
Pemanfaatan Analisis Sentimen dalam Prediksi Harga Saham: Studi pada Investor Retail Indonesia

*Christophorus Bintang Saputra¹
Deddy Priatmodjo Koesrindartoto²

¹Mahasiswa Pascasarjana School of Business and Management,
Institut Teknologi Bandung, Indonesia

²Dosen School of Business and Management, Institut Teknologi Bandung, Indonesia
christophorus.bintangsaputra@gmail.com¹
deddypri@sbm.itb.ac.id²

ABSTRACT

The upswing in engagement from retail investors in the Indonesian stock market aligns with a significant rise in the use of various social media platforms as conduits for stock-related information. In particular, number of content creators shared information about stock in Youtube grows, the information including the effect of corporation actions at stock market. This study sought to leverage sentiment data extracted from particular videos to predict the stock closing prices, especially at the corporate action event using Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) models. In this study also included several classification algorithm result to explore the accuracy in the prediction models. The result indicate that while sentiment from Youtube serves a viable variable for prediction, the Bi-LSTM model shows better performance compared to the based model in forecasting stock prices surrounding corporate action dates. Furthermore, the combination with classification algorithms shows an improvement in refining predictions, where demonstrate a potential accuracy score when incorporated into the predictive model.

This research contributes insights the potential value using sentiment from Youtube platform and machine learning models to predict the time series data, especially in stock market. The findings hold significance for Indonesian retail investors seeking an alternative decision-making tools within the dynamic stock market landscape.

Keywords: LSTM; Bi-LSTM; Machine Learning; Stock; Sentiment.

Article history: Submission date January 31st 2023 Accepted date August 8th 2023

To cite: Saputra, C.B, Koesrindartoto, D.P. (2024). Pemanfaatan Analisis Sentimen dalam Prediksi Harga Saham: Studi pada Investor Retail Indonesia. *Jurnal Manajemen*, 21(1), 1-17.

¹*Corresponding Author: Christophorus Bintang Saputra. Email: christophorus.bintangsaputra@gmail.com

1.PENDAHULUAN

Latar Belakang

Penggunaan informasi media sosial saat ini memiliki manfaat yang beragam, seperti halnya penggunaan untuk mendapatkan pengetahuan atau informasi dalam saham. Investor ritel saat ini dapat menggunakan informasi di berbagai media sosial bertujuan untuk mencapai keuntungan atau hasil yang positif dalam saham yang mereka investasikan. Penelitian sebelumnya menemukan hasil empiris yang signifikan, yaitu abnormal return dan abnormal trading volume yang positif pada hari setelah rekomendasi diberikan di platform media sosial populer di Cina (Zhang, 2018). Analisis sentimen melalui berbagai platform online juga menjadi populer dalam topik pasar keuangan (Daniel, 2017; Gan, 2020; Jing, 2021).

Selama pandemi covid-19 tahun 2020 terjadi peningkatan investor retail domestik, peningkatan tersebut tercatat menjadi jumlah investor tertinggi sepanjang sejarah pasar saham di Indonesia. Peningkatan jumlah investor juga diimbangi oleh meningkatnya informasi terkait pasar saham di berbagai media sosial. Youtube menjadi salah satu media yang digunakan untuk mendapatkan informasi terkait pasar saham. Dalam format video, investor dapat mempelajari informasi yang diberikan oleh berbagai narasumber, seperti praktisi dan akademisi yang membuat konten video terkait pasar saham. Topik yang dibahas dalam Youtube sangat variatif, seperti ilmu dasar teknikal analisis, peristiwa ekonomi, atau perusahaan yang dapat berdampak pada harga saham. Informasi yang didapat investor digunakan sebagai pertimbangan untuk membeli atau menjual saham dalam portofolio investor. Pertimbangan investor dalam mengelola informasi yang didapatkan dari Youtube dapat menjadi setuju maupun tidak setuju. Dalam *platform* tersebut investor dapat memberikan respon atau reaksi melalui fitur komentar dan memberikan opini terkait informasi yang diberikan oleh narasumber.

Tujuan dari penelitian ini untuk investigasi sentimen investor melalui Youtube yang digunakan sebagai variabel untuk memprediksi harga penutupan pada saham dengan topik yang lebih spesifik, yaitu aksi korporasi. Penggunaan sentimen dapat membantu untuk menilai opini positif, negatif ataupun netral terhadap suatu informasi (Amrani,2018). Analisis sentimen akan digunakan dalam model *machine learning neural network* dan dikombinasikan dengan indikator teknikal saham. Dengan mengembangkan pendekatan *machine learning*, khususnya dalam pasar saham, untuk mengatasi keterbatasan pendekatan tradisional yang bersifat sepenuhnya dalam bidang keuangan dan ekonomi. Prediksi merupakan masalah kompleks dalam instrumen keuangan karena melibatkan berbagai variabel. Dalam memprediksi pasar saham, bisa terjadi banyak ketidakpastian karena selalu dipengaruhi oleh sentimen masyarakat (Shah, 2022). Metode *machine learning* yang diterapkan dalam prediksi pasar saham dalam penelitian terkini adalah *Long-Short Term Memory* (LSTM). LSTM merupakan metode

pembelajaran mendalam yang dirancang untuk mengatasi masalah eksplorasi dan penghilangan gradien pada data berurutan (Hochreiter, 1997).

Model yang digunakan dalam penelitian ini membandingkan performa LSTM dan *Bi-directional Long-Short Term Memory* (Bi-LSTM). Sebagai variabel independen, dalam proses sentimen analisis diklasifikasikan dalam tiga kategori: positif, netral, dan negatif. Sentimen akan diberi tag pada setiap kalimat dengan parameter tertentu untuk menentukan apakah kalimat tersebut memiliki sentimen positif, negatif, atau netral terhadap tindakan korporat yang sedang dibahas dalam video. Dalam penelitian ini, dibandingkan sentimen yang diolah dalam pembelajaran mesin untuk mengklasifikasikan sentimen setelah diberi label sebagai variabel input dalam prediksi.

