



**Prediksi Saham dengan *Bidirectional* LSTM dan Analisis
Sentimen**

Nurdi Afrianto

18917218

Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer

Konsentrasi Sains Data

Program Studi Informatika Program Magister

Fakultas Teknologi Industri

Universitas Islam Indonesia

2022


Lembar Pengesahan Pembimbing

Prediksi Harga Saham Dengan *Bidirectional* LSTM dan Analisis Sentimen

Nurdi Afrinto

18917218




Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.


Septia Rani S.T., M.Cs.

Lembar Pengesahan Penguji

Prediksi Harga Saham *Bidirectional* LSTM dan Analisis Sentimen

Nurdi Afrinto

18917218

Yogyakarta, 19 Juli 2022

Tim Penguji,

Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.

Ketua

Chandra Kusuma Dewa, S. Kom., M.Cs., Ph.D.

Anggota I

Ahmad Munasir Rafie Pratama, S.T., M.I.T., Ph.D.

Anggota II

Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



Izzati Muhammadiyah, S.T., M.Sc., Ph.D.

Abstrak

Prediksi Harga Saham Dengan *Bidirectional* LSTM dan Analisis Sentimen

Pertumbuhan ekonomi dunia melemah pada tiga bulan pertama pada tahun 2020 yaitu sebesar 3,3%. Hal ini disebabkan akibat penurunan aktivitas konsumsi dan investasi yang seiring dengan ketidakpastian ekonomi global. Pasar modal salah satu opsi yang sangat penting sebagai penggerak ekonomi dan pertumbuhan suatu negara. Dengan teknologi pembelajaran mesin, permasalahan prediksi saham dapat diselesaikan. Deep learning adalah bagian dari pembelajaran mesin yang memiliki performa baik dalam mengelola data. Data saham diambil dari yahoo *finance*. Sentimen publik diambil dari Twitter. Data twitter yang digunakan adalah *tweet* yang mengandung simbol saham (\$) diikuti dengan kode nama perusahaan (\$BBCA). Simbol saham adalah rangkaian huruf unik pada sekuritas yang bertujuan perdagangan. Perancangan model memiliki dua komponen, yaitu BILSTM sebagai arsitektur dasar dari pengembangan model prediksinya dan VADER sebagai algoritma perhitungan nilai sentimen publiknya. Perhitungan sentimen dengan metode VADER perhitungan nilai *compound score*. penelitian menggunakan data dengan menggabungkan data sentimen dan data tanpa sentimen. Pada pelatihan, empat skenario disusun menggunakan data uji yang sama dengan panjang data latih yang berbeda dari sisi panjang jendela datanya. Evaluasi model dilakukan dengan menghitung MSE, RMSE, dan MAPE. Studi sebelumnya menggunakan model LSTM menghasilkan nilai RMSE sebesar 1120,6517. Pada pemodelan yang dilakukan, prediksi model yang dibangun dengan menggabungkan data sentimen menghasilkan nilai MSE terkecil 0,094, nilai RMSE terkecil 0,306 dan nilai MAPE terkecil 23.85. Prediksi model dengan tanpa menggunakan data sentimen menghasilkan nilai MSE terkecil 0.148, nilai RMSE terkecil 0.38, dan nilai MAPE terkecil 30,57. Prediksi dengan menggabungkan data sentimen memiliki nilai *error* lebih baik dibandingkan dengan data tanpa sentimen.

Kata kunci

Saham, analisis sentimen, *bidirectional* lstm, Pergerakan ekonomi.

Abstract

Stock Price Prediction Using Bidirectional LSTM and Sentiment Analysis

World economic growth weakened as much as 3.3 % in the first three months of 2020. It is due to a decrease in consumption and investment activities along with global economic uncertainty. Capital market is one of the most important options important as economic driver and growth of a country. With machine learning technology, stock prediction problems can be done. Deep learning is part of machine learning that performs well in data management. The stock data taken from yahoo finance. Public sentiment data taken from Twitter with the stock symbol (\$) followed by the codenamecompany (\$BBCA). The stock symbol is a unique set of letters on a security for trading purposes. Model design has two components, namely BILSTM as the basic architecture of its prediction model development and VADER as algorithm for calculating the value of public sentiment. Sentiment calculation used VADER calculation of compound score value. This research used combination data with sentiment data and non-sentiment data. In training, there were four scenarios compiled using the same test data with different lengths of training data from the the length of the data window. Model evaluation was done by calculating MSE, RMSE, and MAPE. Previous studies used LSTM model resulted in an RMSE value of 1120.6517. In the modeling, the predictions of the model built with combining sentiment data produces the smallest MSE value of 0.094, the smallest RMSE value of 0.306 and the smallest MAPE value was 23.85. The model prediction did not used the sentiment data and showed the smallest MSE value of 0.148, the smallest RMSE value of 0.38, and the smallest MAPE value of 30.57. The Predictions by combining sentiment data had more error values better than the data without sentiment.

Keywords

Stocks, sentiment analysis, bidirectional lstm, Economic movements.

Pernyataan Keaslian Tulisan

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

Yogyakarta, 06 Juni 2022



Nurdi Afrianto, S.Kom

Daftar Publikasi

Publikasi yang menjadi bagian dari tesis

Sitasi publikasi 1

Kontributor	Jenis Kontribusi
Nurdi Afrianto	Melakukan komputasi dan analisis model Menulis <i>paper</i>
Dhomas Hatta Fudholi	Melakukan analisis model Menulis <i>paper</i>
Septia Rani	Melakukan analisis model Menulis <i>paper</i>

Halaman Kontribusi

Tidak ada kontribusi dari pihak lain”



Halaman Persembahan

{Halaman ini merupakan bagian yang dipergunakan oleh penulis untuk menyampaikan terima kasih kepada pihak yang telah membantu baik secara akademik maupun non-akademik atau bantuan finansial (beasiswa) selama proses studi di Program Studi Teknik Informatika Program Magister. }



Kata Pengantar

Puji dan syukur penulis ucapkan kehadiran Tuhan Yang Maha Esa, atas segala kasih dan karunia-NYA sehingga penulis dapat menyelesaikan penulisan Tesis ini. Adapun judul penelitian pada tesis ini adalah “Prediksi Harga Saham Menggunakan BiLSTM dengan Faktor Sentimen Publik”. Tesis ini dibuat dalam rangka memenuhi syarat untuk mencapai gelar Magister Informatika pada Konsentrasi Sains Data, Universitas Islam Indonesia, Yogyakarta. Penulis sangat menyadari dan merasakan bahwa terwujudnya Tesis ini tidak terlepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan rasa terima kasih dan penghargaan yang tulus kepada: DThomas Hatta Fudholi, S.T., M.Eng., Ph.D. dan Septia Rani S.T., M.Cs. Selaku pembimbing yang ditengah-tengah kesibukannya masih tetap meluangkan waktu untuk memberikan bimbingan, petunjuk, dan mendorong semangat penulis untuk menyelesaikan tesis ini. Penulis juga ingin menyampaikan ucapan terima kasih kepada semua pihak yang terlibat dalam penyelesaian studi ini. Kepada:

1. Rektor Universitas Islam Indonesia, Bapak Prof. Fathul Wahid, S.T., M.Sc., Ph.D. Atas kesempatan dan fasilitas yang diberikan kepada penulis untuk mengikuti dan menyelesaikan pendidikan.
2. Dekan Fakultas Teknologi Industri, Bapak Prof. Dr. Ir. Hari Purnomo, M.T. Atas kesempatan menjadi mahasiswa Program Magister Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.
3. Ibu Izzati Muhimmah, S.T., M.Sc., Ph.D. Selaku Ketua Program Studi Magister Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia.
4. Bapak DThomas Hatta Fudholi, S.T., M.Eng., Ph.D. Selaku Pembimbing Utama penulis, yang telah meluangkan waktunya, dan dengan penuh perhatian memberikan dorongan, bimbingan, saran, serta masukan yang sangat penting.
5. Ibu Septia Rani S.T., M.Cs. Selaku Pembimbing Kedua penulis, yang juga telah meluangkan waktunya, dan dengan penuh perhatian memberikan dorongan, bimbingan, saran, serta masukan yang sangat penting.
6. Ibu/Bapak dosen-dosen pengampu mata kuliah selama penulis menjalankan pendidikan.

7. Orang tua (Bapak Waluyo dan Ibu Suparmi) serta keluarga tercinta yang telah mendidik dengan penuh rasa kasih sayang, dan senantiasa memberi semangat serta dorongan kepada penulis.
8. Team *Deep learning* (Miftha, Aher, Yurio, Malik, Dede, Magfirah), atas support yang telah diberikan, dan juga telah menjadi teman diskusi, dan berbagi keluh kesah selama penyelesaian tesis ini.
9. Rekan-rekan Sains Data (Putri, Magfirah, Yuan, Yohani, Vita, Miftah, Aher), atas support serta diskusi yang informatif, dan menyenangkan selama perkuliahan berlangsung.
10. Sahabat (Fajar Donny Bachtiar S.Kom), yang selalu menghibur, memberikan semangat, serta berbagi cerita dan keceriaan.
11. Rekan-rekan organisasi (Aziz, Agus, Putri), yang selalu menghibur penulis, dan selalu memberikan support serta di diskusi yang menarik.
12. Rekan-rekan mahasiswa pascasarjana, yang tidak dapat penulis sebutkan satu persatu.

Penulis berharap, semoga tesis ini dapat bermanfaat bagi pengembang ilmu pengetahuan dan bagi banyak orang. Penulis juga memohon maaf jika dalam penulisan ini terdapat kekurangan maupun kekeliruan, penulis menerima kritik serta saran yang membangun demi menyempurnakan tesis ini.

