

Analisis Dinamika Harga Saham yang Dipengaruhi oleh Analisis Sentimen di Media Sosial Menggunakan Algoritma *Support Vector Machine*

Antonio Galileo Tando, dan Mohammad Isa Irawan
Departemen Matematika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: mii.its.ac.id@matematika.its.ac.id

Abstrak—Saham dapat dideskripsikan sebagai tanda penyertaan modal pribadi atau pihak (badan usaha) dalam suatu perusahaan atau perseroan terbatas. Indeks LQ-45 terdiri atas 45 saham yang terpilih berdasarkan likuiditas perdagangan saham dan disesuaikan setiap enam bulan atau dua periode, maka saham yang terdapat dalam indeks tersebut akan selalu berubah. Analisis sentimen atau *opinion mining* merupakan studi komputasi dalam opini, sentimen, dan emosi yang diungkapkan dalam sebuah teks. Algoritma yang digunakan untuk melakukan klasifikasi adalah *Support Vector Machine* yang termasuk dalam algoritma *supervised learning* yang dapat digunakan untuk mengklasifikasikan teks secara otomatis. Pada penelitian ini, *pre-processing* teks yang digunakan adalah *case folding*, *tokenizing*, *normalization*, *stopwords*, dan *stemming*. Hasil klasifikasi untuk analisis sentimen dengan algoritma SVM menghasilkan *accuracy* rata-rata sebesar 75%. Kata-kata yang sering muncul pada masing-masing perusahaan, pada dataset sentimen positif adalah kata “bantu”, “kuat”, dan “sehat”. Sedangkan pada dataset negatif didominasi oleh kata “turun”, “tahan”, dan “bawah”. Hasil korelasi Rank Spearman menunjukkan beberapa perusahaan saham yaitu ANTM, BMRI, dan TLKM menghasilkan bahwa sentimen positif memiliki korelasi yang lemah dengan harga saham, sedangkan sentimen negatif tergolong tidak memiliki korelasi dengan harga saham.

Kata Kunci—Analisis Sentimen, Klasifikasi, Rank Spearman, Saham, *Support Vector Machine*.

I. PENDAHULUAN

SAHAM (*stock*) merupakan instrumen pasar keuangan yang paling terkenal hingga saat ini. Menurut Pradhya dkk., 2018, pergerakan dari harga saham dipengaruhi oleh informasi publik, kondisi sosial, kebijakan makroekonomi, kondisi politik dari suatu negara, dan lainnya [1]. Indeks saham merupakan acuan untuk investor saham di pasar modal untuk meninjau pergerakan harga saham serta menentukan saham dari emiten berdasarkan klasifikasi tertentu. Indeks saham merupakan tolok ukur kinerja pasar modal dari suatu negara. Saham yang terdaftar dalam LQ-45 selalu dipantau dan setiap enam bulan akan dilakukan evaluasi penilaian (awal Februari dan Agustus). Jika terdapat saham yang sudah tidak termasuk dalam kriteria yang sesuai, maka akan diganti dengan saham lain yang memenuhi syarat kriteria.

PT Kustodian Sentral Efek Indonesia (KSEI) telah mencatatkan jumlah *Single Investor Identification* (SID) mencapai 10,47 juta per 19 Oktober 2021. Jumlah tersebut terdiri dari 6,65 juta SID investor pasar modal dan 4,01 SID peserta tabungan perumahan rakyat (Tapera). Jumlah investor pasar modal Indonesia bertambah 71,42% dari tahun sebelumnya yang berjumlah 3,88 juta berdasarkan data akhir tahun 2020.

Di era modern seperti sekarang ini, penggunaan media sosial erat kaitannya dengan pekerjaan yang dilakukan oleh manusia. Media sosial adalah wujud dari perkembangan teknologi. Pesatnya perkembangan teknologi menyebabkan banyak manusia memiliki media sosial yang berbagai ragam jenisnya. Dengan berkembangnya media sosial, investor memiliki banyak peluang untuk memperoleh informasi lengkap seperti berita, data keuangan, dan informasi oleh perusahaan ke bursa efek. Platform media sosial dapat diakses dengan mudah oleh investor untuk berbagi informasi, pengalaman, serta tips untuk berinvestasi. Selanjutnya, dapat berkomunikasi dengan *investor relations* dari perusahaan terbuka yang telah bergabung dalam platform media sosial tersebut untuk mendapatkan informasi langsung terkait saham perusahaan terkait.

Analisis sentimen atau *opinion mining* merupakan studi komputasi dalam opini, sentimen, dan emosi yang diungkapkan dalam sebuah teks [2]. Analisis sentimen pada opini dilakukan untuk mengetahui sentimen yang bersifat positif dan sentimen yang bersifat negatif. Sentimen sendiri adalah suatu kata yang mengacu pada pendapat (opini) atau ekspresi yang didasarkan pada pandangan yang berlebihan terhadap suatu hal.