Dengan menggabungkan pandangan para investor terhadap saham, (Jing, 2021) mengembangkan prediksi yang lebih akurat untuk harga saham di masa depan menggunakan jaringan saraf LSTM. LSTM telah digunakan dalam berbagai penelitian deret waktu, terutama dalam prediksi harga pasar saham. (Moghar, 2020) menemukan hasil yang menjanjikan, di mana model LSTM mampu melacak perubahan harga. Dengan analisis sentimen, dengan mengklasifikasikan data teks yang mencoba menjelaskan opini atau perasaan dari suatu konten dapat membantu memprediksi pergerakan harga saham.

Sentimen yang digunakan dalam penelitian ini adalah opini investor terkait dengan aksi korporasi dalam komentar pada setiap video narasumber di Youtube. Opini investor dapat menggambarkan investor setuju atau tidak setuju terkait informasi aksi korporasi sebagai bahan pertimbangan investor untuk membeli atau menjual pada saat *due date*. Dalam penelitian ini terdapat tiga tahapan: *preprocessing data*, membangun model klasifikasi sentimen, dan model prediksi harga penutupan saham. Hasil model klasifikasi sentimen dikombinasikan dengan indikator teknikal sebagai variabel dalam prediksi harga penutupan saham menggunakan dua model *neural network* LSTM dan Bi-LSTM.

Penelitian ini bertujuan untuk menggunakan analisis sentimen dari Youtube sebagai informasi tambahan dalam pengambilan keputusan dengan memprediksi pergerakan harga saham pada masa aksi korporasi. Prediksi pergerakan harga saham dengan mengintegrasikan klasifikasi polar sentimen, indikator teknikal sebagai variabel menggunakan model LSTM, dan Bi-LSTM untuk prediksi pergerakan harga saham. Metode klasifikasi yang digunakan dalam penelitian ini adalah *RandomForest*, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), dan *Logistic Regression* yang akan dibandingkan dengan data sentimen tanpa melalui proses algoritma klasifikasi.

2. TINJAUAN LITERATUR

Perilaku keuangan adalah bidang yang mencakup berbagai disiplin ilmu yang berakar pada asumsi dan konsep dari perilaku ekonomi, mencakup emosi, sifat, preferensi, serta aspek intelektual dan sosial dari sifat manusia. Dalam perilaku keuangan psikologi kognitif dapat merujuk pada bagaimana cara berpikir manusia (Ritter, 2003). Dalam perilaku tersebut keputusan investor tidak sepenuhnya rasional, yang dipengaruhi oleh preferensi dan kepercayaan. Dalam teori perilaku keuangan terdapat dasar teori *Overconfidence* yang menjelaskan terkait pentingnya *self-awareness* dan analisis yang holistik dalam pengambilan keputusan keuangan. Investor yang menyadari keberadaan bias *overconfidence* dapat mengambil langkah - langkah untuk mengurangi dampaknya dengan mencari pendapat yang beragam, melakukan analisis yang mendalam, dan memperhatikan batasan dan bias (Ricciardi, 2000).

Individu yang *overconfident* berpotensi melakukan perdagangan pada pasar saham secara berlebihan, meyakini bahwa memiliki informasi dan kemampuan yang lebih unggul, sehingga meningkatkan resiko biaya transaksi yang lebih tinggi dan potensi keuntungan yang lebih rendah. Perilaku investor yang *overconfidence* frekuensi dalam melakukan perdagangan dalam pasar saham akan lebih tinggi, yang akan membuat biaya transaksi lebih tinggi dari pada keuntungan (Odean, 1999). Dengan perilaku tersebut, investor kurang mempertimbangkan risiko pada portofolio investasi sehingga menyebabkan peningkatan volatilitas pada keuntungan portofolio. Pada perilaku keuangan juga terdapat kaitan dengan teori *Financial Cognitive Dissonance*, investor memiliki kecenderungan untuk mencari konsistensi dalam keyakinan dan tindakan, ketika dihadapkan pada informasi atau keputusan keuangan yang kontradiktif, individu dapat mengalami ketidakselarasan kognitif yang membuat keputusan keuangan yang bertentangan dengan keyakinan, nilai atau sikap, pada kondisi ini investor akan menyelaraskan keyakinan dengan bukti baru yang bertentangan. Dalam penelitian (Barberis, 1998), pola perdagangan dalam melihat tingkat pengembalian saham dan volume perdagangan konsisten dengan perilaku yang berdasarkan sentimen, sehingga dengan memahami sentimen investor akan membantu untuk melihat pergerakan fluktuasi pasar dan mengidentifikasi kesempatan mendapatkan keuntungan.

Hipotesis Pasar Efisien

Pada pasar keuangan yang efisien mencerminkan semua informasi yang tersedia pada harga aset terkait dengan waktu sehingga bagi investor akan lebih selektif dan berhati - hati dalam memilih saham untuk mendapatkan keuntungan yang lebih tinggi dari rata - rata secara konsisten (Fama, 1970). Teori ini menyatakan bahwa mendapatkan keuntungan secara abnormal adalah irasional karena harga saham tidak menggambarkan aspek psikologis investor

dan hanya mencerminkan nilai fundamental pada harga saham. Terdapat tiga bentuk dalam pasar efisien.

1. *Weak Form of Efficiency*

Harga historis dan volume perdagangan menjadi cermin dalam harga pasar saham. Namun, investor tidak dapat secara konsisten mencapai tingkat pengembalian yang superior dengan menganalisis data harga dan volume historis karena semua informasi yang terkandung dalam harga masa lalu yang sudah menjadi harga pasar saat ini.