Yogyakarta, 19 Juli 2022

Penulis,

Nurdi Afrianto, S.Kom

Daftar Isi

Lembar Pengesahan Pembimbing.....	i
Lembar Pengesahan Penguji.....	ii
Abstrak	iii
Abstract	iv
Pernyataan Keaslian Tulisan	v
Daftar Publikasi	vi
Halaman Kontribusi.....	vii
Halaman Persembahan.....	viii
Kata Pengantar.....	ix
Daftar Isi.....	xi
Daftar Tabel.....	xiii
Daftar Gambar	xiv
Glosarium.....	xv
BAB 1 Pendahuluan.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	4
BAB 2 Tinjauan Pustaka.....	5
2.1 Kajian Pustaka.....	5
2.1.1 Peramalan harga saham dengan sentimen analisis	8
2.2 Landasan Teori	9
2.2.1 Pasar modal	10
2.2.2 <i>Deep Learning</i>	13

2.2.3	<i>Long Short-Term Momory (LSTM)</i>	14
2.2.4	<i>Bidirectional LSTM</i>	16
2.2.5	<i>Vader (Valence Aware Dictionary dan sEntiment Reasoner)</i>	17
2.2.6	Analisis Sentimen	18
2.2.7	Twitter API.....	18
2.2.8	Text Mining.....	19
BAB 3 Metodologi.....		21
3.1	Data.....	21
3.1.1	Lokasi dan Waktu Penelitian.....	21
3.1.2	Populasi Penelitian.....	21
3.2	Langkah - langkah Penelitian	22
3.2.1	Pengumpulan Data.....	22
3.2.2	<i>Pre-processing Data</i>	23
3.2.3	Perancangan.....	28
3.2.4	Implementasi	32
3.2.5	Evaluasi.....	33
BAB 4 Hasil dan Pembahasan.....		35
4.1	Data <i>Pre-processing</i>	35
4.2	Pelatihan Model dan Evaluasi	37
4.3	Pengujian.....	42
4.3.1	Evaluasi kualitatif	46
4.4	Pembahasan	49
BAB 5 Kesimpulan dan Saran.....		51
5.1	Kesimpulan.....	51
5.2	Saran	51
Daftar Pustaka.....		52

Daftar Tabel

Tabel 2.1 Ulasan Kritis	7
Tabel 3.1 Aturan <i>Case Folding</i>	25
Tabel 3.2 Ilustrasi Tahapan <i>Case Folding</i>	25
Tabel 3.3 Aturan Tahapan Tokenisasi	26
Tabel 3.4 Contoh <i>Tokenisasi</i>	26
Tabel 3.5 Kamus Kata Dasar Anomali	27
Tabel 3.6 Ilustrasi <i>Stemming</i>	27
Tabel 3.7 Aturan tahap <i>Filtering</i>	28
Tabel 3.8 Contoh <i>Filtering</i>	28
Tabel 3.9 Rentang waktu pelatihan.....	30
Tabel 4.1 Parameter Dan <i>Loss</i> Pemodelan Dengan Menggunakan Sentimen.....	40
Tabel 4.2 Parameter Dan <i>Loss</i> Pemodelan Tanpa Menggunakan Sentimen.....	40
Tabel 4.2 Hasil Pengujian Empat Skenario Model.....	42
Tabel 4.2 Hasil Pengujian Empat Skenario Model Tanpa Sentimen.....	43

Daftar Gambar

Gambar 2.1	Arsitektur LSTM (Hochreiter and Jürgen Schmidhuber, 1997).....	15
Gambar 2.2	Arsitektur <i>Bidirectional</i> LSTM (Graves et al., 2013).....	17
Gambar 3.1	Tahapan Penelitian.....	22
Gambar 3.2	Model yang diusulkan.....	29
Gambar 3.3	Penggabungan data history saham dan data twitter.....	30
Gambar 3.4	Pemilihan <i>input</i> untuk <i>Bidirectional Long Short-Term Memory</i>	31
Gambar 3.5	<i>Sentiment score</i> menggunakan VADER	32
Gambar 4.1	<i>Kode pre-processing</i> data	35
Gambar 4.2	<i>Code</i> pelabellan menggunakan Vader	36
Gambar 4.3	<i>Code</i> normalisasi data, penggabungan data saham, dan nilai <i>compound</i>	38
Gambar 4.4	Potongan kode proses <i>training</i> data	39
Gambar 4.5	Tangkapan layar saat pelatihan model	40
Gambar 4.6	Nilai <i>Loss</i> pada proses pelatihan	41
Gambar 4.7	Nilai <i>Loss</i> pada proses pelatihan model tanpa data sentimen	42
Gambar 4.8	Gambar Hasil Prediksi	44
Gambar 4.9	Gambar Hasil Prediksi tanpa data sentimen	45
Gambar 4.10	Statistik <i>tweet</i>	46

Glosarium

RNN	- Recurrent Neural Network
LSTM	- Long Short-Term Memory
BILSTM	- Bidirectional Long Short-Term Memory
DLSTM	- Deep Long Short-Term Memory
LSTMP	- Long Short-Term Memory Projected
DLSTMP	- Deep Long short-term memory Projected
AAPL	- Apple Inc.
GOOG	- Google
TSLA	-Tesla
BBCA	- Bank Central Asia
SVM	- Support Vector Machine
covid-19	- Corona Virus 19
SID	- <i>Single Investor Identification</i>
MCDM	- Multi-Criteria Decision-Making
RBF	- <i>Radial Basis Functions</i>

BAB 1

Pendahuluan

1.1 Latar Belakang

Pertumbuhan ekonomi Indonesia mengalami penurunan dari 5,2% pada tahun 2019 dan menjadi 2,1% pada tahun 2020 (Word Bank, 2020). Pelemahan ekonomi dipicu karena penurunan aktivitas konsumsi dan investasi ketidakpastian global (Sumarni, 2020). Ekonomi global menurun dari 1,3% di tahun 2015 menjadi 0,3% pada tahun 2020 (Ahmed et al., 2017). Krisis ekonomi sangat berdampak dari hampir semua sektor, yaitu sektor pembangunan, sektor listrik, sektor gas, sektor industri pengolahan dan perdagangan (Deeyaf & Padjadjaran, 2019). Skenario pertumbuhan ekonomi sebelum covid-19 hanya sebesar 5,3%. Namun, Skenario pertumbuhan ekonomi setelah covid-19 lebih buruk menjadi -0,4% (Sihaloho, 2020).

Pasar modal sangat penting sebagai penggerak ekonomi, pertumbuhan dan perkembangan suatu negara dalam modal jangka Panjang (Osaze BE, 2000). Pasar modal mempunyai dua fungsi dalam perekonomian. Pasar modal sebagai sarana pendanaan usaha dari segi perusahaan dan pasar modal sebagai sarana investasi pada instrumen keuangan (Nasution, 2015). Selain itu, pasar modal mempunyai pengaruh secara signifikan terhadap pertumbuhan ekonomi dan Produk Domestik Bruto (PDB) (Suchy & Soltes, 2018). Pasar modal berguna sebagai pembiayaan alternatif bagi perusahaan dan investasi bagi investor. Oleh sebab itu, pasar modal diperkuat dengan Undang-undang Nomor 8 Tahun 1995 tentang Pasar Modal (Permata & Ghoni, 2019). Berdasarkan laporan Pertumbuhan Single Investor Identification (SID) pada bulan Juli 2018 sampai dengan bulan Juni 2019 mengalami peningkatan yaitu sebanyak 44% dari 1.369.810 investor menjadi 1.971.213. Investor pasar modal meningkat 44% dari bulan juli 2018 hingga juni 2019 (PT Kustodian Sentral Efek Indonesia, 2019). Para investor bertumbuh dari 1.369.810 menjadi 1.971.213 (PT Kustodian Sentral Efek Indonesia, 2019). Investor pasar modal mencakup reksa dana, saham dan surat berharga negara.

Di dalam pasar modal, diperdagangkan berbagai macam instrumen keuangan jangka panjang, salah satunya adalah pasar yang memperjualbelikan saham. Pasar saham dapat mendatangkan keuntungan sebab transaksi jual beli saham merupakan bentuk investasi yang dilakukan banyak orang. Para investor berusaha memilih investasi dengan tingkat resiko yang kecil. Sehingga, investor perlu memiliki ketelitian dan melakukan penilaian terhadap

suatu perusahaan. Para investor membutuhkan suatu cara untuk memprediksi harga saham dalam suatu perusahaan.

Penelitian tentang prediksi dapat menggunakan *data mining* dan *deep learning*. Penelitian menggunakan metode *linear regresi*, *polynomial* dan *radial basis functions* (RBF) dalam memprediksi pergerakan saham (Bhuriya et al., 2017). *Dataset* yang digunakan berasal dari *Tata Consultancy Historical* (TCS) di India. Data saham diurutkan berdasarkan tanggal dengan *library numpy* dan *pandas*. Variabel harga *open*, harga *high*, harga *low* dan jumlah tren digunakan untuk variabel *input* dan harga *close* digunakan untuk variabel target. Nilai *confidence* didapatkan model *linear regresi* sebesar 0.9774. Nilai *confidence* didapatkan model *polynomial* sebesar 0.468. Nilai *confidence* diperoleh model *radial basis functions* (RBF) sebesar 0.5652.

Penelitian menggunakan metode *recurrent neural network* (RNN) dalam memperkirakan pergerakan saham (Jahan & Sajal, 2018). Model menggunakan *dataset* harga penutup saham *Advanced Micro Device* (AMD). *Dataset* yang digunakan selama 168 hari kerja. Data 156 hari digunakan untuk proses *training* dan data 12 hari digunakan untuk proses pengujian. Peeliti menggunakan beberapa *library* antara lain *library numpy*, *pandas*, *Scikit-learn* dan *tensorflow* dalam membangun model. Model yang dibangun dengan *neuron* 500 dan iterasi sebanyak 5000. Nilai *mean squared error* (MSE) didapatkan dibawah 5%.

Penelitian selanjutnya (Mathur et al., 2019) membangun beberapa model *Long short-term memory* (LSTM) dalam memprediksi harga saham. Peneliti menggunakan harga *Adjacent Close* dari saham S&P 500. Variasi model *Long short-term memory* (LSTM) yang dibangun antara lain *Deep Long short-term memory* (DLSTM), *Long Short-Term Memory Projected* (LSTMP), *Deep Long short-term memory Projected* (DLSTMP). *Loss function* digunakan dalam *mean squared error* (MSE) dan optimisasi adam. *Mean squared error* (MSE) pada model LSTMP setelah 3 *epoch* adalah 0,5770. Nilai *mean squared error* (MSE) pada model DLSTM setelah 3 *epoch* adalah 0,00031464. *Deep Long short-term memory* (DLSTM) terpilih menjadi model terbaik. Model DLSTM memiliki arsitektur berupa *cells* 128 *layer* 2. Peneliti menerapkan model pada saham AAPL, GOOG dengan *epoch* 10.

Deep learning baik dalam menangani permasalahan data *time series*. Namun, penelitian terdahulu memberikan saran perlu menambahkan indikator lain. *Sentiment* investor adalah salah satu indikator saham. Sentimen investor mampu membantu memperkirakan pola index saham (Renault, 2017). Indikator inflasi memiliki pengaruh secara positif dan signifikan terhadap pertumbuhan ekonomi Indonesia (Ningsih & Andiny, 2018). Penelitian lain mengungkapkan bahwa sentimen investor dan volatilitas *return* saham Bursa Efek

Johannesburg memiliki hubungan signifikan (Rupande et al., 2019). Bahkan, kasus *covic-19* menyebabkan saham mengalami penurunan dengan nilai signifikansi sebesar $0.00 < 0.05$ (Nurmasari, 2020).