Support Vector Machine (SVM) merupakan salah satu metode *machine learning* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah class pada *input space*. Algoritma SVM termasuk dalam Algoritma *supervised learning* yang dapat digunakan untuk mengklasifikasikan teks secara otomatis. Penggunaan Algoritma SVM yang bertujuan untuk klasifikasi teks dengan menggunakan bobot index term sebagai fitur, dirintis oleh Thorsten Joachim. Pembelajaran SVM telah dipopulerkan sejak tahun 1992 oleh Boser, Guyon, dan Vapnik [3]. Koefisien korelasi Rank Spearman digunakan untuk mengidentifikasi dan menguji kekuatan suatu hubungan antara dua set data. Metode ini sering digunakan sebagai metode statistik untuk membantu pembuktian atau menyangkal hipotesis.

Dalam penelitian ini, LQ-45 sebagai salah satu indeks saham yang memiliki nilai likuiditas tinggi dan nilai cenderung tetap menjadikan objek penelitian ini menarik untuk diteliti. Penelitian ini mengambil sampel atas 45 perusahaan yang ada pada LQ-45 tahun 2021. Pada penelitian ini dilakukan analisis untuk melihat korelasi antara sentimen dari media sosial dengan harga saham dari perusahaan-perusahaan yang terdaftar dalam indeks LQ-45. Algoritma SVM menjadi pilihan metode yang akan digunakan penulis untuk klasifikasi analisis sentimen. Hasil klasifikasi sentimen

ini kemudian di analisis menggunakan metode uji korelasi *Rank Spearman*. Hal ini menjadi topik penulis untuk mengetahui pengaruh sentimen masyarakat melalui media sosial terhadap dinamika kerja saham dari LQ-45.

II. TINJAUAN PUSTAKA

A. Analisis Sentimen

Analisis sentimen merupakan salah satu bidang ilmu dari *Natural Language Processing* (NLP), *text mining*, dan *artificial intelligence*. Selain itu, analisis sentimen merupakan bidang ilmu yang bertujuan untuk menganalisis opini, sikap, evaluasi, dan penilaian terhadap suatu kejadian, topik, organisasi, individu, dan kegiatan lainnya [4]. Analisis sentimen mengelompokkan sebuah teks yang ada dari suatu kalimat atau dokumen kemudian menentukan opini yang telah dikemukakan dalam kalimat atau dokumen tersebut bersifat positif atau negatif.

B. Text Mining

Text mining adalah suatu ilmu penggalian data untuk memproses teks sehingga menjadi informasi yang diperoleh dari suatu peramalan pola dan cenderung melalui pola statistik. Teks yang diolah dapat berupa teks terstruktur dan teks tidak terstruktur. *Text mining* mengacu pada *information retrieval*, *data mining*, *machine learning*, statistik, dan komputasi linguistik [5].

C. Pre-processing

Pre-processing teks merupakan tahapan awal dalam pengolahan teks yang digunakan untuk perubahan bentuk dokumen menjadi data yang terstruktur sesuai kebutuhannya agar dapat diolah lebih lanjut dalam proses *text mining*. Tahapan praproses teks dalam klasifikasi bertujuan untuk meningkatkan akurasi klasifikasi data. *Pre-processing* dalam *text mining* cukup rumit karena dalam Bahasa Indonesia terdapat berbagai macam aturan dalam penulisan kalimat maupun pembentukan kata berimbuhan. Tahapan dalam *pre-processing* adalah sebagai berikut.

1) Case Folding

Case folding adalah proses untuk mengubah semua karakter teks menjadi huruf kecil serta menghilangkan tanda baca dan angka. Cara kerja *case folding* adalah memproses huruf alphabet dari “a” hingga “z” saja sehingga karakter selain huruf tersebut akan dihapus [6].

2) Tokenizing

Tokenizing merupakan proses memecah kalimat menjadi kata-kata atau memutus urutan string menjadi potongan-potongan, seperti kata-kata berdasarkan tiap kata yang menyusunnya.

3) Normalization

Normalization merupakan proses normalisasi bahasa terhadap kata yang tidak atau kurang baku. Proses ini mengembalikan bentuk penulisan dari masing-masing kata sesuai dengan KBBI.

4) Stopwords

Stopwords merupakan daftar kata umum yang tidak memiliki arti penting dan tidak digunakan. Pada proses ini kata umum akan dihapus untuk mengurangi jumlah kata yang disimpan oleh sistem [7].

5) Stemming

Stemming adalah proses untuk mencari root (kata dasar) dari kata hasil *stopword removal* (filtering). Terdapat dua aturan dalam melakukan *stemming* yaitu dengan pendekatan kamus dan pendekatan aturan.

D. Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode yang digunakan untuk menghitung bobot setiap kata yang telah diekstrak. Penggunaan metode ini dilakukan untuk menghitung kata umum yang ada pada *information retrieval*. Model pembobotan *TF-IDF* merupakan metode yang mengintegrasikan model *term frequency* (*tf*) dan *inverse document frequency* (*idf*), dimana *term frequency* (*tf*) merupakan proses untuk menghitung jumlah kemunculan term dalam satu dokumen dan *inverse document frequency* (*idf*) digunakan untuk menghitung term yang muncul di berbagai dokumen (komentar) yang dianggap sebagai term umum, yang dinilai tidak penting. Berikut Persamaan 1 dan Persamaan 2 merupakan rumus untuk menemukan pembobot dengan *TF-IDF*.

$$w_{ij} = tf_{ij} \times idf \quad (1)$$

$$idf = \log \left(\frac{N}{df_j} \right) \quad (2)$$

Dimana w_{ij} adalah bobot dari kata i pada dokumen ke j , N adalah jumlah dokumen, tf adalah jumlah kemunculan kata i pada dokumen j , dan df adalah jumlah dokumen j yang mengandung kata i .

E. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode *machine learning* yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah class pada *input space*. Algoritma SVM termasuk dalam Algoritma *supervised learning* yang dapat digunakan untuk mengklasifikasikan teks secara otomatis. Penggunaan Algoritma SVM yang bertujuan untuk klasifikasi teks dengan menggunakan bobot index term sebagai fitur, dirintis oleh Thorsten Joachim. SVM merupakan metode yang dapat menyelesaikan permasalahan secara linier maupun permasalahan non-linier. Dalam menyelesaikan permasalahan non-linier digunakan konsep kernel pada ruang kerja berdimensi tinggi, dimana akan mencari *hyperplane* yang dapat memaksimalkan margin antar kelas data. *Hyperplane* berguna dalam memisahkan 2 kelompok *class +1* dan *class -1* dimana setiap class memiliki *pattern* masing-masing.

F. Rank Spearman

Koefisien korelasi Rank Spearman digunakan untuk mengidentifikasi dan menguji kekuatan suatu hubungan antara dua set data. Metode ini sering digunakan sebagai metode statistik untuk membantu pembuktian atau menyangkal hipotesis.

Korelasi Rank Spearman dapat dicari dengan menggunakan rumus Persamaan 3 sebagai berikut.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

Tabel 1.
Interval korelasi rank spearman

Nilai Interval	Interpretasi
0.90–1.00	Korelasi Sangat Kuat
0.70–0.89	Korelasi Kuat
0.40–0.69	Korelasi Sedang
0.10–0.39	Korelasi Lemah
0.00–0.10	Korelasi Sangat Lemah

Dimana ρ adalah nilai korelasi Rank Spearman, d adalah beda urutan dari setiap pasangan, dan n adalah jumlah pasangan.

Langkah dalam melakukan uji hipotesis dalam korelasi adalah sebagai berikut [8]. Pertama yaitu menentukan uji hipotesis dengan menetapkan dua variabel yang akan dihubungkan. $H_0: \rho = 0$ (Tidak ada korelasi yang signifikan antara kedua variabel) dan $H_1: \rho \neq 0$ (Ada korelasi yang signifikan antara kedua variabel).

Kedua, nilai dari nilai signifikansi yang digunakan dalam uji korelasi yaitu 5% atau 0.05.

Ketiga, dasar pengambilan keputusan dari nilai signifikansi adalah sebagai berikut. Jika nilai Signifikansi < 0.05, maka H_0 ditolak dan H_1 diterima yaitu kedua variabel berkorelasi. Jika nilai Signifikansi > 0.05, maka H_0 diterima dan H_1 ditolak yaitu kedua variabel tidak berkorelasi.

Keempat, hubungan antar dua variabel dilihat dari arahnya dapat dibagi menjadi dua, yaitu hubungan yang sifatnya searah dan berlawanan arah. Hubungan searah disebut korelasi positif, artinya jika satu variabel naik, maka variabel lainnya akan ikut naik. Sementara yang berlawanan arah disebut korelasi negatif, artinya jika satu variabel naik, maka variabel lainnya akan turun.

Hasil dari uji korelasi ini digunakan untuk mencari korelasi antara jumlah sentimen dan harga saham. Hasilnya ditentukan dari seberapa besar nilai koefisien korelasi Rank Spearman. Berikut merupakan tabel interval korelasi Rank Spearman pada Tabel 1 [9].

G. Evaluasi

Evaluasi merupakan tahapan dalam upaya untuk mengukur keberhasilan suatu sistem dengan membandingkan hasil perolehan implementasi dengan kriteria standar yang telah ditetapkan [10]. Pengukuran evaluasi dilakukan berdasarkan *confusion matrix* pada Tabel 2.

Pengukuran yang sering digunakan untuk menghitung ketepatan klasifikasi adalah *Accuracy*, *Precision* dan *Recall* [11]. Perhitungan yang dilakukan dalam tahap evaluasi berupa *Accuracy*, *Precision*, *Recall*, dan *F-Measure* didefinisikan pada Persamaan 4 - 7 berikut:

$$Accuracy = \frac{TN+TP}{TN+TP+FP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

Tabel 2.
Confusion matrix

Classification	Predicted Positives	Predicted Negatives
Actual Positives Cases	Number of True Positives Cases (TP)	Number of False Negatives Cases (FN)
Actual Negatives Cases	Number of False Positives Cases (FP)	Number of True Negatives Cases (TN)

III. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah kumpulan data kualitatif dari media sosial yaitu Twitter mengenai sentimen terkait LQ-45 pada bulan Februari 2021 hingga Januari 2022. Data harga saham LQ-45 didapat dari *website Yahoo Finance* pada bulan Februari 2021 hingga Januari 2022 mengambil data harga saat *closed*.