2. *Semi-strong Form of Efficiency*

Bentuk ini memperluas bukan hanya pada informasi teknikal historis, informasi publik, laporan keuangan, dan indikator ekonomi menjadi pendukung dalam pengambilan keputusan investor. Dalam kondisi ini investor cenderung akan beradaptasi dengan cepat terhadap informasi baru sehingga tingkat pengembalian abnormal akan menjadi inkonsisten.

3. *Strong Form Efficiency*

Pada bentuk efisiensi yang kuat, semua informasi yang bersifat publik dan internal menjadi refleksi harga saham. Oleh karena itu, untuk mencapai keuntungan abnormal pada pasar saham tidak dapat dilakukan karena tidak ada informasi yang memberikan investor keuntungan yang kompetitif.

Pada teori Hipotesis Pasar Efisien berpendapat bahwa kemampuan prediksi menggunakan sentimen investor tidak dapat digunakan, namun indikator sentimen bisa memberikan sinyal dalam *trend* pasar dalam jangka waktu pendek (Tetlock, 2007). (Shleifer, 2000) berpendapat sentimen investor refleksi keputusan emosional daripada analisis rasional sehingga melemahkan asumsi teori Hipotesis Pasar Efisien. Dalam bentuk *Semi-Strong Form of Efficiency* menyatakan bahwa harga saham mencerminkan semua informasi yang tersedia secara publik, termasuk data historis dan informasi fundamental. Namun, jika sentimen investor memiliki informasi yang tidak sepenuhnya tercermin dalam harga saham, hal ini bertentangan dengan bentuk efisiensi tersebut. Investor yang dapat menafsirkan dan menggunakan sinyal sentimen dapat memperoleh keunggulan informasi dibandingkan investor yang hanya mengandalkan analisis fundamental (Baker, 2006).

Literature Review

Sistem keuangan dipengaruhi oleh interaksi sosial dan mencerminkan persepsi investor terhadap suatu peristiwa yang menciptakan sentimen yang masuk ke dalam polar positif, netral maupun negatif, fluktuasi dalam sentimen investor pada akhirnya membentuk perilaku ekonomi secara keseluruhan (Rasheed, 2023). Sentimen melalui media sosial memiliki peran dalam menghubungkan investor dengan pasar saham dengan sifat informasi asimetris (Ali, 2018). Sentimen tersebut merefleksikan reaksi investor yang telah mendapatkan informasi dan

ekspektasi terhadap keuntungan pada saham (Gao, 2023). Penggunaan sentimen sebagai variabel untuk melakukan *forecasting* volatilitas dan *return* saham (Antweiler, 2004).

Berbagai media sosial dapat menjadi sumber untuk mendapatkan sentimen terkait pasar saham. Youtube merupakan media video digital yang memiliki fitur kolom komentar untuk pengguna memberikan opini. Komentar yang diberikan merepresentasikan apresiasi atau bentuk ekspresi tidak setuju terkait dengan topik yang dibahas pada video (Kavitha, 2020). Penelitian terdahulu menggunakan *unsupervised machine learning* untuk menyatakan Youtube dapat memberikan sampel yang relevan berbagai macam bentuk emosi dan ekspresi (Meyer, 2023), dimana Youtube memiliki efek emosi yang menular terhadap video yang memiliki topik terkait Bitcoin, kreator konten video dapat mempengaruhi emosi penonton.

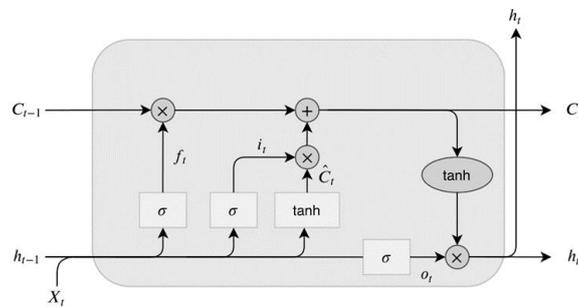
Penggunaan sentimen pada analisis prediksi harga saham telah digunakan dalam penelitian terdahulu menggunakan LSTM sebagai salah satu *machine learning neural network*. (Jing, 2021) menggunakan metode *hybrid* yang menggabungkan polar sentimen dengan indikator saham pada industri keuangan dalam prediksi pergerakan harga saham menemukan model memiliki performa yang lebih baik jika menggunakan sentimen dengan menggunakan pengukuran *mean absolute percentage error*. (Siami-Namini et al, 2018) menyatakan LSTM menghasilkan rata - rata *root mean square error* dibandingkan model dengan metode ARIMA. (Aamir et al, 2018) menemukan LSTM memberikan performa lebih baik dibandingkan dengan model *time series* tradisional dalam akurasi prediksi dengan menggunakan pengukuran *mean absolute error* dan *mean squared error* karena mampu menangkap pola yang kompleks dan dependensi jangka panjang yang dapat menghasilkan prediksi yang lebih akurat. Model LSTM mampu mengatasi jangka panjang dependensi dalam data sekuensial sehingga dapat meningkatkan akurasi dalam prediksi. (Jain, 2018) menemukan LSTM memiliki performa yang lebih baik dalam memprediksi harga saham dibandingkan model ANN dan ARIMA dalam memprediksi harga saham jangka pendek. (K Chen, 2015) menggunakan LSTM untuk prediksi harga pada pasar saham China menggunakan teknikal indikator, seperti, *close price*, *high*, *low*, dan *volume*. Pada model LSTM memerlukan *hyperparameter tuning* untuk menghasilkan hasil yang lebih efektif, (Chen, 2019) menggunakan indikator *relative strength index*, *moving average*, *exponential moving average*, dan *moving average convergence / divergence* serta menguji *hyperparameter tuning* menghasilkan akurasi prediksi yang lebih baik. Struktur memori dalam LSTM dirancang untuk mengatasi masalah gradien pada data yang berurutan (Hochreiter & Schmidhuber, 1997). Parameter yang digunakan pada model LSTM adalah *epochs*, *batch size*, *hidden layer*, dan *activation function*. Setiap sel LSTM mempertahankan vektor keadaan sel yang bertanggung jawab untuk menyimpan aliran data yang telah dilewati, dimana setiap pada langkah waktu selanjutnya sel dapat memilih untuk membaca atau mengulang sel dengan mekanisme *gating*. Terdapat tiga input gerbang biner yang ada pada setiap sel: *Input gate*, *Forget gate*, dan *Output*

