* setelah kondisi *bearish* terlewati.

DAFTAR PUSTAKA

- Abdillah, J., & Handoko, H. . (2012). Motivation and Consequence of Individual's Involvement in Social Network Sites : A Study of Social Computing of Inter Collectivist-Individualist Cultural Value. *Jornal of Economy and Business*, 27, 31.
- Adamopoulos, D. B., V. (2013). Exploring the Social Networks of Online Investors. Paper presented at the 24 th Australasian Conference on Information Systems, Melbourne.
- Akerlof, G. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, No. 3, 84, 14.
- Anandarajan, M., Hill, & Nolan, T. (2019). Practical Text Analytics: Maximizing the Value of Text Data. In Vol. 2. R. Sharda (Ed.). doi:https://doi.org/10.1007/978-3-319-95663-3
- Bahasa, B. P. d. P. (2020, Jan 2021). Kamus Besar Bahasa Indonesia -KBBI-Kamus versi daring. Retrieved from https://kbbi.web.id/sentimen
- Bar-Haim, D., E., Feldman, R., Fresko, & Goldstein, G. (2011). Identifying and Following Expert Investors in Stock Microblogs. Paper presented at the Conference on Empirical Methods in Natural Language Processing, Edinburgh.
- Brown, E. (2012, January 2012). Will twitter make you a better investor? A look at sentiment, user reputation and Their Effect on Stock Market. Paper presented at the Proceedings of the

- Southern Association for Information Systems Conference, Atlanta, GA, USA.
- Chiarello, F., Bonaccorsi, A., & Fantoni, G. . (2020). Technical Sentiment Analysis: Measuring Advantages and Drawbacks of New Products Using Social Media. *Computers in Industry*(103299), 12. doi:<https://doi.org/10.1016/j.compind.2020.103299>
- Christopoulos, G., Tobler, P., Bossaerts, Dolan, & Schultz, W. (2009). Neural Correlates of Value, Risk, and Risk Aversion Contributing to Decision Making under Risk. *The Journal of Neuroscience*. doi:10.1523/JNEUROSCI.2614-09.2009
- Ciccone, S. J. (2011). Investor Optimism, False Hopes and the January Effect. *Journal of Behavioural Finance*, pp. 158-168,, 12, 158.
- Creswell, J. C., P.V. (2018). Designing and Conducting Mixed Method Research. In.
- Dong, Y., Wang, J., Chen, T. (2019). Price Linkage Rumors in the Stock Market and Investor Risk Contagion on Bilayer-Coupled Networks. *Complexity*, 21. doi:<https://doi.org/10.1155/2019/4727868>
- Duan, J. Z., Jianping. (2013). Mining Opinion and Sentiment for Stock Return Prediction based on WebForum Messages. 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 984.
- Dube, R. (2021). Characteristics of Social Networks. Retrieved from https://socialnetworking.lovetoknow.com/Characteristics_of_Social_Networks
- Felton, J., Gibson, & Sanbonmatsu. (2003). Preference for Risk in Investing as a Function of Trait Optimism and Gender. *Journal of Behavioural Finance*.
- Han, J. K., Micheline ; Pei Jian. (2012). *Data Mining, Concepts and Techniques* (3 ed.). USA: Elsevier.
- Hoffman, M., G. (1999). Problem with Peirce's Concept of Abduction *Journal of Foundation Science*, 4, 33. doi:<https://doi.org/10.1023/A:1009675824079>
- Hridoy, S. A., Ekram, M. T., Islam, Mohammad, Faysal, A., Rahman, Rashedur. (2015). Localized twitter opinion mining using sentiment analysis, *Decision Analytics. ECONSTOR*, 2(1), 1-19. doi:<http://dx.doi.org/10.1186/s40165-015-0016-4>
- Ignatouw, G. M. (2018). An Introduction to Text Mining: Research Design, Data Collection, and Analysis. In H. Salmon (Ed.). Retrieved from <https://study.sagepub.com/introtextmining>
- Iqbal, N. (2015). Impact of Optimism Bias on Investment Decision: Evidence from Islamabad Stock Exchange, Pakistan. *Research Journal of Finance and Accounting*, 9(19), 6. doi:<https://core.ac.uk/download/pdf/234631087.pdf>
- Jefferson, J. S., N. (2014). *Shopping Saham Modal Sejuta* (2 ed.). Jakarta: Elex Media Komputindo.
- JKSE. (2020). Yahoo Finance. Retrieved from <https://finance.yahoo.com/quote/%5EJKSE/history?period1=1580515200&period2=1590710400&interval=1d&filter=history&frequency=1d>
- Kuhnen, C. K., B. (2005). The Neural Basis of Financial Risk Taking. *Neuron*, 47, 763-770.
- Kumar, S. R., V. (2016). A survey of the applications of text mining in financial domain. *Knowledge based-System*, 114, 128-147.