Metode dan tahapan penelitian yang dilakukan adalah sebagai berikut.

A. Studi Literatur

Studi literatur dilakukan untuk memperoleh dan memahami teori-teori yang berhubungan dengan penelitian.

B. Pengambilan Data

Pengambilan data dilakukan untuk memperoleh bahan penelitian sesuai dengan tujuan penelitian yang telah ditetapkan.

C. Melakukan Pre-Processing Data

Tahap pertama adalah menghapus data yang mengandung dua sentimen, positif dan negatif. Tahap kedua adalah menghilangkan *mention*, *URL*, *hashtag*, angka, tanda baca, dan koma. Tahap berikutnya adalah melakukan pre-processing data, antara lain yaitu melakukan *case folding*, yaitu mengubah semua teks dengan huruf kecil (non-kapital) serta menghilangkan tanda baca. Selanjutnya, melakukan *tokenizing* untuk memecah data menjadi kata per kata. Kemudian, melakukan *normalization* untuk mengoreksi kata yang kurang tepat. Dilanjutkan dengan menghapus kata yang terdapat dalam daftar *stopwords*. Terakhir, melakukan *stemming* untuk menghilangkan kata imbuhan dan mendapatkan kata dasar Klasifikasi data menggunakan SVM.

D. Ekstraksi Fitur Menggunakan TF-IDF

Dalam tahapan ini, teks akan diubah menjadi vektor dan akan dilakukan pembobotan kata dengan menggunakan TF-IDF untuk menghitung bobot setiap kata yang telah diekstrak. Kemudian akan ditentukan pembobot parameter pada SVM dan membangun model SVM dengan fungsi kernel linier.

E. Evaluasi Hasil Klasifikasi dengan Algoritma SVM

Evaluasi model SVM dilakukan untuk menunjukkan seberapa baik kemampuan algoritma untuk mengklasifikasikan data. Evaluasi klasifikasi akan dilakukan dengan menggunakan *confusion matrix*.

Dengan pengukuran *confusion matrix*, akan dihasilkan pengukuran antara lain, yang pertama, *accuracy*, yaitu pengukuran untuk menghitung seluruh keadaan yang diprediksikan dengan nilai yang benar terhadap seluruh keadaan yang diprediksi. *Accuracy* menunjukkan seberapa

Tabel 3.
Jumlah objek penelitian

No	Nama Saham	Kode Saham	Jumlah Data
1	Aneka Tambang Tbk.	ANTM	233
2	Bank Central Asia Tbk.	BBCA	194
3	Bank Rakyat Indonesia (Persero) Tbk.	BBRI	499
4	Bank Mandiri (Persero) Tbk.	BMRI	100
5	Indo Tambangraya Megah Tbk.	ITMG	129
6	Bukit Asam Tbk.	PTBA	100
7	PP (Persero) Tbk.	PTPP	100
8	Timah Tbk.	TINS	510
9	Telkom Indonesia (Persero) Tbk.	TLKM	155
10	Unilever Indonesia Tbk.	UNVR	114

akurat sistem dapat mengklasifikasikan data secara benar.

Kedua, *precision*, yaitu perhitungan pada kondisi benar, yaitu kelas aktual dan kelas prediksi yang sama (positif) terhadap seluruh kondisi yang diprediksi positif. *Precision* menunjukkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif.

Ketiga, *recall*, yaitu perhitungan pada kondisi benar yaitu, merupakan kelas data positif terhadap seluruh kondisi aktual yang bernilai positif. *Recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem.

Keempat, *F-Measure*, yaitu perhitungan yang melibatkan *precision* dan *recall* untuk dicari nilai tengah pada kedua evaluasi tersebut. *F-Measure* dibutuhkan untuk mempresentasikan *precision* dan *recall*.

F. Melakukan Visualisasi Data

Melakukan visualisasi data menggunakan *word cloud* untuk melihat kata-kata dari setiap perusahaan LQ-45.

G. Analisis Menggunakan Metode Uji Korelasi Rank Spearman.

Melakukan analisis dinamika harga saham pada LQ-45 menggunakan metode uji korelasi Rank Spearman yang dilakukan untuk melihat hubungan dari sentimen terhadap data harga saham dengan *software Statistical Package for the Social Science* (SPSS). Hasil dari tahapan ini akan ditentukan seberapa besar nilai koefisiensi korelasi Rank Spearman untuk mendapatkan hasil analisis.