gate. Setiap gerbang memiliki aktivasi *sigmoid* sehingga dapat membentuk kurva halus dalam rentang 0 dan 1 dan model tetap memiliki diferensial untuk menentukan informasi yang akan diolah. Setiap gerbang memiliki fungsi sebagai berikut.

Input gate: memiliki fungsi untuk mengendalikan sel memori perlu diperbaharui, dimana biner 1 menunjukkan untuk mempertahankan memori dan biner 0 menunjukkan untuk mengabaikan memori.

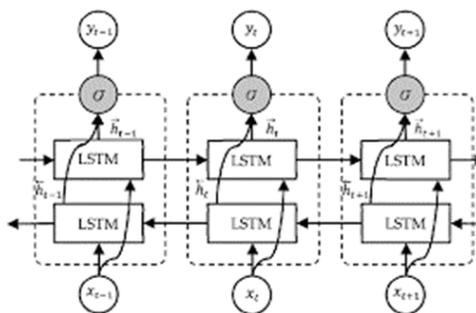
Forget gate: memiliki fungsi untuk mengendalikan berapa banyak informasi pada sel sebelumnya untuk diabaikan.

Output gate: mengendalikan informasi pada sel akan dibuat terlihat dari *hidden state* yang berasal dari sel sebelumnya menjadi keluaran ke sel berikutnya disaring dan data baru berdasarkan keadaan sel.



Gambar 1. Struktur Sel Long Short-Term Memory

Pada penelitian ini membandingkan LSTM dengan model modifikasi *Bidirectional-LSTM*, dimana arsitektur pada model ini memiliki *hidden layer output* untuk mengolah data kedua arah yang mengembangkan nilai fitur lebih panjang dibandingkan dengan model LSTM. Fitur yang lebih panjang dapat meningkatkan kemampuan proses informasi dalam tahap *feed forward neural* untuk mengklasifikasikan dengan lebih akurat.



Gambar 2. Arsitektur Bi-directional LSTM

3. METODE PENELITIAN

Penelitian ini menggunakan metode analisis prediktif. Data yang digunakan merupakan data sekunder yang diperoleh melalui **Youtube** dan **yahoo finance** untuk data saham. Emiten yang digunakan dalam penelitian ini adalah BBKA dan BBRI untuk memprediksi harga penutupan menggunakan sentimen dari fitur komentar Youtube pada video yang membahas terkait aksi korporasi, BBKA melakukan *stock split* pada 13 Oktober 2021 dan BBRI melakukan *right issue* pada 9 September 2021. Pada data saham yang digunakan dari yahoo finance adalah harga penutupan, harga pembukaan, dan volume. teknik indikator, seperti *simple moving average*, *exponential moving average*, *moving average convergence/divergence* dan *relative strength index* diproses menggunakan *library Technical Analysis* dari Python.

Data sentimen diperoleh dari Youtube menggunakan API melalui Rstudio. Proses ekstrak data pada Youtube dengan menggunakan *keyword* yang berkaitan dengan aksi korporasi emiten. Data yang diperoleh dari setiap emiten mencapai 1300 komentar, kemudian setiap komentar akan dipolarisasi menjadi tiga kategori: positif, negatif, dan netral. Selanjutnya, hasil polarisasi digunakan dalam metode *machine learning* klasifikasi yang hasilnya digunakan sebagai *input* dalam prediksi harga saham setiap emiten. Data sentimen diproses melalui tahapan sebagai berikut.

1. *Text Cleaning*

Pada tahap ini karakter teks yang tidak diperlukan akan dihilangkan dalam kalimat seperti tanda baca, angka, k.

2. *Case Folding*

Huruf dalam dokumen teks setiap kalimat akan diubah menjadi huruf kecil.

3. *Tokenization*

Proses pemisahan teks menjadi potongan setiap kata dalam kalimat.

4. *Filtering*

Pada tahap ini mengambil kata-kata tertentu dari hasil tokenization.

5. *Stemming*

Kata yang melewati proses filtering akan dibentuk menjadi bentuk dasar.

Data sentimen yang telah melalui tahapan *preprocessing* akan ditransformasi menjadi bentuk vektor atau *bag of words*. Pada tahap ini merepresentasikan frekuensi istilah yang terdapat dalam kumpulan dokumen, baris matriks mewakili kalimat dari data yang akan dianalisis dan kolom matriks mewakili kata. Dalam matriks ini, total jumlah kemunculan setiap kata dalam seluruh korpus akan dikumpulkan, kemudian data *text* akan diolah menggunakan *supervised*

machine learning untuk klasifikasi dengan menggabungkan polar yang sudah diberikan pada setiap kalimat.