- Leon, F. M. P., R. R. (2019). The Impact of Overconfidence and Optimism on Investment Decision on Individual Investor in Indonesia. *Jurnal Akuntansi*, 9(2), 93-102.
- Luo, B., Zeng, J., & Duan, J. (2015). Emotion Space Model for Classifying Opinions in Stock Message. *Expert Systems With Applications*, 12. doi:10.1016/j.eswa.2015.08.023
- Luo, Y., & Yang, X. (2014). Rumor Clarification and Stock Returns: Do Bull Market Behaves Differently from Bear Markets? *Emerging Markets Finance & Trade*, 50, 197-209. doi:10.2753/REE1540-496X500111
- Maisam A., R., Tabassamand, & Abbas, T. (2017). Effect of Optimism, Overconfidence, and Status Quo Bias on Stock Market Perceived Performance: Evidence from Islamabad Stock Exchange. *International Journal of Economics & Finance Research & Applications*, 1(2).
- Makhabel, M., Pradeepta, Danneman, & Heimann. (2017). *R: Mining Spatial, Text, Web, and Social Media Data* (T. Shriyan Ed.). Birmingham, UK: Packt Publishing Ltd.
- Margono H., Y., & Raikundalia, G. (2015). Clustering Indonesian Cyberbullying Words in Social Network. *International Journal of Computer Science and Electronics Engineering (IJCSEE)*, 3(4), 274.
- Morse, J., Barrett, Mayan, Olson, K., & Spiers, J., . (2002). Verification Strategies for Establishing Reliability and Validity in Qualitative Research. *International Journal of Qualitative Methods*, 1(2), 11.
- Nemati H.R, B. (2005). *Organizational Data Mining*. In. Retrieved from https://doi.org/10.1007/0-387-25465-X_50
- Nikfarjam, A. E., & Muthaiyah. (2010). Text mining approaches for stock market prediction. Paper presented at the 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE).
- Nuzula, R. M. F., N. . (2018). Does Investor Sentiment Affect Large-Cap and Small-Cap Stock. *Jurnal Administrasi Bisnis (JAB)*, 59, 51-61.
- Oppusunggu, L. S., I. P. . (2020). Coronavirus: How to Behave in Investment. *Journal of Advanced Research in Dynamical & Control Systems*, 12(06), 246-250.
- Prayogi, I. S., Y. (2018). Jargon Pedagang Saham di Telegram. *TOTOBUANG*, 6(2).
- Provost, F. F. T. (2013). *Data Science for Business* (M. L. a. M. Blanchette Ed. 1st ed.). Chicago: O'reilly.
- Raorane, K., & Jitkar. (2012). Association Rule – Extracting Knowledge Using Market Basket Analysis. *Research Journal of Recent Sciences*, 1, 9. doi:http://www.isca.in
- Raut, R. K. K., R. (2018). Investment Decision-Making Process between Different Groups of Investors : A Study of Indian Stock Market. *Asia-Pacific Journal of Management Research and Innovation*, 14(1-2), 11. doi:10.1177/2319510X18813770
- Sadeq, A. J. (2019). Factors influencing Individual Investor Behaviour: Evidence from the. *Asian Social Science*, 15(3).
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44 (4), 35. doi:10.1177/0539018405058216
- Shefrin, H. (2002). *Beyond Greed and Fear : Understanding Behavioural Finance and Psychology of Investing*: Oxford University Press.
- Shi, Y., Tang, Y., Cui, & Long, W. (2019). A Text Mining Based Study of Investor

- Sentiment and Its Influence on Stock. Economic Computation and Economic Cybernetics Studies and Research, 52(1), 183.
- Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. The Journal of Economic Perspectives, 17(1), 83-104.
- Slovic & Kerjan, E. (2010). The Irrational Economist: Making Decisions in a Dangerous World. In (pp. 336). New York Public Affairs.
- Song, C. S., C. (2007). SOPS: Stock Prediction using Web Sentiment. Paper presented at the Seventh IEEE International Conference on Data Mining - Workshops, Maryland, USA.
- Spence, M. (1974). Job Market Signaling. Quarterly Journal of Economics, 20.
- Statman, M. (2019). Behavioural Finance : The Second Generation. In (pp. 248). doi:
<http://dx.doi.org/10.2139/ssrn.3668963>
- Suwardjono. (2010). Teori Akuntansi: Perekayasa Pelaporan Keuangan. Yogyakarta: BPFE UGM.
- Tauni, M. Z., HongXing, & Iqbal, A. (2016). Personality Matter?, Qualitative Research in Financial Markets. Qualitative Research in Financial Markets, 8, 35. doi:<http://dx.doi.org/10.1108/QRFM-08-2015-0031>
- Turban, E., Volonino, L., & Wood, G. (2018). Information Technology for Management: On Demand Strategies for Performance, Growth, and Sustainability (B. L. Golub Ed. 10th ed.). Danvers, MA: Wiley.
- Wardana, R. (2020). Apa Itu IHSG? Ini Pengertian, Manfaat, dan Cara Hitungnya. Retrieved from <https://lifepal.co.id/media/ihsg/>
- Widyartati, P., Setyawati, I., & Ariyani, I. (2021). Kurs dan Nilai Perdagangan Saham di BEI Sebelum dan Pada Masa Pandemi COVID-19. Balance: Economic, Business, Management, and Accounting Journal, XVII(1), 10. doi:<http://dx.doi.org/10.30651/blc.v18i1.5446>
- Yang, S., Mo, S. Y. Kevin, & Liu, A. (2015). Twitter financial community sentiment and its Predictive Relationship. Quantitative Finance Journal, 15(10), 1637-1656.