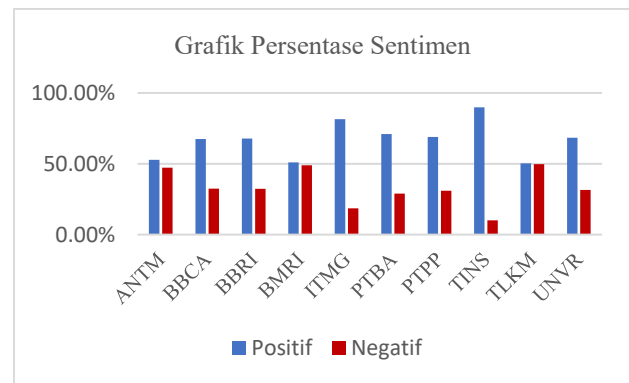
H. Penarikan Kesimpulan

Setelah melakukan interpretasi dan analisis data, maka ditarik suatu kesimpulan yang merupakan ringkasan untuk menjawab tujuan penelitian.

IV. ANALISIS DAN PEMBAHASAN

A. Perancangan dan Pre-processing Data

Tahap awal pada perancangan dimulai dengan mengumpulkan seluruh data berupa kalimat opini mengenai 45 perusahaan LQ-45 yang dibutuhkan dari media sosial untuk tahap pengolahan data selanjutnya. Pengambilan data akan dicari dengan menggunakan kode perusahaan LQ-45. Pengambilan data dilakukan melalui web *export comments* yang dirancang untuk mengumpulkan data dari media sosial untuk proses selanjutnya yang akan disimpan ke dalam format file CSV. Data yang sudah diinput dan dikumpulkan untuk dilakukan *pre-processing*. Data sentimen yang telah terkumpul dari 45 perusahaan LQ-45 dilakukan proses



Gambar 1. Grafik persentase klasifikasi sentimen perusahaan LQ-45.

cleaning data terlebih dahulu untuk membersihkan beberapa parameter yang tidak relevan dengan isi data. Peneliti mengambil batas 100 sentimen dalam setiap kode perusahaan yang ada sebab jumlah data yang terlalu sedikit akan memberikan hasil penelitian yang kurang sesuai. Setelah dilakukan proses *cleaning data*, didapatkan total 10 perusahaan yang memiliki jumlah data yang sesuai untuk diteliti. Berikut merupakan perusahaan saham yang menjadi objek penelitian dalam penelitian pada Tabel 3.

B. Pelabelan Data

Data teks yang sudah dibersihkan pada tahap *pre-processing* selanjutnya akan dilakukan proses pelabelan data. Dimana setiap opini akan diberi label berdasarkan sentimen positif atau negatif. Jumlah kategori sentimen positif dan negatif masing-masing perusahaan memiliki jumlah yang beragam. Berikut merupakan grafik persentase klasifikasi sentimen pada 10 perusahaan saham objek penelitian pada Gambar 1.

C. Klasifikasi SVM

Langkah pertama dalam mengklasifikasikan data adalah melatih model menggunakan data training. Data dari setiap media yang telah dilakukan *pre-processing* digunakan untuk model menggunakan *software* Python 3 dan Jupyter Notebook. Untuk melakukan klasifikasi data, diperlukan sebuah proses sebagai berikut.

1) Split Data

Pada proses ini semua data perusahaan yang telah dilakukan *pre-processing* akan dibagi menjadi 2 bagian yaitu sebagai *data training* dan *data testing* dengan rasio *data training* 80% dan *data testing* 20%. Proses *split data* bertujuan untuk membagi data secara otomatis dan acak berdasarkan ketentuan rasio yang ditentukan.

2) Pembentukan Model

Pada proses ini dibutuhkan *data training* dan *data testing*. *Data training* adalah data yang berfungsi untuk membentuk sebuah model klasifikasi. *Data training* merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru. Pada *data training* sentimen sudah diberikan pelabelan data yaitu kelas sentimen positif dan kelas sentimen negatif.

3) Apply Model

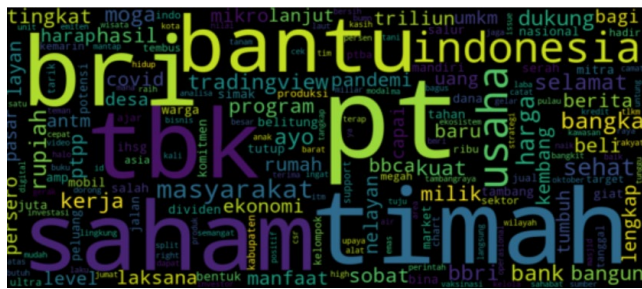
Proses *apply model* merupakan proses pengujian untuk mengetahui apakah model yang dibuat dengan *data training* sudah memiliki standar akurasi yang baik. Pada proses ini