Data sentimen yang telah melalui proses klasifikasi menggunakan *machine learning* digabungkan dengan data harga dan indikator teknikal saham yang akan digunakan dalam prediksi menggunakan LSTM dan Bi-LSTM. Dalam model *neural network* data akan dibagi untuk *training* dan *test*, Emiten BBKA menggunakan rentang 20 hari untuk *training model* LSTM dan Bi-LSTM, sedangkan BBRI menggunakan rentang 26 hari. Setelah dibagi menjadi *training* dan *test*, data akan dinormalisasi. Pada tahap *training network sequence* menggunakan lima *layer* pada kedua metode, kemudian dilakukan *hyperparameter tuning* untuk mendapatkan (/) namun, menggunakan fitur *EarlyStopping* untuk menghentikan training data pada *epoch* tertentu saat hasil tidak memberikan perkembangan, fitur *ReduceLROnPlateu* diaplikasikan untuk mengurangi *learning rate* pada saat hasil tidak memberikan perkembangan. Pada hasil prediksi akan menggunakan *mean absolute error percentage* sebagai metrik untuk mengukur akurasi hasil prediksi terhadap harga aktual.

4. HASIL DAN PEMBAHASAN

4.1 Statistik Deskriptif

Data harga saham setiap emiten yang diekstrak dari Yahoo Finance memiliki periode waktu yang berbeda berdasarkan tanggal jatuh tempo setiap aksi korporasi. Dalam penelitian ini variabel dependen adalah harga penutupan, statistik deskriptif dari harga penutupan saham yang digunakan terdapat pada tabel berikut.

Tabel 1. Statistik Deskriptif Harga Penutupan Saham

	BBKA	BBRI
<i>Mean</i>	7111.77	3889.21
<i>Standard Error</i>	54.67	35.30
<i>Median</i>	7300	3850
<i>Standard Deviation</i>	440.76	277.98
<i>Sample Variance</i>	194269.87	77273.93
<i>Kurtosis</i>	-1.62	-1.03
<i>Skewness</i>	-0.35	0.41

Tabel 2. Statistik Deskriptif Teknikal Indikator BBCA

	<i>RSI</i>	<i>SMA_20</i>	<i>EMA_5</i>	<i>SMA_5</i>	<i>EMA_30</i>	<i>MACD</i>
<i>Mean</i>	57.93	7190.86	6984.31	7122.87	7175.86	91.44
<i>Standard Error</i>	1.80	45.11	48.30	55.36	34.75	11.15
<i>Median</i>	57.37	7226.67	7033.39	7380.00	7210.31	74.80
<i>Standard Deviation</i>	12.83	270.63	389.41	432.37	208.48	89.92
<i>Sample Variance</i>	164.59	73241.67	151637.35	186942.08	43462.76	8086.16
<i>Kurtosis</i>	-0.60	-1.22	-1.83	-1.66	-0.22	-1.03
<i>Skewness</i>	-0.16	-0.42	-0.01	-0.43	-0.87	0.50

Statistik harga penutupan menggunakan periode rentang 65 hari untuk BBCA dan BBRI 62 hari. Pada rentang tersebut kedua saham memiliki rata - rata yang positif menunjukkan bahwa harga penutupan pada kedua saham cenderung memiliki *return* yang positif pada periode sebelum dan sesudah aksi korporasi dilakukan. Namun, BBCA memiliki volatilitas yang lebih tinggi ditunjukkan melalui Standard Deviation (440) yang lebih tinggi dibandingkan dengan BBRI (277). BBCA menunjukkan *skewness* negatif (-0.35 yang artinya bahwa potensi *return* pada saham tersebut cenderung rendah atau dalam *trend* menurun sepanjang rentang periode observasi, sementara BBRI menunjukkan *skewness* positif (0.41) yang menunjukkan harga saham sedang *uptrend* pada periode rentang observasi.

Tabel 3. Statistik Deskriptif Teknikal Indikator BBRI

	<i>RSI</i>	<i>SMA_20</i>	<i>EMA_5</i>	<i>SMA_5</i>	<i>EMA_30</i>	<i>MACD</i>
<i>Mean</i>	59.45	3789.02	3819.30	3879.47	3851.60	34.75
<i>Standard Error</i>	2.66	21.32	20.35	34.89	25.77	9.50
<i>Median</i>	59.29	3738.71	3777.79	3847.63	3786.34	31.23
<i>Standard Deviation</i>	18.41	122.48	160.22	265.70	148.06	74.79
<i>Sample Variance</i>	338.91	15000.82	25670.75	70597.47	21920.71	5592.92
<i>Kurtosis</i>	-1.10	-0.63	0.26	-0.88	-1.18	-0.91
<i>Skewness</i>	0.02	0.90	1.15	0.58	0.64	0.46

Indikator teknikal yang umum digunakan oleh investor dipilih dengan periode jangka pendek karena pada pengamatan setiap saham memiliki periode tidak lebih dari 65 hari, rata - rata *Simple Moving Average* (SMA), dan rata - rata *Exponential Moving Average* (EMA) menggunakan periode 30 hari ke bawah.

Pada tahap model prediksi harga saham menggunakan LSTM dan Bi-LSTM dengan menggabungkan data sentimen yang diekstrak dari Youtube. Sentimen BBKA diambil dari 18 video yang diseleksi berdasarkan topik bahasan *stock split* dan sembilan video yang membahas terkait *right issue* BBRI. Data sentimen diberikan label dengan polar sentimen, interpretasi polar diberikan berdasarkan setiap komen setuju atau tidak setuju terkait dengan aksi korporasi:

1. Polar sentimen positif berdasarkan komentar yang menunjukkan ekspektasi akan menghasilkan *return* yang positif pada saat aksi korporasi, menunjukkan keinginan untuk membeli saham, setuju atau memiliki sudut pandang yang sama terkait topik yang disampaikan pada video.
2. Polar sentimen negatif berdasarkan komentar yang menunjukkan ekspektasi akan menghasilkan *return* yang negatif pada saat aksi korporasi, menunjukkan tidak ada keinginan untuk membeli saham, tidak setuju atau memiliki sudut pandang yang sama terkait topik yang disampaikan pada video.
3. Polar sentimen netral menunjukkan opini atau ekspresi yang netral terhadap aksi korporasi.