Berdasarkan latar belakang tersebut, didapatkan penelitian yang telah dilakukan dalam prediksi saham. Kontribusi penelitian penulis adalah menambahkan data sentimen investor dalam memprediksi saham. Perbedaan penelitian yang akan dilakukan dengan penelitian sebelumnya adalah terletak pada metode serta data. Penelitian yang akan dilakukan menggunakan *variance* lain *Long short-term memory* (LSTM) yaitu *bidirectional Long short-term memory* (BLSTM) dalam memprediksi saham. Model BILSTM dapat memprediksi lebih baik dibandingkan model LSTM dalam data *time series* (Siami-Namini et al., 2019). Selain itu, model BILSTM mencapai keseimbangan lebih lambat dari pada model LSTM. *Dataset* yang digunakan adalah saham yang berasal dari Indonesia. Perusahaan Indonesia yang terdaftar dalam anggota saham LQ45 yaitu bank BCA (BBCA). Oleh karena itu, dalam penelitian ini penulis bermaksud untuk membuat model prediksi harga saham dan sentimen publik menggunakan algoritma *bidirectional LSTM* dan algoritma *Vader*. model diharapkan menjadi referensi dalam memprediksi pergerakan harga saham.

1.2 Rumusan masalah

Rumusan masalah yang menjadi fokus penelitian ini yaitu bagaimana membangun model prediksi harga saham dengan menambahkan faktor sentimen publik?

1.3 Tujuan Penelitian

Tujuan dilakukannya penelitian ini adalah membandingkan model prediksi saham dengan ditambahkan sentimen analisis dan model prediksi tanpa ditambahkan sentimen analisis yang diharapkan menjadi referensi saintifik dan praktisi dalam menganalisis pergerakan saham kedepan.

1.4 Batasan masalah

Batasan masalah yang menjadi fokus penelitian ini adalah:

1. Penelitian menggunakan data saham perusahaan Bank BCA (BBCA).
2. Data saham diambil dalam situs *yahoo finance* dan sentimen publik diambil dari twitter.
3. Data saham yang diambil adalah data saham pada hari perdagangan saham tersebut aktif.
4. Data twitter yang digunakan hanya *tweet* yang mengandung simbol saham (\$) dan diikuti dengan kode nama perusahaan (\$BBCA).

5. Penelitian hanya mengambil data twitter tanpa lebih lanjut menganalisis portofolio yang dimiliki setiap akun karena hal tersebut bersifat privasi.
6. Masih adanya kemungkinan data twitter yang kotor, akun palsu, dan jumlah data sentimen terbatas karena hanya berasal dari twitter.
7. Tidak ada pembatasan *user* tertentu dalam *tweet* yang dipilih.
8. Data dikumpulkan selama kurang lebih 1 tahun 8 bulan mulai dari 19 februari 2020 sampai dengan 29 september 2021.
9. Algoritma vader digunakan secara general pada data *tweet*.

1.5 Manfaat Penelitian

1. Bagi investor, memberikan pandangan terkait dengan pergerakan saham sehingga dapat memperkecil kerugian dan sebagai pendukung keputusan berinvestasi.
2. Bagi para peneliti, model yang didapatkan kelak dapat dikembangkan dan menjadi referensi baru dalam menganalisis pergerakan saham dengan menambahkan beberapa aspek terkait saham.

BAB 2

Tinjauan Pustaka

2.1 Kajian Pustaka

Pasar modal (*capital market*) adalah merupakan salah satu instrumen ekonomi di suatu negara yang dipengaruhi oleh berbagai peristiwa (Wibowo & Darmanto, 2017). Pasar modal (*capital market*) menjadi sumber pembiayaan yang dibutuhkan oleh pelaku usaha yang memerlukan tambahan modal, juga sebagai alternatif pembiayaan bagi masyarakat investor (Rahmah, 2019).

Adanya sentimen yang digambarkan dalam bentuk berita memiliki pengaruh kuat terhadap pergerakan saham (Alamsyah et al., 2019). *Deep learning* adalah bidang *artificial intelligence* yang terbukti mampu memberikan peningkatan akurasi sebanding dengan penambahan jumlah data (Suyanto, 2019).

Studi yang berkaitan dengan *deep learning* dan *sentiment analysis* (analisis sentimen) telah banyak dilakukan. Namun, penelitian terdahulu berada dalam dua hal yang berbeda. Penelitian ini menggabungkan sentimen investor dan *deep learning* menggunakan metode *bidirectional long-short term memory* (BLSTM) dengan menambahkan faktor sentimen untuk membaca akurasi saham.

Penelitian dalam membangun sebuah model untuk mengamati pergerakan harga saham menggunakan metode regresi linear (Bhuriya et al., 2017). *Dataset* (kumpulan data) yang digunakan adalah saham *Tata Consultancy Historical* (TCS) di India. Variabel *independent* yang digunakan dalam penelitian ini adalah harga pembuka, harga tertinggi, harga terendah, dan jumlah tren. Sementara itu, variabel *dependent* target yang digunakan adalah harga penutup. Penelitian ini menggunakan tiga metode regresi linear yang mendapatkan nilai *confidence* sebesar 0,97, metode polynomial yang mendapatkan nilai *confidence* sebesar 0,468., dan Metode RBF yang mendapatkan nilai *confidence* sebesar 0,5652.

Dalam penelitiannya, (Izzah et al., 2017) menggunakan metode regresi linear untuk memprediksi pergerakan harga saham perusahaan. *Dataset* yang digunakan adalah saham *Jakarta Composite Index* (JCI). Metode regresi linear berganda (*multiple linear regression*) menunjukkan nilai *mean square error* sebesar 15087,465, nilai *root mean square error* sebesar 122,831, serta nilai *mean absolute percentage error* sebesar 3,255. Namun, data saham yang bersifat deret waktu mengakibatkan penggunaan model linear masih belum

cukup, sehingga model *deep learning* lebih unggul daripada model linear (Hiransha et al., 2018).

Salah satu metode *deep learning* yaitu *recurrent neural network* (RNN) dapat digunakan dalam memprediksi saham (Jahan & Sajal, 2018). Penelitian ini menggunakan data harga penutup saham *Advanced Micro Device* (AMD). Data diambil selama 168 hari kerja. Data *training* yang digunakan sebanyak 168 hari, sedangkan data *testing* yang digunakan sebanyak 12 hari. Model dibangun menggunakan *neuron* 500 dan iterasi sebanyak 5000. Model tersebut berhasil mendapatkan hasil *mean squared error* (MSE) dibawah 5%.

Meskipun demikian, arsitektur metode RNN juga memiliki kekurangan, yaitu tidak dapat memproses *sequence* yang panjang. Model RNN juga memiliki permasalahan terkait *vanishing gradient* dan *exploding gradient*. Nilai *gradient* didapatkan dari hasil aktivasi fungsi *tanh* dengan *range* [-1,1]. *Vanishing gradient* adalah keadaan ketika hasil perkalian beberapa *gradient* bernilai nol. Sementara itu, *exploding gradient* adalah keadaan ketika bila nilai bobot pada $W > 1$ sehingga, nilai *gradient* akan terus membesar (Suyanto et al., 2019).

Metode *Long Short-Term Memory* (LSTM) dapat mempelajari pola dari data deret waktu. Arsitektur LSTM juga dapat mengatasi permasalahan *vanishing gradient* dan *exploding gradient*.

Dalam penelitiannya, (Ghosh et al., 2019) menggunakan metode LSTM telah dilakukan menggunakan *dataset Bombay Stock Exchange* (BSE). Harga penutupan diprediksi dari 5 perusahaan yang bergerak di berbagai sektor. Model dilatih menggunakan data selama 3 bulan, 6 bulan, 1 tahun dan 3 tahun. Nilai error pada setiap perusahaan menunjukkan penurunan selama 3 bulan hingga 3 tahun. Nilai error terbaik yang didapatkan yaitu sebesar 0,874805.

Dalam sebuah penelitian, (Mathur et al., 2019) menggunakan beberapa arsitektur LSTM, antara lain model *deep long short-term memory* (DLSTM), model *long short-term memory projected* (LSTMP), dan model *deep long short-term memory Projected* (DLSTMP). *Dataset* saham yang digunakan adalah saham Apple Inc. (AAPL), saham Google (GOOG) dan saham Tesla, Inc (TSLA). Percobaan dilakukan sebanyak 5 kali. Nilai *epoch* yang dipakai sebanyak 10 *epoch*. Arsitektur LSTMP dan DLSTMP menghasilkan nilai *loss function* MSE masing-masing sebesar 0,5770 dan 0,00031. Model LSTM dan RNN menunjukkan hasil lebih baik dibandingkan dengan *machine learning* pada umumnya.

Metode *autoregressive integrated moving average* (ARIMA) dan LSTM juga diimplementasikan pada perusahaan luar negeri (Joosery & Deepa, 2019). *Dataset* yang digunakan adalah saham GOOGL (*alphabet.inc*), saham NKE (*nike.inc*), saham NOK (*nokia*

oyj) dan saham SNE (*sony corp*). Model dilatih dengan menggunakan data selama 1 bulan, 3 bulan, 6 bulan, 1 tahun, 5 tahun dan 10 tahun. Evaluasi pada penelitian menggunakan *mean squared error* (MSE). Model ARIMA mendapatkan akurasi sebesar 96,766%. Model LSTM mendapatkan akurasi sebesar 97,549%. Model *attention* LSTM mendapatkan akurasi sebesar 98,070%. Penulis mendeskripsikan ulasan kritis penelitian terdahulu seperti pada Tabel 2.1.

Table 2.1 Ulasan Kritis

No.	Sub Tema	Keyword	Saran Penelitian	Pustaka
1.	<i>Stock Market Prediction Using a Linear Regression</i>	<i>Stock Market, Forecast, Data Mining, linear regression, Data set.</i>	Penelitian kedepan mencoba metode lain yang bersifat <i>nonlinear</i> .	(Bhuriya et al., 2017)
2.	<i>Mobile App for Stock Prediction Using Improved Multiple Linear Regression</i>	<i>Android, Mobile App, Moving Average, Prediction, Regression, Stock</i>	Penelitian ke depan mencoba metode <i>non linier</i> untuk prediksi agar menambah variasi metode.	(Izzah et al., 2017)
3.	<i>Stock Price Prediction using Recurrent Neural Network (RNN) Algorithm on Time-Series Data</i>	<i>Stock market, RNN, finance, Big Data.</i>	Penelitian ke depan menambahkan variabel lain (<i>sentiment investor</i>) guna mendukung akurasi dan varians data.	(Jahan & Sajal, 2018)
4.	<i>Extreme Market Prediction for Trading Signal with Deep Recurrent Neural Network</i>	<i>Recurrent neural networks, Deep learning, Financial</i>	Penelitian ke depan menggunakan variance RNN guna menambah variance model dan akurasi.	(Zhichen Lu et al., 2018)
5.	<i>Stock Price Prediction Using LSTM on Indian Share Market</i>	<i>LSTM, Stock price, Indian market</i>	Parameter penelitian perlu ditambahkan lain (contoh: <i>sentiment analysis</i>) yang tidak berkorelasi dengan harga saham supaya meningkatkan akurasi prediksi.	(Ghosh et al., 2019)

6.	<i>Stock Market Price Prediction Using LSTM RNN</i>	<i>Recurrent neural network, long short-term memory, Trading, Portfolio optimization</i>	Pengumpulan data harus lebih relevan. Penelitian kedepan perlu menambahkan ulasan produk agar ke depan prediksi semakin akurat.	(Mathur et al., 2019)
7.	<i>Comparative analysis of time-series forecasting algorithms for stock price prediction</i>	<i>Computing Classification System (CCS), Computing methodologies, Machine learning, Machine learning algorithms, Ensemble methods, Boosting</i>	Penelitian ke depan menggunakan <i>variance</i> dari LSTM (GRU atau Bilstm). Faktor eksternal perlu ditambahkan seperti sentimen pasar agar hasil lebih akurat.	(Joosery & Deepa, 2019)

2.1.1 Peramalan harga saham dengan sentimen analisis

Penelitian dalam memperkirakan harga saham dengan sentimen analisis menggunakan metode Naive Bayes dan SVM (Kordonis et al., 2016). *Dataset* yang digunakan adalah 16 perusahaan teknologi teratas dari *yahoo finance*. Pengumpulan data twitter menggunakan Twitter Search API. Pada proses *preprocessing* data dilakukan *tokenization*, *removing stopwords* dan penghapusan simbol pada twitter. Metode SVM memiliki akurasi yang tinggi dalam memprediksi sentimen setiap *tweet*. Setelah memprediksi sentimen *tweet*, peneliti membuat confusion matrix untuk prediksi saham menggunakan skor sentimen dan perubahan harga history saham. SVM digunakan untuk prediksi harga saham. Hasil akurasi mencapai 87% dengan tingkat error di bawah 10%.