Tabel 4.
Ketepatan klasifikasi terbaik dengan SVM linier

Kode Saham	Accuracy	Precision	Recall	F-Measure
ANTM	0.72340	0.875	0.6774	0.7636
BBCA	0.69230	0.23076	0.6	0.33333
BBRI	0.75757	0.48571	0.73913	0.58620
BMRI	0.65	0.69230	0.75	0.71999
ITMG	0.9090	0.5	0.5	0.5
PTBA	0.7	0.33333	0.5	0.4
PTPP	0.75	0.16666	1.0	0.28571
TINS	0.90849	0.17647	1.0	0.3
TLKM	0.70212	0.64285	0.81818	0.72000
UNVR	0.78260	0.93333	0.77777	0.84848

Tabel 5.
Perubahan sentimen harian ANTM

Tanggal	Tweet	Sentimen
1 Februari 2021	investor lokal gantung manuver prank bandar bandar gantung panik selling murah investor lokal hard landing arb tarik ulur smp mental kuat prank bandar game changer investor lokal tahan ikut pengaruh panik selling murah bris antm smp arb	Negatif
2 Februari 2021	Wuahahaha kok ngakak ya. Lhais keee #antm	Negatif
3 Februari 2021	#ANTM cakep nih buat yg mau hari ini n besok potensi ada kenaikan... Accum tipis Volume oke... Diskon. on: hanya sharing saja Balik ke analisa pribadi masing masing...	Positif
3 Februari 2021	pertarungan buyer vs seller masih sangat ketat di antm, dimana #antm juga sdh out dari range trading sebelumnya yg hrsnya bisa menahan penurunan lbh lanjut baginya	Negatif
3 Februari 2021	Kayaknya besok manggung #ANTM	Positif
4 Februari 2021	Sayang sekali Antam #ANTM tdk mau memonetisasi #eksplorasi-nya. Padahal bisa banget! Entah kenapa. Laporan eksplorasi isinya biaya eksplorasi	Negatif
5 Februari 2021	#kafein #ANTM-2600 #antm tampaknya mulai positif kembali setelah hari ini bisa bertahan di atas support MA5-2440 dan berhasil berada di atas support naik 2550/2530, dgn potensi awal uji resisten MA20-2750, sambil menunggu konfirmasi lanjutan dari 3MA nya.	Positif
5 Februari 2021	ANTM penyelamat portofolio gasss terus sentimen #Tesla masih panjang,,, semoga #ANTM #TINS meroket	Positif
5 Februari 2021	Juara hari ini! #antm	Positif



Gambar 2. Word cloud sentimen positif.



Gambar 3. Word cloud sentimen negatif.

dibutuhkan *data testing*. *Data testing* adalah data yang belum mendapatkan pelabelan data guna untuk menguji tingkat akurasi model.

4) *Evaluasi Model*

Performansi model klasifikasi akan mengukur seberapa baik kemampuan sistem tersebut dalam mengklasifikasikan data. Pengukuran ketepatan klasifikasi dilakukan dengan membentuk *confusion matrix*. Pada dasarnya *confusion matrix* membandingkan hasil klasifikasi yang dilakukan sistem dengan hasil klasifikasi seharusnya. Kemudian dari *confusion matrix* akan menghasilkan empat pengukuran yaitu *accuracy*, *precision*, *recall*, dan *f-measure*.

Hasil perhitungan ketepatan klasifikasi setiap perusahaan selanjutnya dirangkum menjadi satu tabel. Berikut merupakan hasil pengukuran ketepatan klasifikasi terbaik dari setiap perusahaan menggunakan algoritma SVM menggunakan kernel linier pada Tabel 4.

D. *Visualisasi Word Cloud*

Visualisasi data teks menggunakan *word cloud* digunakan untuk mengetahui kata-kata yang paling sering muncul pada data. Ukuran *font* pada *word cloud* menunjukkan frekuensi kemunculan kata. Semakin besar ukuran font berarti semakin besar frekuensi kemunculan kata tersebut. Berikut merupakan contoh *word cloud* sentiment positif dan negatif pada perusahaan LQ-45 pada Gambar 2 dan Gambar 3.

Berdasarkan kata-kata yang sering muncul pada masing-masing perusahaan, dataset sentimen positif adalah kata

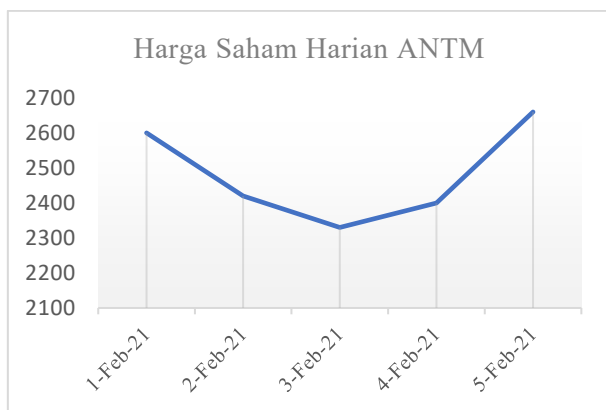
“bantu”, “kuat”, dan “sehat”. Kata dalam dataset tersebut menunjukkan bahwa bantuan yang diberikan pemerintah dan masyarakat memberikan pengaruh yang kuat dan sehat untuk Indonesia. Sedangkan pada dataset negatif didominasi oleh kata “turun”, “tahan”, dan “bawah”. Kata dalam dataset negatif menunjukkan bahwa harga saham yang sedang turun, tahan, dan sedang dibawah mempengaruhi masyarakat untuk memberikan sentimen negatif di media sosial.