Tabel 4. Proporsi Polar Sentimen Saham

	<i>BBKA Sentiment Polarity</i>	<i>BBRI Sentiment Polarity</i>
<i>Positive</i>	208	182
<i>Neutral</i>	1088	1328
<i>Negative</i>	82	107

Tabel 4 menunjukkan polar sentimen positif pada BBKA dan BBRI lebih tinggi daripada polar negatif sentimen. Setelah data sentimen diberikan label digunakan sebagai input pada tahap *preprocessing* dan ditransformasi menjadi bentuk vektor atau *bag of words*, pada vektorisasi untuk merubah data teks menjadi angka yang merepresentasikan frekuensi kata pada setiap kalimat. Vektor diolah dengan algoritma klasifikasi menggunakan *supervised machine learning*, yaitu dengan *Random Forest*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, dan *Logistic Regression*. Model klasifikasi menggunakan 80% dari data untuk tahap *training* dan 20% tahap validasi model.

Tabel 5. Akurasi Model Klasifikasi Sentimen

	<i>Accuracy (BBKA)</i>	<i>Accuracy (BBRI)</i>
<i>RandomForest</i>	76%	82.4%
<i>KNN Classifier</i>	79%	83.3%
<i>SVM</i>	79%	83.6%
<i>Logistic Regression</i>	75.7%	80.2%

Model klasifikasi pada kedua data sentimen saham menunjukkan KNN dan SVM menghasilkan nilai akurasi yang lebih tinggi daripada model lainnya. Pada sentimen BBKA menghasilkan akurasi 79%, sedangkan pada BBRI menghasilkan akurasi yang lebih tinggi sebesar 83.6%. JEvaluasi model klasifikasi selain akurasi, pada penelitian ini juga menggunakan pengukuran *precision*, *recall*, dan *F1-score*.

Tabel 6. Evaluasi Pengukuran Model Klasifikasi Sentimen BBKA

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>RandomForest</i>	0.77	0.77	0.77
<i>KNN Classifier</i>	0.78	0.78	0.78
<i>SVM</i>	0.78	0.78	0.78
<i>Logistic Regression</i>	0.75	0.75	0.75

Tabel 7. Evaluasi Pengukuran Model Klasifikasi Sentimen BBRI

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>RandomForest</i>	0.82	0.82	0.82
<i>KNN Classifier</i>	0.83	0.83	0.83
<i>SVM</i>	0.83	0.83	0.83
<i>Logistic Regression</i>	0.80	0.80	0.80

Pada evaluasi model klasifikasi sentimen KNN dan SVM pada emiten BBKA menunjukkan *precision* merepresentasikan bahwa model dapat memprediksi 78% sentimen dengan tepat dan untuk BBRI sebesar 83%. Skor tersebut juga diperoleh pada metrik *recall* dan *F1-score* untuk setiap emiten. Pada *recall* merepresentasikan bahwa model dapat mengidentifikasi dengan tepat sebesar 83% untuk BBKA dan 78% untuk BBRI.

4.2 Long-Short Term Memory Model

Penelitian ini bertujuan untuk memprediksi rentang satu minggu waktu perdagangan sebelum tanggal jatuh tempo aksi korporasi setiap emiten. Data sentimen dan indikator teknikal menjadi input untuk memprediksi harga penutupan saham H+1. Rentang periode data yang digunakan untuk tahap *training* dalam prediksi harga penutupan BBKA 20 hari dan BBRI 26 hari sebelum rentang satu minggu waktu perdagangan sebelum tanggal jatuh tempo. Data perlu melalui tahapan normalisasi kemudian dipisahkan ke dalam *array X_train* dan *Y_train* dimana *X_train* sebagai training untuk membuat sekuensial input pada target variabel *Y_train*. Model menggunakan parameter *look_back* satu yang memiliki fungsi untuk model mengolah data menggunakan H-1 untuk memprediksi data poin berikutnya yang kemudian dikonversikan menjadi *array* untuk proses dan model *training* berikutnya.

Pada tahapan prediksi model LSTM melalui proses *hyperparameter tuning* untuk mendapatkan hasil yang terbaik, pada setiap model emiten menggunakan lima layer, parameter

untuk BBRI menggunakan 70 *hidden unit* untuk layer pertama sampai layer keempat, pada layer output menggunakan 50 *hidden unit*, 32 *batch size*, dan 100 *epochs*. untuk emiten BBKA menggunakan 80 *hidden unit* pada *layer* pertama dan kedua, dan 70 *hidden unit* pada *layer* ketiga dan keempat, untuk *layer output* menggunakan 60 *hidden unit*, 16 *batch size* dan 100 *epochs*. Pada *sel neural network* model LSTM menggunakan dasar *activation function* yang memiliki fungsi kontrol pada masalah *vanishing gradient*. Pada fase training, model menggunakan parameter *EarlyStopping* untuk kontrol pada *epochs* saat metrik tidak menunjukkan perkembangan. Proses training data emiten BBRI menghasilkan metrik terbaik pada 60 *epochs* dan 61 *epochs* pada data emiten BBKA.

Setelah melalui tahapan training, kemudian model digunakan pada proses validasi untuk memprediksi harga penutupan saham saat jatuh tempo aksi korporasi. Pada model BBKA pada menghasilkan skor MAPE 0.003 dan BBRI menghasilkan skor MAPE 0.011 untuk model yang tidak menggunakan sentimen yang telah melalui proses model klasifikasi dan dibandingkan dengan hasil validasi yang menggunakan sentimen yang telah melalui klasifikasi.