Dalam penelitiannya, (Khedr et al., 2017) memprediksi tren pasar dimasa depan menggunakan metode K-Nearest Neighbor (KNN). *Dataset* yang digunakan adalah saham yahoo, Microsoft Corporation MSFT dan Facebook. Analisis sentimen berita dilakukan dengan menggunakan algoritma Naive Bayes. Vector berita dan data numerik saham digabungkan berdasarkan tanggal yang sama. vektor gabungan yang berisi data numerik dan sentimen dilatih menggunakan algoritma K-Nearest Neighbor untuk memprediksi harga

saham di masa depan. Hasil yang didapatkan model untuk memprediksi pasar saham di masa depan memperoleh akurasi hingga 89,80%.

Dalam sebuah penelitian, (Batra & Daudpota, 2018) menggunakan SVM untuk memprediksi pergerakan harga saham. *Dataset* yang digunakan adalah perusahaan Apple. Data sentimen diambil dari stockTwits. *StockTwits* adalah platform jejaring sosial yang penggunaannya mencapai 40 juta. Data tweet diambil dari tahun 2010 hingga tahun 2017. Pengambilan data saham Apple dari yahoo finance dari tahun 2010 sampai 2017. Pada proses *preprocessing* data dilakukan *tokenization*, *removing stopwords* dan penghapusan simbol pada twitter. Analisis sentimen dilakukan dengan menggunakan metode SVM. Dalam satu hari jika *tweet* positif lebih banyak dari *tweet* negatif maka sentimen bernilai positif. Penggabungan data sentimen dan data saham berdasarkan tanggal yang sama. Hasil menunjukkan akurasi proses pelatihan sebesar 75,22% dan akurasi testing sebesar 76,68%.

Penelitian menganalisa hubungan antar sentimen berita dan return saham telah dilakukan (Alamsyah et al., 2019). *Dataset* dikumpulkan dari 5 situs berita yaitu Detik.com, Kompas.com, Liputan News, dan Merdeka.com yang terkait dengan 20 perusahaan di indeks LQ45. Data dikumpulkan selama periode 2013 sampai 2018. Proses klasifikasi sentimen menggunakan metode Naïve Bayes dan Support Vector Machine. Pada *preprocessing* data dilakukan *tokenization*, *stemming*, dan *stopword removal*. Pergerakan return saham periode Februari 2013-2018 sebesar 46% menunjukkan sentimen positif. Selanjutnya 43% terbentuk sentimen negatif. Prediksi klasifikasi Naïve Bayes didapatkan sebesar 71,42% sedangkan Support Vector Machine didapatkan sebesar 78,87%.

Perbedaan penelitian yang akan dilakukan dengan penelitian sebelumnya adalah terletak pada metode serta data. Penelitian yang akan dilakukan menggunakan *variance* lain *Long short-term memory* (LSTM) yaitu *bidirectional Long short-term memory* (BLSTM) dalam memprediksi saham. Model BILSTM dapat memprediksi lebih baik dibandingkan model LSTM dalam data *time series* (Siami-Namini et al., 2019). Sentimen analisis menggunakan algoritma Vader.

2.2 Landasan Teori

Landasan teori menjadi dasar terpenting di setiap menjalankan penelitian ilmiah. Landasan teori wajib digunakan dalam setiap penelitian.

2.2.1 Pasar Modal

Pasar Modal adalah kegiatan yang bersangkutan dengan penawaran dan perdagangan efek, perusahaan publik yang berkaitan dengan efek yang diterbitkan, serta lembaga dan profesi yang berkaitan dengan efek (Undang-Undang Republik Indonesia, 2016). Investasi pada sektor pasar modal adalah sumber alternatif pendanaan baik bagi pemerintah maupun swasta. Pihak pemerintah menerbitkan obligasi untuk dijual kepada masyarakat melalui pasar modal. Demikian swasta yang membutuhkan dana dapat menerbitkan saham dan obligasi (Nasution, 2015). Instrumen keuangan yang diperdagangkan di pasar modal merupakan instrumen jangka Panjang (lebih dari satu tahun) yaitu: saham, obligasi, reksadana, dan berbagai instrumen derivative seperti *option*, *futures*, *waran*, *right*, dan lain-lain.

1. Jenis-jenis pasar modal

Pasar modal dibagi menjadi empat macam yaitu:

a. Pasar perdana (*Primary market*).

Pasar perdana adalah pasar di mana efek-efek diperdagangkan untuk pertama kalinya, sebelum dicatat di bursa efek. Di sini, saham dan efek lainnya untuk pertama kalinya ditawarkan kepada investor oleh pihak penjamin emisi (*Underwriter*) melalui perantara pedagang efek (*Broker-Dealer*) yang bertindak sebagai agen penjual saham.

b. Pasar sekunder.

Pasar sekunder merupakan pasar dari efek yang telah dicatatkan di bursa. Dengan kata lain, pasar sekunder merupakan tempat investor dapat melakukan jual beli efek, setelah efek tersebut dicatatkan di bursa sebagai kelanjutan dari pasar perdana.

c. Pasar ketiga (*third market*).

Pasar ketiga atau disebut juga *over the counter* (OTC) adalah sarana transaksi jual-beli efek antara anggota bursa (*market maker*) serta investor dan harga dibentuk oleh market maker.

d. Pasar keempat (*fourth market*).

Pasar keempat adalah sarana transaksi jual-beli antara investor jual dan investor beli tanpa melalui perantara efek. Pasar keempat dilaksanakan oleh para investor besar karena dapat menghemat biaya transaksi daripada jika dilakukan di pasar sekunder karena dapat menghemat biaya.

2. Manfaat pasar modal.

Melalui fungsi ekonomi, pasar modal menyediakan tempat atau fasilitas yang mempertemukan dua pihak. Pihak yang memiliki kelebihan dana (investor) dan pihak yang memerlukan dana (emiten). Pihak emiten dapat memperoleh sejumlah dana dari investor dan investor mengharapkan adanya imbal hasil (return) (Muklis, 2016). Emiten dapat memanfaatkan dana yang didapat untuk keperluan operasi maupun investasi perusahaan tanpa harus menunggu tersedianya dana dari operasi perusahaan. Melalui fungsi keuangan, pasar modal memberikan kesempatan untuk memperoleh return bagi investor, sesuai dengan investasi yang dipilih (Muklis, 2016).

3. Saham

Saham didefinisikan sebagai tanda penyertaan atau kepemilikan seseorang atau badan pada suatu perusahaan (Muklis, 2016). Dengan menyertakan modal tersebut, maka pihak tersebut memiliki klaim atas pendapatan perusahaan, dan berhak hadir dalam rapat umum pemegang saham (RUPS) (Putri, 2015). Pada sisi yang lain, saham merupakan instrumen investasi yang banyak dipilih para investor karena mampu memberikan tingkat keuntungan yang menarik (Saputra, 2018). Dilihat dari kinerja perdagangan, saham dapat dibedakan atas:

- a. *Blue-Chip Stock* adalah saham biasa dari perusahaan yang memiliki reputasi tinggi, sebagai leader di industri sejenis, memiliki pendapatan yang stabil dan konsisten dalam membayar *dividen*.
- b. *Income Stocks* adalah saham dari suatu perusahaan yang memiliki kemampuan membayar *dividen* lebih tinggi dari rata-rata *dividen* yang dibayarkan pada tahun sebelumnya.
- c. *Growth Stocks* adalah saham dari perusahaan yang memiliki pertumbuhan pendapatan yang tinggi, sebagai leader di industry sejenis yang mempunyai reputasi tinggi.
- d. *Speculative Stocks* adalah saham dari perusahaan yang tidak secara konsisten memperoleh penghasilan dari tahun ke tahun, akan tetapi mempunyai kemungkinan penghasilan yang tinggi di masa yang akan datang meskipun belum pasti.
- e. *Counter Cyclical Stocks* adalah saham yang tidak terpengaruh oleh kondisi ekonomi makro maupun situasi bisnis secara umum.

4. Keuntungan membeli saham.

Secara umum keuntungan yang diperoleh investor dengan membeli saham atau keuntungan investasi saham adalah:

a. *Capital Gain*

Capital gain adalah selisih positif antara harga jual dengan harga beli. Contoh: Tuan Dika membeli saham Telkom Tbk. Pada harga Rp4,200, - per lembar saham, kemudian satu hari kemudian satu hari kemudian saham tersebut dijual pada harga Rp4,300,- per lembar saham. Penjualan saham tersebut tuan Dwiky mendapatkan *capital gain* sebesar Rp100, - per lembar saham (belum dengan perhitungan pajak dan komisi). Formulasinya terdapat pada persamaan (2.1).

$$\text{Capital Gain} = \frac{P_1 - P_0}{P_0} \quad (2.1)$$

Dimana P_1 adalah harga saham periode t , P_0 adalah harga saham awal *Capital gain* adalah diperoleh hanya ketika menjual aset dengan harga lebih tinggi.

b. Dividen

Dividen adalah sebagian keuntungan perusahaan yang dibagikan kepada pemegang saham. Jumlah dividen yang akan dibagikan diusulkan oleh dewan direksi dan disetujui di dalam rapat umum pemegang saham. Jenis *dividen* terdiri atas:

1. *Dividen* tunai, *dividen* yang dibagikan oleh perusahaan kepada pemegang saham dalam bentuk uang untuk setiap saham yang dimiliki.
2. *Dividen* saham, *dividen* yang dibagikan oleh perusahaan kepada para pemegang saham dalam bentuk saham, dibayar dari laba tunai yang ditahan.
3. *Dividen* bonus, *dividen* yang dibagikan oleh perusahaan kepada para pemegang saham dalam bentuk saham dan dibayar dari selisih agio saham.