E. *Dinamika Harga Saham*

Menurut Kartono, 2007 pengertian dinamika adalah suatu bentuk perubahan secara dinamis, baik itu yang sifat besar-besaran atau kecil-kecilan, maupun secara cepat atau lambat yang sifatnya nyata dan berhubungan dengan suatu kondisi keadaan [12]. Harga saham yang memiliki dinamika naik turun harga saham menggambarkan kondisi pasar saham yang bersifat fluktuatif. Untuk menunjukkan suatu dinamika harga saham, berikut disajikan grafik dinamika harga saham harian dari tanggal 1 Februari – 5 Februari 2021 perusahaan ANTM pada Gambar 4.

Pada perusahaan ANTM, perubahan dinamika nilai harga saham dapat dilihat dari grafik diatas. Dinamika harga saham yang terus berubah juga memengaruhi tweet masyarakat mengenai perusahaan saham pada hari yang sama. Berikut merupakan perubahan sentimen harian pada ANTM pada Tabel 5.

Berdasarkan Tabel 5 dapat dilihat bahwa terjadi perubahan sentimen pada tanggal 3 sampai 4 Februari. Sentimen positif



Gambar 4. Grafik Dinamika Harga Saham Harian ANTM

yang diberikan masyarakat memberikan pengaruh pada harga saham di hari berikutnya. Dapat dilihat pada tanggal 3 Februari 2021 bahwa para investor memberikan sentimen positif dan negatif, namun pada tanggal 4 Februari harga saham pada ANTM naik dari Rp.2.330 menjadi Rp.2.400. Ini menunjukkan bahwa sentimen positif memiliki pengaruh terhadap harga saham. Kemudian pada tanggal 4 Februari terdapat sentimen negatif yang diberikan oleh masyarakat, namun harga saham pada tanggal 5 Februari signifikan naik dari yang sebelumnya Rp.2.400 menjadi 2.660. Hal tersebut menunjukkan bahwa sentimen negatif belum tentu memberikan pengaruh terhadap harga saham perusahaan. Harga saham yang naik dari hari sebelumnya membuat para masyarakat atau investor memberikan sentimen yang positif di hari yang sama. Dapat dilihat dari grafik harian dan grafik mingguan perusahaan ANTM bahwa harga saham yang selalu berubah naik turun cenderung membentuk dinamika pada harga saham dan memberikan pengaruh pada sentimen yang diberikan oleh masyarakat. Hal ini juga terjadi pada perusahaan saham LQ-45 lainnya.

F. Uji Korelasi Rank Spearman

Teknik korelasi digunakan untuk mencari hubungan dan membuktikan hipotesis hubungan dua variabel bila data kedua variabel berbentuk interval atau rasio. Untuk menguji hipotesis, maka diperlukan hipotesis penelitian sebagai berikut: H_0 : Jumlah sentimen tidak berkorelasi dengan harga saham, H_1 : Jumlah sentimen berkorelasi dengan harga saham. Pengujian dilakukan dengan menggunakan nilai signifikansi dengan ketentuan sebagai berikut: H_0 diterima jika nilai Signifikansi > 0.05 , maka H_1 ditolak, H_0 ditolak jika nilai Signifikansi < 0.05 , maka H_1 diterima.

Hasil korelasi dengan sentimen positif dari setiap perusahaan menggunakan uji korelasi Rank Spearman pada Tabel 6. Sedangkan hasil korelasi dengan sentimen negatif dari setiap perusahaan menggunakan uji korelasi Rank Spearman pada Tabel 7.

Berdasarkan hasil penelitian ini dapat dikatakan bahwa sentimen-sentimen di media sosial Twitter tidak memiliki cukup pengaruh terhadap harga saham. Hasil penelitian ini tidak sejalan dengan teori dan hasil penelitian-penelitian terdahulu yang menyatakan bahwa sentimen mempengaruhi harga saham yang berarti terdapat hubungan yang cukup kuat antara keduanya. Hal ini mungkin terjadi akibat beberapa faktor seperti kehadiran *influencer marketing* dalam mempromosikan saham tertentu, kondisi fundamental ekonomi makro, misalnya tingkat inflasi dan suku bunga,

Tabel 6.
Hasil korelasi sentimen positif

Kode Perusahaan	Korelasi	Signifikansi
ANTM	0.298	0.027
BBCA	0.055	0.690
BBRI	0.233	0.087
BMRI	0.344	0.010
ITMG	0.238	0.081
PTBA	0.229	0.093
PTPP	0.000	0.999
TINS	0.076	0.583
TLKM	0.340	0.011
UNVR	0.227	0.96