Tabel 8. Skor MAPE Prediksi LSTM dengan Model Klasifikasi

	<i>BBKA</i>	<i>BBRI</i>
<i>LSTM-RandomForest</i>	0.003	0.003
<i>LSTM-KNN Classifier</i>	0.004	0.005
<i>LSTM-SVM</i>	0.005	0.002
<i>LSTM-Logistic Regression</i>	0.011	0.005

Skor MAPE pada model LSTM dengan model klasifikasi BBRI menunjukkan hasil yang lebih baik. Namun, pada BBKA model tanpa klasifikasi menghasilkan skor yang lebih kecil, yang merepresentasikan *error* yang lebih kecil terhadap harga penutupan saham yang aktual. Model *LSTM-SVM* BBRI menghasilkan skor MAPE 0.002 dibandingkan dengan model lain memiliki skor yang lebih baik dari model lainnya. Pada emiten BBKA *LSTM-RandomForest* menghasilkan skor yang sama pada model yang tidak menggunakan sentimen dari model klasifikasi dengan MAPE skor 0.003.

4.3 *Bi-Directional Long-Short Term Memory Model*

Model *Bi-LSTM* menggunakan tahapan metode yang sama dengan *LSTM*. Setiap model emiten menggunakan 5 layer dan proses *hyperparameter tuning*. Pada model emiten BBRI, setiap layernya menggunakan 85 *hidden unit*, 100 *epochs*, dan 16 *batch size*. Proses *training* model menghasilkan valuasi terbaik pada 42 *epochs* untuk model yang tidak menggunakan model klasifikasi pada sentimen Pada model BBKA menggunakan 80 *hidden unit* pada layer pertama

dan 60 *hidden unit* pada layer kedua sampai layer lima, dengan 100 *epochs* dan 16 *batch size*, training pada model ini menghasilkan metrik terbaik pada 45 *epochs*. Pada tahap validasi model *Bi-LSTM* BBRI menghasilkan skor MAPE 0.0048 dan BBKA 0.0001.

Tabel 9. Skor MAPE Prediksi Bi-LSTM dengan Model Klasifikasi

	<i>BBKA</i>	<i>BBRI</i>
<i>Bi-LSTM-RandomForest</i>	0.005	0.001
<i>Bi-LSTM- KNN Classifier</i>	0.001	0.002
<i>Bi-LSTM-SVM</i>	0.001	0.0004
<i>Bi-LSTM-Logistic Regression</i>	0.0003	0.0005

Komparasi hasil skor MAPE dengan model *Bi-LSTM* dengan model klasifikasi menunjukkan model emiten *Bi-LSTM* BBRI lebih superior dibandingkan dengan tidak menggunakan klasifikasi, pada model *Bi-LSTM-SVM* dan *Bi-LSTM-Logistic Regression* menghasilkan skor MAPE paling rendah 0.0004 dan 0.0005. Pada model emiten BBKA, model yang tidak menggunakan model klasifikasi menghasilkan skor MAPE yang lebih baik.

Hasil prediksi menggunakan sentimen pada model LSTM dan *Bi-LSTM* dapat mengacu pada teori *efficient market hypothesis* pada kategori *semi-strong form*. Sentimen yang diperoleh dari fitur komentar pada Youtube video merupakan respon terhadap informasi umum yang dipublikasikan melalui *content creator* yang memberikan pandangan, data, informasi dan opini terkait aksi korporasi pada video. Prediksi pada kedua emiten dapat mendekati trend pada masa jatuh tempo aksi korporasi sehingga hasil prediksi dapat dijadikan sebagai alternatif acuan investor dalam pengambilan keputusan investasi sebagai nilai tambah dalam pertimbangan. Penggunaan sentimen memiliki pengaruh terhadap skor MAPE yang mendekati harga saham aktual, sentimen pada Youtube dapat memberikan nilai tambah yang cukup signifikan untuk membantu investor melihat trend pada pergerakan harga saham.

5. SIMPULAN

Analisis prediksi menunjukkan *Bi-LSTM* memberikan performa prediksi yang lebih baik, menghasilkan skor MAPE yang lebih rendah dibandingkan dengan model LSTM. Model prediksi yang menggunakan sentimen melalui algoritma klasifikasi menghasilkan skor MAPE yang baik, namun pada model BBKA hanya beberapa model klasifikasi yang dapat memberikan skor yang lebih baik, sedangkan pada model BBRI memberikan performa yang lebih dalam prediksi saat menggunakan algoritma klasifikasi. Dapat diasumsikan untuk menghasilkan performa yang konsisten, hasil akurasi pada algoritma klasifikasi perlu menghasilkan skor minimal 80%.

Integrasi algoritma klasifikasi sentimen memiliki dampak yang signifikan pada akurasi prediksi, terutama sekitar tanggal jatuh tempo, dan menyoroti nilai potensial analisis sentimen dalam peramalan harga saham. Menghubungkan temuan ini dengan Hipotesis Pasar Efisien (EMH), studi ini menyarankan bahwa sementara EMH menganggap harga saham mencerminkan semua informasi yang tersedia dan oleh karena itu, tidak mungkin diprediksi secara akurat. Temuan ini menunjukkan potensi model pembelajaran mesin, seperti LSTM dan Bi-LSTM untuk mengekstrak pola dan tren dari data historis dan analisis sentiment. Dengan demikian, dapat meningkatkan akurasi prediksi di luar apa yang dapat diharapkan dalam skenario pasar yang murni efisien.