5. Risiko memiliki saham.

Saham dapat memberikan peluang keuntungan tinggi, tetapi memiliki risiko yang tinggi juga. Risiko memiliki saham adalah:

a. Tidak mendapatkan *dividen*.

Jika perusahaan tidak bisa menghasilkan keuntungan, maka perusahaan tidak akan membagikan *dividen* kepada pemegang saham karena laba yang diperoleh akan dipergunakan untuk ekspansi usaha.

b. *Capital loss*

Capital Loss adalah selisih negatif antara harga jual dengan harga beli. Investor akan mengalami *capital loss*, jika harga beli saham lebih besar dari harga jual.

c. Risiko Likuiditas

Jika emiten bangkrut atau dilikuidasi, para pemegang saham memiliki hak klaim terakhir terhadap aktivitas perusahaan setelah seluruh kewajiban emiten dibayar. Skenario terburuk adalah jika tidak ada lagi aktivitas yang tersisa, maka para pemegang saham tidak memperoleh apa-apa.

d. Saham *Delisting* dari Bursa.

Karena beberapa alasan tertentu, saham dapat dihapus pencatatannya (*delisting*) di bursa, sehingga pada akhirnya saham tersebut tidak dapat diperdagangkan.

2.2.2 *Deep Learning*

Deep learning pertama kali diperkenalkan oleh Geoffrey Hinton pada tahun 2006, dengan sebuah jaringan saraf yang disebut *deep belief nets*. Perkembangan *Deep learning* semakin cepat setelah ditemukan teknik implementasi yang lebih praktis menggunakan *graphics processing unit* (GPU) pada tahun 2009. *Big data* dapat digunakan dalam meningkatkan kinerja *deep learning* terlebih perangkat dan arsitektur komputasi dari segi CPU dan GPU memungkinkan pelatihan model *deep learning* dengan fitur data besar (Zhang et al., 2018). *Deep learning* memiliki banyak kelebihan diantaranya adalah universal, tahan terhadap variasi data, serta memiliki tingkat generalisasi dan skalabilitas yang tinggi.

Universal deep learning dapat diimplementasikan pada hampir setiap ranah aplikasi. *Deep learning* memiliki kemampuan *universal learning*. *Robust deep learning* bersifat tahan terhadap berbagai variasi data yang secara alami memang sesuai fakta yang ada. *Deep learning* tidak memerlukan fitur yang didesain secara artifisial. Bahkan, fitur optimal dapat dipelajari secara otomatis. *Generalization deep learning* bisa digunakan pada beragam aplikasi dan tipe data karena *deep learning* memiliki kemampuan *transfer learning*, Hal ini sangat berguna ketika tidak tersedia data yang cukup untuk suatu masalah yang akan diselesaikan. *Scalability deep learning* memiliki skalabilitas yang tinggi. Sebuah jaringan ResNet, yang dibuat oleh Microsoft terdiri atas 1.202 lapisan dan diimplementasikan pada sebuah supercomputer (Y. J. Kumar et al., 2016). Namun, Resnet dapat diimplementasikan pada sistem paralel dengan ribuan komputer, seperti yang dilakukan oleh Lawrence Livermore National Laboratory (Van Essen et al., 2015).

Deep learning merupakan pembaharuan konsep dan ide-ide lama, mulai dari *Convolution Neural Networks* (CNN), *capsule networks* atau *capsNet*, *deep recurrent Networks*, hingga *deep reinforcement learning*. *Deep learning* bisa dikatakan sebagai modernisasi machine learning untuk menangani *big data* (Suyanto et al., 2019).

Secara umum, berdasarkan Teknik pembelajarannya, *deep learning* dapat dikelompokkan menjadi empat pendekatan (Suyanto et al., 2019). *Deep unsupervised learning*, *deep supervised learning*, *deep semi supervised learning*, dan *deep learning reinforcement learning*.

1. *Supervised Learning*.

Pada pendekatan ini, *deep learning* membangkitkan suatu fungsi yang memetakan *input* (dengan sejumlah atribut) ke *output* (label kelas) berdasarkan data berlabel yang diberikan. *Deep learning* jenis ini digunakan untuk menyelesaikan masalah klasifikasi maupun regresi.

2. *Unsupervised Learning*.

Pada pendekatan ini, *deep learning* memodelkan sekumpulan *input* secara otomatis tanpa panduan (yang berupa *output* yang diinginkan). Artinya, data-data yang dipelajari hanya berupa *input* tanpa label kelas. *Deep learning* jenis ini biasanya digunakan untuk masalah klusterisasi (*Clustering*).

3. *Semi supervised learning*.

Pada pendekatan ini, *deep learning* menggunakan *sample-sample input* yang sebagian memiliki label dan Sebagian memiliki label dan Sebagian lainnya tidak berlabel. *Deep learning* jenis ini membangkitkan suatu fungsi atau pengklasifikasi yang tepat berdasarkan semua *sample input* yang diberikan.

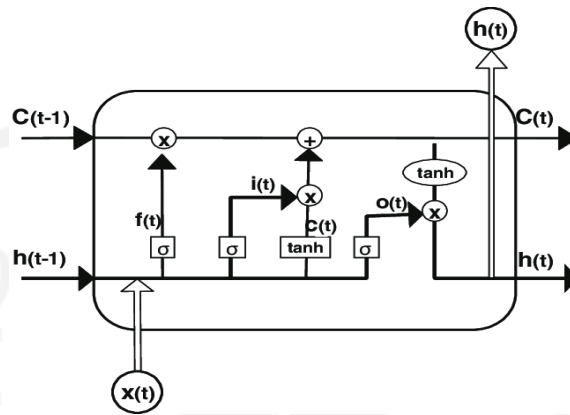
4. *Reinforcement Learning*.

Pendekatan *deep reinforcement learning* ini mempelajari suatu kebijakan bagaimana melakukan aksi berdasarkan hasil pengamatan terhadap lingkungan yang ada. Setiap aksi menghasilkan akibat bagi lingkungan tersebut, dan lingkungan memberikan umpan balik (*feedback*) untuk memandu *deep reinforcement learning*.

2.2.3 *Long Short-Term Memory* (LSTM)

Long Short Term Memory (LSTM) pertama kali dikenalkan oleh Hochreiter dan Schmidhuber dan saat ini menjadi arsitektur RNN yang populer digunakan (Suyanto et al., 2019). LSTM dapat melakukan pelatihan dan mengatasi masalah *vanishing gradient* yang menjadi kesulitan pada *Recurrent Neural Network* (RNN). LSTM dapat mempelajari pola

panjang dari data berurut karena mencegah situasi *vanishing gradient* (Putra, 2019). Tetapi, LSTM tetap memiliki prinsip yang sama dengan RNN dan yang membedakan dengan RNN yaitu isi *cellnya*. *Recurrent Neural Network* (RNN) sederhana karena dengan *cell* yang hanya berisi 1 *layer neuron* dengan fungsi aktivasi *tanh*. Pada LSTM yang dilihat pada Gambar 2.1, menjadi lebih kompleks karena isi *cell* lebih dari satu *layer neuron*.



Gambar 2.1 Arsitektur LSTM (Hochreiter and Jürgen Schmidhuber, 1997)

Ide dasar dari LSTM adalah adanya jalur yang menghubungkan antara *cell state* (C_{t-1}) sebelumnya dengan *cell state* yang sekarang (C_t). Dengan jalur tersebut, suatu informasi pada *cell state* dapat dengan mudah diteruskan ke *cell state* berikutnya dengan beberapa modifikasi yang diperlukan. Nilai *cell state* merupakan vektor yang dirancang untuk menyimpan informasi tentang konteks dari suatu dari suatu sekuen data.

Langkah pertama dalam LSTM adalah menentukan informasi yang akan dibuang dari *cell state* (C_{t-1}) menggunakan fungsi sigmoid yang disebut sebagai *forget gate* (f_t). Nilai nol menandakan bahwa informasi akan dibuang sedangkan satu berarti informasi diteruskan. Formulasinya terdapat pada persamaan (2.1).

$$f_t = \sigma(w_f \cdot [s_{t-1}, x_t] + b_f) \quad (2.2)$$

Di mana f_t adalah *forget gate*, σ adalah fungsi sigmoid, w_f adalah nilai *weight* untuk *forget gate*, s_{t-1} adalah nilai *output* sebelum order ke t , x_t adalah nilai *input* pada order ke t , b_f adalah nilai bias pada *forget gate*.

Langkah berikutnya adalah menentukan informasi yang akan ditambahkan ke *cell state* (C_t). Langkah ini memproses hasil penggabungan dari s_{t-1} dan x_t menggunakan dua fungsi, yaitu fungsi *sigmoid* sebagai *input gate* dan fungsi *tanh* sebagai *intermediate gate*. Hasil dari kedua fungsi tersebut dikalikan untuk mendapatkan informasi yang akan ditambahkan pada *cell state* (C_t). Persamaan matematika dari kedua fungsi terdapat pada persamaan (2.2).

$$i_t = \sigma(w_i \cdot [s_{t-1}, x_t] + b_i) \quad (2.3)$$

$$C_t = \tanh(w_c \cdot [s_{t-1}, x_t] + b_c) \quad (2.4)$$

$$C_t = f_t * C_{t-1} + i_t * C_t \quad (2.5)$$

Dimana i_t adalah *input gate*, w_i adalah nilai *weight* untuk *input gate*, s_{t-1} adalah nilai *output* sebelum order ke t , x_t adalah nilai *input* pada order ke t , b_i adalah nilai *bias* pada *input gate*. C_t merupakan *cell state*, \tanh merupakan fungsi *tanh*, w_c merupakan nilai *weight* untuk *cell state*, b_c merupakan nilai *bias* untuk *cell state*, C_{t-1} adalah *cell state* sebelum order ke t .

Setelah itu, tambahkan dengan *output* dari *forget gate* pada langkah pertama. Langkah terakhir adalah menentukan *output* dari unit LSTM. Untuk menghasilkan *output*, perlu menghitung *sigmoid* dari gabungan s_{t-1} dan x_t yang disebut sebagai *output gate*. *Output gate* ini menentukan seberapa besar nilai dari *cell state* akan dihasilkan pada s_t . kemudian hitung nilai fungsi *tanh* dari *cell state* (C_t) dan kalikan dengan nilai dari *output gate*. Hasil perkaliannya tersebut menjadi *output* dari unit LSTM yang terdapat pada persamaan (2.5)

$$o_t = \sigma(w_o \cdot [s_{t-1}, x_t] + b_o) \quad (2.6)$$

$$s_t = o_t * \tanh(C_t) \quad (2.7)$$

Dimana o_t adalah *output gate*, σ adalah fungsi aktivasi *sigmoid*, w_o adalah nilai *weight* untuk *output gate*, s_{t-1} adalah nilai *output* sebelum order ke t . x_t adalah nilai *input* pada order ke t . b_o adalah nilai *bias* pada *output gate*, s_t adalah nilai *output* order t , o_t adalah *output gate*, \tanh adalah fungsi *tanh* C_t adalah *cell state*.