Tabel 7.
Hasil korelasi sentimen negatif

Kode Perusahaan	Korelasi	Signifikansi
ANTM	-0.126	0.357
BBCA	-0.130	0.343
BBRI	-0.158	0.248
BMRI	-0.027	0.845
ITMG	0.177	0.197
PTBA	-0.034	0.806
PTPP	-0.123	0.369
TINS	0.223	0.102
TLKM	-0.198	0.148
UNVR	0.140	0.310

masyarakat tidak mudah dipengaruhi oleh isu-isu yang beredar, fluktuasi kurs Rupiah terhadap mata uang asing, masyarakat memiliki kepercayaan yang besar terhadap perusahaan-perusahaan LQ-45, dan kebijakan pemerintah yang menimbulkan volatilitas harga saham, seperti kebijakan perseroan, kebijakan ekspor impor, dan kebijakan Penanaman Modal Asing (PMA).

V. KESIMPULAN

A. Kesimpulan

Kesimpulan dari penelitian ini yang pertama yaitu, cara menerapkan analisis sentimen di media sosial yaitu dengan melakukan *export data*, *pre-processing data* yaitu *clean text*, *case folding*, *tokenizing*, *normalization*, *stopwords*, dan *stemming* selanjutnya pelabelan data, ekstraksi fitur, dan pembentukan model klasifikasi menggunakan algoritma SVM dengan kernel linier parameter C . Hasil klasifikasi untuk analisis sentimen dengan algoritma SVM menghasilkan *accuracy* rata-rata sebesar 75%. Kedua, kata-kata yang sering muncul pada masing-masing perusahaan, pada dataset sentimen positif adalah kata “bantu”, “kuat”, dan “sehat”. Kata dalam dataset tersebut menunjukkan bahwa bantuan yang diberikan pemerintah dan masyarakat memberikan pengaruh yang kuat dan sehat untuk Indonesia. Sedangkan pada dataset negatif didominasi oleh kata “turun”, “tahan”, dan “bawah”. Kata dalam dataset negatif menunjukkan bahwa harga saham yang sedang turun, tahan, dan sedang dibawah mempengaruhi masyarakat untuk memberikan sentimen negatif di media sosial. Ketiga, hasil korelasi Rank Spearman menunjukkan beberapa perusahaan saham yaitu ANTM, BMRI, dan TLKM menghasilkan bahwa sentimen positif memiliki korelasi yang lemah dengan harga saham, sedangkan sentimen negatif tergolong tidak memiliki

korelasi dengan harga saham.

B. Saran

Adapun saran untuk penelitian selanjutnya yang pertama yaitu, menambahkan sumber data media sosial untuk sentimen yang akan diambil seperti Facebook, Instagram, dan media sosial lainnya agar dapat memperkuat tingkat penelitian. Kedua, menambahkan hipotesis dengan variabel penelitian seperti variabel tingkat inflasi dan variabel kurs Rupiah terhadap Dollar Amerika. Ketiga, dapat menggunakan algoritma yang lebih baik dan cocok untuk gaya bahasa investor saham.

DAFTAR PUSTAKA

- [1] I. C. Pradhya, D. Iskandar, and R. C. Tarumingkeng, "Analisis faktor faktor yang mempengaruhi indeks harga saham gabungan di bursa efek Indonesia," *J. Manaj. Bisnis Kompetensi*, vol. 133, no. 1, pp. 43–56, 2018.
- [2] B. Liu, *Sentiment Analysis and Subjectivity, Handbook of Natural Language Processing*, 2nd ed. Chicago: Department of Computer Science, University of Illinois at Chicago, 2010.
- [3] Paramita, "Penerapan Support Vector Machine Untuk Ekstraksi Informasi Dari Dokumen Teks," Departemen Teknik Informatika, Institut Teknologi Bandung, 2017.
- [4] B. Liu, *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & Claypool, 2012.
- [5] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. Massachusetts: Morgan Kaufmann, 2012.
- [6] S. M. Weiss, *Text Mining: Predictive Methods for Analyzing Unstructured Information*. New York: Springer, 2010.
- [7] H. Manning, C., Raghavan, P. & Schütze, *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2009.
- [8] A. Anwar, *Statistika untuk Penelitian Pendidikan*. Kediri: IAIT Press, 2009.
- [9] S. LA Schober P, Boer C, "Correlation coefficients: appropriate use and interpretation," *Anesth Analg*, vol. 126, no. 5, pp. 1763–1768, 2018, doi: 10.1213/ANE.0000000000002864.
- [10] R. Parikh and M. Movassate, *Sentiment Analysis of User-Generated Twitter Updates using Various Classification Techniques*. California: Stanford University, 2009.
- [11] A. Hotho, A. Nürnberger, and G. Paass, "A brief survey of text mining," *LDV Forum*, vol. 20, pp. 19–62, 2005, doi: 10.21248/jlcl.20.2005.68.
- [12] Kartono, *Pengantar Desain Komunikasi Visual*. Yogyakarta: Andi Offset, 2007.