Dalam kaitannya terkait teori *behavioral finance*, studi ini menambahkan pada tubuh penelitian yang berkembang yang menantang asumsi ketat rasionalitas dan efisiensi dalam pasar keuangan. Dengan menggabungkan analisis sentimen ke dalam model prediksi harga saham, studi ini mengakui pengaruh faktor psikologis, seperti sentimen investor, pada pergerakan pasar. Hal ini sejalan dengan teori *behavioral finance*, yang mengindikasikan bahwa peserta pasar dapat menunjukkan perilaku irasional yang dipicu oleh emosi dan bias kognitif, mempengaruhi harga saham. Sebagai simpulan, studi ini menekankan efektivitas model LSTM dan Bi-LSTM dalam prediksi harga saham, terutama ketika ditambah dengan analisis sentimen. Dengan memanfaatkan teknik pembelajaran mesin yang canggih dan wawasan dari Keuangan Perilaku, studi ini berkontribusi pada pengembangan metodologi dalam prediksi pasar saham, yang potensial menawarkan alat berharga bagi investor dan peserta pasar dalam membuat keputusan yang berbasis informasi.

Keterbatasan Penelitian

Keterbatasan pada penelitian ini terdapat pada jumlah emiten yang dapat diobservasi. Data sentimen yang diperoleh dari Youtube mencakup dua emiten, yaitu BBCA dan BBRI yang pada periode tersebut melakukan aksi korporasi dan memiliki ketersediaan jumlah data yang dapat diobservasi. Pada data sentimen, penelitian ini hanya melakukan observasi pada satu *platform* media sosial sehingga hasil tidak dapat merepresentasikan sentimen melalui *platform* media sosial lain, seperti twitter, Instagram, dan media lainnya yang memiliki fitur komentar serupa.

Implikasi Manajerial

Sentimen dapat dijadikan variabel input dalam prediksi harga saham berdasarkan hasil prediksi pada penelitian ini. Untuk retail, investor dapat mempertimbangkan respon publik pada Youtube yang membahas suatu saham terkait pada aksi korporasi sebagai informasi tambahan dalam pengambilan keputusan investasi. Hasil prediksi juga dapat berkontribusi untuk perusahaan yang berkaitan dengan pasar saham. Penggunaan algoritma LSTM dan Bi-

LSTM dapat memberikan fitur untuk rekomendasi pada pengguna jasa perusahaan sebagai informasi tambahan.

Saran

Berdasarkan hasil penelitian, terdapat saran yang dapat diberikan untuk pengembangan dan penggunaan metode serupa bagi penelitian selanjutnya ialah melakukan komparasi dengan menggunakan data sentimen dari berbagai platform media sosial sehingga dapat melihat media sosial yang dapat memberikan informasi prediksi lebih akurat. Penggunaan sentimen juga dapat digunakan tidak hanya terbatas pada aksi korporasi, namun dapat juga digunakan pada peristiwa lainnya atau secara umum respon publik terhadap suatu berita tentang emiten atau peristiwa ekonomi.

DAFTAR RUJUKAN

- Aamir, M., Khan, S. A., Saleem, Y., & Khalid, S. (2018). Stock Market Prediction Using LSTM. *International Journal of Computer Applications*, 181(38), 34-38.
- Ahmad, Ibrahim, Said., et al. (2020). Movie Revenue Prediction Based on Purchase Intention Mining Using Youtube Trailer Reviews. *Information Processing & Management*, 57.
- Amrani, Yassine., et al. (2018). Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis. *Procedia Computer Science*, 127.
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 61(4), 1645-1680.
- Barberis, N., Shleifer, A., & Vishny, R. W. (1998). A Model of Investor Sentiment. *Journal of Financial Economics*, 49(3), 307-343.
- Daniel, Mariana., Neves, Rui, Ferreira., Horta, Nuno. (2017). Company event popularity for financial markets using Twitter and sentiment analysis. *Expert Systems with Application*, 111-124.
- Fama, E. (1970) Efficient Capital Market: A Review of Theory and Empirical Work. *Journal of Finance*, 25, 382-417.
- Hochreiter, S., Schmidhuber, J. (1997). Long-short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Jain, A., Gupta, P., & Sharma, V. (2018). Comparative Study of LSTM and ARIMA Models for Stock Price Prediction. *Journal of Financial Forecasting*, 3(2), 45-58.
- Jing, Nan., Wu, Zhao., Wang, Hefei., 2021. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert System with Applications*, 178.

- K. Chen, Y. Zhou, F. Dai. (2015). A LSTM-based method for stock returns prediction: A case study of China stock market. *Proceedings of the IEEE International Conference on Big Data, IEEE Big Data 2015 (2015)*, pp. 2823-2824
- Li, Youru., Zhu, Zhenfeng., Kong, Deqian., Han, Huan., Zhao, Yao., (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowledge-Based Systems*, 181.
- Li, Yelin., et al. (2020). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting*, 36.
- Li, Xiadong., Wu, Pangjing., Wang, Wenpeng. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, Vol 57.
- Meyer, Eva, Andrea., Sandnerr, Phillip., Cloutier, Bernard., Welp, Isabell, M. (2023). High on Bitcoin: Evidence of emotional contagion in the YouTube crypto influencer space. *Journal of Business Research*, vol.164.
- Moghar, Adil., Hamiche, Mhamed. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science*, 170.
- Odean, T. (1999). Do Investors Trade Too Much? *The American Economic Review*, 89(5), 1279-1298.
- Ricciardi, Victor., Simon, Helen K. (2000). What is Behavioral Finance? *Business, Education & Technology Journal*, Vol. 2.
- Shleifer, A. (2000). *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press.
- S. Siami-Namini, N. Tavakoli, A. Siami Namin. (2018). A comparison of ARIMA and LSTM in forecasting time series Proceedings of the 17th IEEE. *International Conference on Machine Learning and Applications, ICMLA 2018 (2018)*, pp. 1394-1401
- Zhang, Yin., Jin, Rong., Zhou, Zhi-Hua. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 43-52.