2.2.4 Bidirectional Long Short-Term Memory (BLSTM)

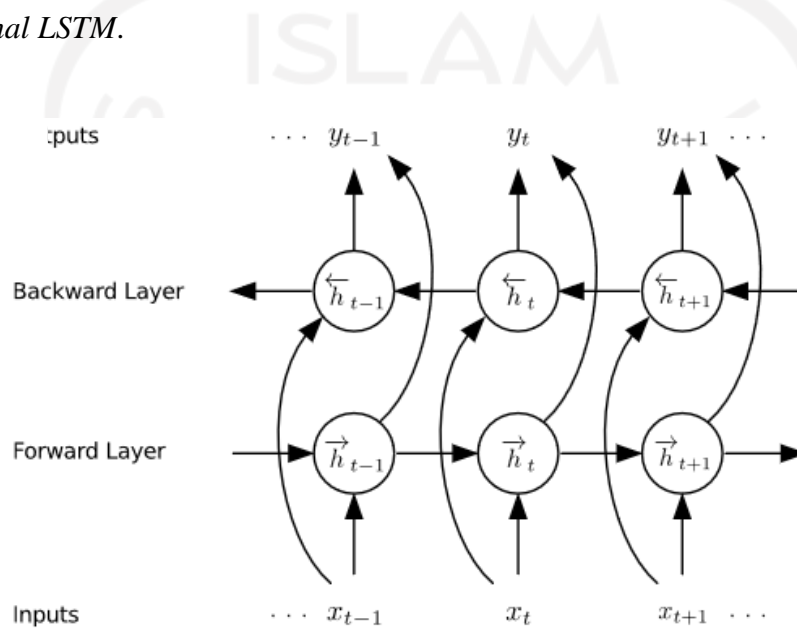
Bidirectional lstm adalah variasi dari *recurrent neural network* (RNN) yang dikembangkan oleh Schuster dan Paliwal untuk melatih model *neural network* menggunakan *sequential data* dengan *input* masa lalu dan masa depan (Mike Schuster and Kuldip K. Paliwal, 1997). *Bidirectional LSTM* adalah jaringan syaraf dari *long short-term memory* (LSTM) yang terdiri dari dua lapisan jaringan saraf LSTM. Lapisan LSTM lanjutan untuk memodelkan konteks sebelumnya dan lapisan LSTM terbelakang untuk memodelkan masing-masing berikutnya (Ertugrul & Karagoz, 2018). Pada kasus klasifikasi, *bidirectional lstm* lebih akurat dan lebih cepat dibandingkan LSTM (Graves & Schmidhuber, 2005). *Bidirectional LSTM* memanfaatkan konteks sebelumnya dan konteks setelahnya dengan memproses data dari dua arah dengan *hidden layer* terpisah (Maimaiti et al., 2017). *Bidirectional LSTM* dikatakan tumpukan dari LSTM. *Forward layer* digunakan untuk konteks sebelumnya dan

backward layer digunakan untuk konteks setelahnya (Lample et al., 2016). Keluaran dari kombinasi dua arah *hidden layer* terdapat pada persamaan (2.7).

$$y_t = w_{\vec{h}_y} \vec{h}_t + w_{\leftarrow h_y} h_t^{\leftarrow} \quad (2.8)$$

Dimana y_t adalah *output gate* LSTM dua arah, $w_{\vec{h}_y}$ adalah nilai bobot *output gate* LSTM maju, \vec{h}_t adalah nilai keluaran LSTM maju, $w_{\leftarrow h_y}$ adalah nilai bobot untuk *output gate* LSTM mundur, h_t^{\leftarrow} adalah nilai keluaran LSTM mundur.

Pada Gambar 2.2 Arsitektur *Bidirectional LSTM* memperlihatkan arsitektur dari *Bidirectional LSTM*.



Gambar 2.2 Arsitektur *Bidirectional LSTM* (Graves et al., 2013).

2.2.5 Valence Aware Dictionary dan Sentiment Reasoner (Vader)

Valence Aware Dictionary dan Sentiment Reasoner (Vader) adalah *library* berbasis *lexicon* yang memiliki keunggulan untuk mengakses sentimen pada tiap teks yang diberikan tanpa memerlukan label pada teks (U et al., 2018). Vader berbeda dengan pendekatan *machine learning*. Vader diperkenalkan oleh C.J Hutton dan Eric pada tahun 2014. Vader metode pembentukannya didasarkan pada pendekatan *human-centric*, menggabungkan analisis kualitatif dan validasi empiris menggunakan kebijaksanaan dan penilaian manusia (Hutto, C.J. and Gilbert, 2014).

Kamus *lexicon* merupakan daftar leksikal kemampuan seperti kata yang dapat dikategorikan secara umum dengan orientasi semantik mereka sebagai positif, negatif atau netral. Pendekatan berbasis leksikon merupakan salah satu metode pada sentiment analisis yang memanfaatkan kamus yang berisi daftar kata yang mengandung opini (Nafan &

Amalia, 2019). dimana setiap kata pada kamus tersebut telah memiliki skor polaritas yang diberi nilai dari -1 (untuk kelas negatif) sampai dengan +1 (untuk kelas positif). Pendekatan berbasis *lexicon* dapat melatih model tanpa perlu menggunakan data berlabel.

Vader adalah salah satu metode *lexicon*. Keuntungan dari menggunakan *vader polarity detection* adalah sudah tersedia kamus yang berisi nilai dari setiap kata. Proses penentu polaritas kalimat didapatkan dari penyatuan attribute “*compound*” dari setiap kata yang tersedia (Ghiassi & Lee, 2018). Perhitungan sentimen dikelompokkan menjadi empat kelas yaitu positif, negatif, netral dan *compound score*. *Compound score* adalah jumlah keseluruhan skor positif, negatif dan netral yang telah dinormalisasi antara -1 dan +1. Jika nilai *compound score* ≥ 0.05 maka sentiment positif, yang diwakilkan dengan angka 1. Nilai *compound score* > -0.05 dan nilai *compound score* < 0.05 maka sentimen netral, jika nilai *compound score* ≤ -0.05 maka sentimen negative yang diwakilkan dengan angka -1 (Karim & Das, 2018).

2.2.6 Analisis Sentimen

Analisis sentimen merupakan sebuah studi komputasi tentang opini, evaluasi, sikap, subjektifitas dan pandangan yang diungkapkan dalam teks (Chaturvedi et al., 2017). Analisis sentimen menjadi populer dalam proses data sosial seperti blog, Twitter dan media kolaboratif online lainnya (Joshi et al., 2017).

Analisis sentimen adalah bidang studi yang menganalisis pendapat orang, opini, evaluasi, penilaian, sikap, dan emosi terhadap entitas (Hussein, 2018). Objek analisis sentimen berupa produk atau layanan yang ulasannya telah dipublikasikan di internet. Analisis sentimen adalah bidang yang menarik untuk dikembangkan dalam dunia digital sebab pada saat ini masyarakat secara umum mengutarakan perasaan melalui dunia maya. Menganalisis sentimen masyarakat dapat menghasilkan informasi-informasi baru yang dapat diolah sehingga menghasilkan pengetahuan baru.

2.2.7 Twitter API

Twitter merupakan layanan microblogging yang dioperasikan oleh Twitter, inc. Twitter memungkinkan penggunaannya mengirimkan dan membaca pesan *tweet* berupa teks, gambar atau sebuah video. Media sosial selain Twitter tidak memiliki Batasan dalam dalam penulisan karakter sedangkan Twitter memberikan 280 karakter yang dapat ditulis sebagai status atau cuitan. *Tweets* adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. *Tweets* dapat dilihat secara publik, namun pengguna Twitter dapat

membatasi pengiriman *tweet* hanya bisa dilihat temannya atau *follower*. *Application Programming Interface* (API) adalah seperangkat aturan dan spesifikasi tertentu yang dapat diikuti oleh program perangkat lunak untuk berkomunikasi satu sama lain (Gu et al., 2017). Fungsi yang dibuat dengan menggunakan API tersebut kemudian akan memanggil system calls sesuai dengan sistem operasinya (Trupthi et al., 2017).

Pada awalnya perusahaan Summize yang menyediakan fasilitas mencari data di Twitter. Perusahaan Summize ini diakuisisi dan diganti merek menjadi *twitter Search* sehingga *search* API terpisah sebagai entitas sendiri. API Twitter terdiri dari 2 (dua) bagian (Campan et al., 2019).

1. Search API

Search API dirancang untuk memudahkan *user* dalam mengelola *query search* di konten Twitter. *User* dapat menggunakannya untuk mencari *tweet* berdasarkan *keyword* khusus atau mencari *tweet* lebih spesifik berdasarkan *username* Twitter. *Search API* juga menyediakan akses pada data *trending topic*.

2. Streaming API

Streaming API digunakan *developer* untuk kebutuhan yang lebih intensif seperti melakukan penelitian dan analisis data. *Streaming API* dapat menghasilkan aplikasi yang dapat mengetahui statistik status *update*, *follower* dan lain sebagainya.

2.2.8 Text Mining

Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction* dan *information retrieval* (Berry & Kogan, 2010). *Text mining* menggabungkan teknik data *mining*, *machine learning*, *natural language processing*, informasi dan manajemen informasi (Yang et al., 2018).

Text mining berfungsi untuk mengekstrak informasi dari text tidak terstruktur sehingga dapat mentransfer pengetahuan antar domain dan umumnya digunakan dalam intelijen bisnis dalam mendukung pengambilan keputusan (Westergaard et al., 2018). Pada dasarnya proses kerja *text mining* mengadopsi dari *data mining*. Perbedaan *text mining* dan *data mining* adalah pola yang digunakan oleh *text mining* diambil dari sekumpulan bahasa alami yang tidak terstruktur. *Data mining* pola yang diambil dari *database* yang terstruktur. Tahapan-tahapan pada *text mining* adalah:

1. Text Preprocessing

Preprocessing data adalah langkah-langkah yang digunakan untuk persiapan data dan transformasi data yang berguna untuk *output* lebih baik dan efisien (Alasadi & Bhaya,

2017). *Preprocessing* data adalah langkah-langkah yang digunakan untuk persiapan data dan transformasi data yang berguna untuk *output* lebih baik dan efisien (Jamshed et al., 2019). Proses *preprocessing* memiliki beberapa tahapan, yaitu:

a. *Case Folding*.

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter. Fungsi *lower()* digunakan dalam tahap *case folding*.

b. *Tokenisasi*.

Tokenisasi adalah memecah kalimat menjadi kumpulan kata. *Tokenizing* memotong string *input* berdasarkan tiap kata yang menyusunnya. *Library* NLTK (*Natural Language Toolkit*) digunakan dalam proses *tokenisasi*.

c. *Normalization*.

Spelling normalization (normalisasi) merupakan proses identifikasi kata silang dan penulisan kata berlebihan kemudian diganti dengan kata kamus KBBI (pusat bahasa departemen pendidikan nasional, 2008).

d. *Filtering Stopword Removal*.

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil *token*. *Filtering* dapat menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *Library* python sastrawi dan NLTK (*Natural Language Toolkit*) digunakan pada tahap *stopword removal*.

e. *Lemmatization*.

Lemmatization adalah proses menghilangkan infleksi kata ke bentuk dasarnya. Bentuk dasar tersebut tidak berarti sama dengan akar kata (*root word*). Misalnya kata "mendengarkan", "dengarkan", "didengarkan" akan ditransformasi menjadi kata "dengar".

BAB 3

Metodologi

Bab ini menguraikan penelitian yang digunakan dalam memperoleh hasil temuan. Secara umum, tahapan penelitian yang dilakukan meliputi pengumpulan data, *pre-processing*, perancangan model, implementasi dan evaluasi model.

3.1 Data

Data dalam penelitian ini adalah sumber informasi yang akan diseleksi sebagai bahan analisis.

3.1.1 Lokasi dan Waktu Penelitian

Objek penelitian dalam penelitian ini adalah prediksi harga saham dan analisis sentimen. Penelitian ini menggunakan saham Bank Central Asia (BBCA) yang merupakan perusahaan terdaftar LQ45. Berdasarkan data yang diolah infobank saham-saham perbankan indonesia yang menduduki peringkat lima besar salah satunya adalah emiten adalah Bank Central Asia dengan kode saham BBCA (Handayani & Prastyo, 2020). LQ45 merupakan indeks saham yang berisi 45 perusahaan besar dengan prospek bisnis yang baik. Adapun periode penelitian ini adalah selama satu tahun.

Penelitian ini menggunakan dua data sekunder, yaitu data saham dari *Yahoo Finance* dan data sentimen dari Twitter. Data saham diperoleh melalui situs Yahoo Finance mulai dari Februari 2020 hingga September 2021. Data sentimen diambil dari Twitter mulai dari Februari 2020 sampai September 2021. Modul *yahoo_historical* digunakan untuk mengambil data saham dan beberapa modul python yang dapat digunakan untuk *crawling* data Twitter seperti *Twitter api*, *library twitterscraper* dan *libray getoldtweets3* digunakan untuk mengambil data sentimen.

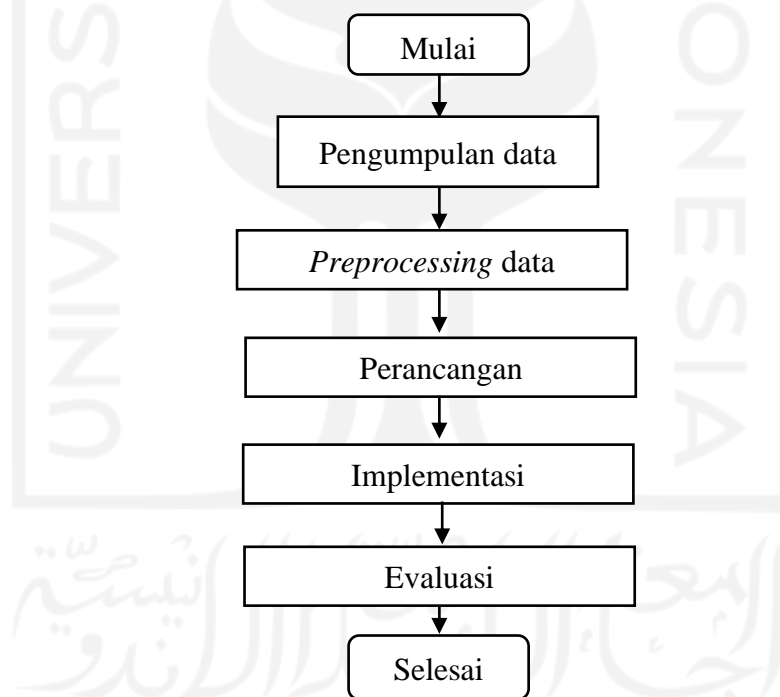
3.1.2 Populasi Penelitian

Populasi yang digunakan dalam penelitian ini adalah perusahaan perbankan bank BCA yang terdaftar sebagai anggota LQ45 selama periode Februari 2020 sampai September 2021. Data *close history* saham didapatkan dari Yahoo Finance sedangkan sentimen analisis diambil dalam Twitter. Data Twitter yang diambil adalah kalimat yang mengandung Ticker (\$) saham perusahaan bank BCA. Data *tweet* diambil setiap hari. Perhitungan sentimen menggunakan algoritma vader. Dalam 1 hari terdapat setidaknya 2 sampai 5 *tweet* yang membahas saham. *Tweet* kemudian dilakukan *preprocessing* data sebelum masuk perhitungan *score sentimen*. Perhitung *score sentimen* menggunakan algoritma vader.

Sentimen positif dalam satu hari lebih banyak dari sentimen negatif, maka sentimen pada hari tersebut bernilai positif. Sentimen negatif dalam satu hari lebih banyak dari sentimen positif, maka sentimen pada hari tersebut bernilai negatif. Kode saham ticker (\$) diberikan oleh Bursa efek Indonesia kepada perusahaan untuk keperluan perdagangan saham di Bursa (Ramadani, 2020). Kode saham (\$) akan dikenal investor secara luas dalam melakukan transaksi saham perusahaan di Bursa Efek Indonesia.

3.2 Langkah – Langkah Penelitian

Secara umum, metodologi penelitian merupakan acuan, pedoman dan tahapan yang akan dilakukan pada sebuah penelitian untuk mencapai tujuan dari penelitian tersebut. Metodologi yang diusulkan untuk penelitian prediksi saham berupa studi literatur, analisis dan pengumpulan data, *pre-processing* data, perancangan, pengujian, interpretasi hasil, serta pelaporan. Gambar 3.1 menunjukkan tahap aliran penelitian.



Gambar 3.1 Tahapan Penelitian

3.2.1 Pengumpulan Data

Pada tahap pengumpulan data, ditentukan jenis data apa yang digunakan untuk penelitian. Tahap pengumpulan data merupakan tahap yang penting dalam suatu penelitian. Data yang digunakan untuk penelitian harus benar-benar akurat dan jelas sumbernya. Penelitian ini menggunakan data saham perusahaan Bank BCA (BBCA). Data saham diambil dalam situs

Yahoo Finance sedangkan sentimen publik diambil dari Twitter. Data dikumpulkan selama 1 tahun 8 bulan mulai dari Februari 2020 sampai dengan September 2021. Modul yang digunakan untuk mengambil data Twitter, antara lain Twitter api, library twitterscraper dan library GetOldTweets3.

1 Studi Literatur

Pada tahapan ini, penulis mengumpulkan berbagai literatur terkait hasil penelitian sebelumnya. Peneliti mengumpulkan jurnal, buku, dan *library machine learning* terkait dengan penelitian. Kesimpulan dari penelitian-penelitian terdahulu yang relevan dijadikan dasar dari penulisan saat ini.

3.2.2 Pre-processing Data

Preprocessing data adalah langkah-langkah yang digunakan untuk persiapan data dan transformasi data. Proses *preprocessing* berguna untuk *output* lebih baik dan efisien (Alasadi & Bhaya, 2017). *Preprocessing* data adalah langkah-langkah yang digunakan untuk persiapan data dan transformasi data yang berguna untuk *output* lebih baik dan efisien (Jamshed et al., 2019). Proses tahapan dalam *preprocessing*, antara lain:

1 *Preprocessing* data *history* saham.

Pre-processing data *history* saham bertujuan agar data tersebut sesuai dengan yang dibutuhkan dalam penelitian ini. Langkah-langkah dalam *pre-processing* data *history* saham adalah:

a Menghilangkan *missing value*.

Missing value dapat ditangani dengan cara mengabaikan nilai *missing value*, mengisi nilai dengan manual, menggunakan konstanta global, dan menggunakan nilai rata-rata / median (Suyanto et al., 2019).

b Menggunakan kemungkinan nilai untuk mengisi *missing value*.

Nilai untuk mengisi *missing value* dapat ditentukan dengan menggunakan regresi, alat berbasis inferensi menggunakan *bayes formalism*, atau *decision tree*. Namun, pada penelitian ini, cara yang digunakan untuk mengatasi *missing value* yaitu dengan menghilangkan data yang memiliki *missing value* yang terdapat pada *data history* saham.

c Normalisasi dengan metode min-max.

Normalisasi data dilakukan dalam rentang yang kecil seperti [0,1] atau [-1, 1] sehingga semua atribut mempunyai bobot yang sama. *Min-max* merupakan metode

normalisasi dengan melakukan transformasi linear terhadap data asli (Suyanto et al., 2019).

Metode *min-max* menggunakan nilai minimum dan maksimum untuk melakukan konversi data secara *linear*. Misalkan, A adalah atribut bertipe numerik, maka \min_A adalah nilai minimum dalam atribut A dan, \max_A adalah nilai maksimum dalam atribut A. Suatu nilai x_i dapat dinormalisasi menjadi nilai baru X_i^1 yang berada dalam rentang $[\min_{\text{baru } A}, \max_{\text{baru } A}]$ sebagaimana disajikan dalam persamaan 3.1.

$$x_i^1 = \frac{X_i - \min_A}{\max_A - \min_A} \quad (3.1)$$

Di mana x_i^1 adalah data atribut yang akan dinormalisasi, X_i adalah data yang belum dinormalisasi, \min_A adalah nilai terkecil pada suatu atribut, dan \max_A adalah nilai terbesar pada suatu atribut.

Metode *min-max* telah banyak digunakan secara praktis. Dengan metode *min-max* dapat dilakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses.

d Split Dataset

Split dataset adalah pembagian komposisi data menjadi dua, yaitu data *training* dan data *testing*. Proporsi data *training* dan data *testing* masing-masing sesuai dengan keempat skenario.

2 Pre-processing data twitter

Pre-processing dilakukan supaya data sesuai dengan yang dibutuhkan dalam penelitian ini. Pada tahap ini, dilakukan *preprocessing* data twitter meliputi:

a Case Folding

Case folding adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Huruf 'a' sampai dengan 'z' dapat diterima, sedangkan karakter selain huruf dihilangkan dan dianggap *delimiter*. Fungsi *lower()* digunakan dalam tahap *case folding*. Aturan – aturan *case folding* dapat dilihat pada Tabel 3.1.

Tabel 3.1 Aturan *Case Folding*

No	Sebelum	Sesudah
1.	<i>Input</i> data latih memiliki huruf kapital [A...Z].	Maka akan mengubah semua inputan tersebut menjadi huruf kecil [a...z] semua.
2.	<i>Input</i> data latih memiliki karakter <i>symbol</i> .	Maka akan menghapus karakter simbol tersebut dari inputan.
3.	<i>Input</i> data latih memiliki huruf kecil.	Tidak ada aksi.
4.	<i>Input</i> data latih memiliki spasi.	Tidak ada aksi.

Pada tahap ini tidak menggunakan *external library* apapun, peneliti memanfaatkan modul yang tersedia di python. Contoh tahapan *case folding* diilustrasikan pada Tabel 3.2.

Tabel 3.2 Ilustrasi Tahapan *Case Folding*

No	Sebelum <i>case folding</i>	Setelah <i>case folding</i>
1.	#CariAlamat -> Alamat Bank Central Asia (BCA) Jl Hayam (∇)/ Wuruk Jakarta https://t.co/SYFZ6GK4z4	cari Alamat Bank Central Asia Jalan Hayam Wuruk Jakarta

b Tokenisasi

Tokenisasi yaitu proses memecah kalimat menjadi kumpulan kata. *Tokenizing* memotong *string input* berdasarkan tiap kata yang menyusunnya. *Library NLTK (Natural Language Toolkit)* digunakan dalam proses *tokenisasi*. Beberapa aturan proses agar hasil sesuai dengan yang diinginkan disajikan pada Tabel 3.3.

Tabel 3.3 Aturan Tahapan Tokenisasi

No	Sebelum	Sesudah
1.	Jika <i>input</i> data uji bertemu spasi.	Maka akan memecah dari deskripsi data latih menjadi bab-bab per bagian kata atau <i>string</i> .
2.	Jika <i>input</i> data latih memiliki huruf.	Tidak ada aksi.

Dalam hal ini, peneliti menghilangkan semua tanda baca dan segala sesuatu yang bukan huruf. Contoh tahapan *tokenizing* disajikan pada Tabel 3.4.

Tabel 3.4 Contoh Tokenisasi

No	Sebelum <i>tokenizing</i>	Setelah <i>tokenizing</i>
1.	#CariAlamat-> Alamat Bank Central Asia (BCA) Jl Hayam \(\nabla\)/ Wuruk Jakarta https://t.co/SYFZ6GK4z4	cari Alamat Bank Central Asia Jalan Hayam Wuruk Jakarta

c Normalization (*Stemming* dan *Lemmatization*)

Spelling normalization (normalisasi) merupakan proses identifikasi kata silang dan penulisan kata berlebihan yang kemudian diganti dengan kata baku berdasarkan Kamus Besar Bahasa Indonesia KBBI (Pusat Bahasa Departemen Pendidikan Nasional, 2008). *Lemmatization* adalah proses menghilangkan infleksi kata ke bentuk dasarnya. Bentuk dasar kata tersebut tidak berarti sama dengan akar kata (root word). Misalnya kata “mendengarkan”, “dengarkan”, “didengarkan” akan ditransformasi menjadi kata “dengar”. Sebagian kata bias pada kamus dasar bias tersaji pada Tabel 3.5.

Tabel 3.5 Kamus Kata Dasar Anomali

No	Kata bias	Kata dasar
1.	Jg	Juga
2.	Jd	Jadi
3.	Wkt	Waktu
4.	Brp	Berapa
5.	Jml	Jumlah

Dalam hal ini, peneliti melakukan proses pengambilan berbagai bentuk kata kedalam suatu representasi yang sama. Contoh tahapan *stemming* tersaji pada Tabel 3.6.

Tabel 3.6 Ilustrasi *Stemming*

No	Sebelum <i>normalization</i>	Setelah <i>normalization</i>
1.	BCA merupakan bank swasta yg jg melayani beragam segmen nasabah	bca Merupakan Bank Swasta yang Juga melayani beragam segmen nasabah

d Filtering.

Tahap *filtering* adalah tahap pengambilan kata-kata penting dari hasil tokenisasi. Proses *filtering* dapat menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *Library python* sastrawi dan *Natural Language Toolkit* (NLTK) digunakan pada tahap *stopword removal*. Beberapa proses agar hasil *stopword* sesuai yang diinginkan dapat dilihat pada Tabel 3.7.

Tabel 3.7 Aturan Tahap *Filtering*

No	Kondisi	Aksi
1.	Jika <i>input</i> data latih mengandung kata pada <i>database stopwords</i>	Maka akan menghapus kata atau string dalam data latih.
2.	Jika <i>input</i> data latih tidak mengandung kata pada kamus <i>stopword</i>	Maka tidak akan dihapus kata atau string data latih.

Penghilangan *stopword* dapat mengurangi ukuran *index* dan waktu pemrosesan. Contoh tahapan *filtering* penulis sajikan pada Tabel 3.8.

Tabel 3.8 Contoh *Filtering*

No	Sebelum <i>filtering</i>	Setelah <i>filtering</i>
1.	bagaimanapun juga dukung Jokowi Jadi Presiden Agar Indonesia Hebat	dukung Jokowi presiden Indonesia hebat

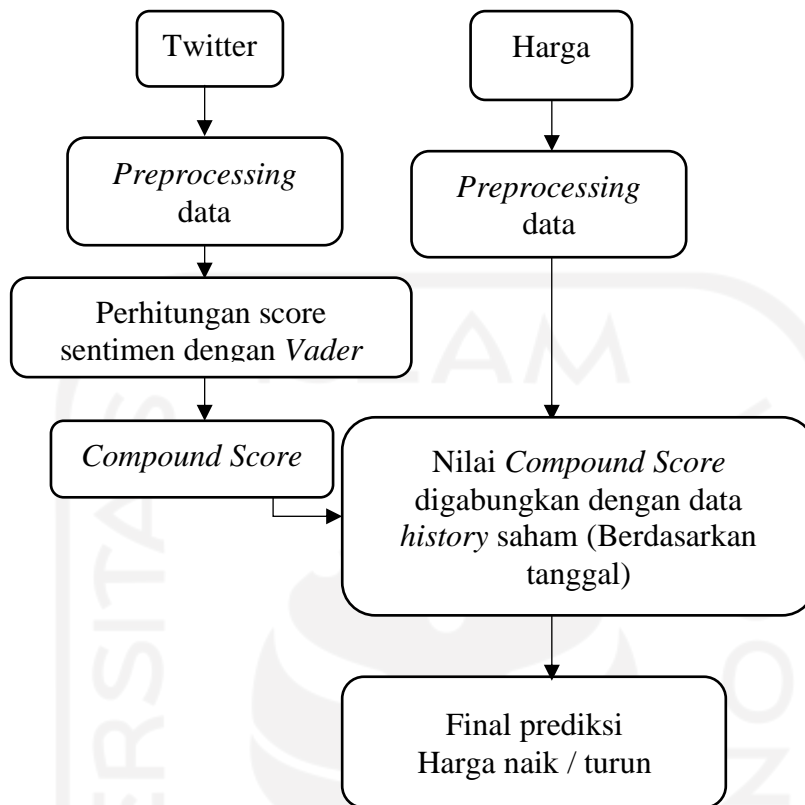
3.2.3 Perancangan

Perancangan merupakan penentuan proses dan data yang diperlukan dalam penelitian.

1. Metode prediksi saham yang diusulkan.

Model yang diusulkan diharapkan dapat membantu para investor untuk mengetahui risiko dan membantu investor dalam membuat keputusan dalam berinvestasi. Model yang dibangun, kemudian diujikan dari sisi keakuratan dan kegunaannya. Pada tahap ini, ditambahkan sentimen investor dalam proses *training* dan *testing*. Sentimen investor dimasukkan untuk melihat seberapa jauh mempengaruhi prediksi harga saham. Pada proses pelabellan data dilakukan *translate* bahasa dari Indonesia kedalam Inggris. Vader berkinerja baik dalam mengatasi text yang bercampur emoticon, akronim, dan bahasa gaul di sosial media (B. Caluza, 2017). Vader adalah campuran dari pendekatan kualitatif dan kuantitatif untuk menghasilkan dan memvalidasi leksikon sentimen. *Output* dari akhir prediksi adalah harga saham naik ataupun turun berdasarkan harga hari

sebelumnya. Langkah-langkah pada model yang diusulkan dapat dilihat pada Gambar 3.2.



Gambar 3.2 Model yang diusulkan

2. Komposisi *dataset*

Dalam langkah ini, dilakukan komposisi data *training* dan *testing* untuk melihat seberapa baik komposisi yang didapatkan. Komposisi data *training* dan *testing* masing masing sesuai dengan keempat skenario.

3. Skenario Pelatihan

Deep Learning membutuhkan data sebagai pelatihan (*training*) sebelum mengeluarkan *output*. Pelatihan Skenario pelatihan dilakukan sebanyak empat kali. Keempat skenario digunakan untuk dikombinasikan pada *hyperparameter tuning* untuk menemukan model terbaik. Rentang tanggal pelatihan dapat dilihat pada Tabel 3.9.

Tabel 3.9 Rentang waktu pelatihan

No.	Rentang waktu pelatihan
1.	05-07-2021 sampai dengan 16-08-2021
2.	05-04-2021 sampai dengan 16-08-2021
3.	19-11-2020 sampai dengan 16-08-2021
4.	19-02-2020 sampai dengan 16-08-2021.

Pada Tabel 3.9 Skenario pertama dilaksanakan dengan menggunakan data pada rentang waktu dari 05-07-2021 sampai dengan 16-08-2021. Skenario kedua dilakukan dengan menggunakan data pada rentang waktu 05-04-2021 sampai dengan 16-08-2021. Skenario ketiga dilakukan dengan menggunakan data pada rentang waktu dari 19-11-2020 sampai dengan 16-08-2021. Skenario keempat dilakukan menggunakan data pada rentang waktu dari 19-02-2020 sampai dengan 16-08-2021.

Keempat skenario tersebut memiliki data tes yang sama, yaitu data pada rentang waktu dari 18-08-2021 hingga 29-09-2021, dengan kurang lebih berjumlah 30 data. Dalam tahap ini juga dilakukan parameter jumlah *hidden state*, *learning rate* dan *max epoch* yang dilaksanakan dengan menganalisis parameter komposisi dataset.

Harga saham dilakukan tahap *pre-processing* sebelum menggunakan algoritma Bidirectional LSTM. Penelitian lain membuat matriks gabungan yang berisi skor sentiment dan perubahan harga saham (Kordonis et al., 2016). Selain itu, *output* dari data saham dan sentimen adalah vektor yang yang berisi Twitter positif atau negatif serta data harga saham (Khedr et al., 2017). Vektor berisi sentimen dan stock saham digabungkan berdasarkan tanggal yang sesuai (Batra & Daudpota, 2018). Skor sentimen positif dan skor sentimen negatif dapat digunakan untuk menganalisis perubahan tren harga saham (Wang et al., 2019). Data Twitter dan data history saham digabungkan berdasarkan tanggal. Pada python, digunakan `merge = df1.merge(df2, how='inner', on = 'Date', left_index=True)`. Code penggabungan data dapat dilihat pada Gambar 3.3.

```

1 #merge = saham.merge(hasil_twitter, how='inner', on='Date', left_index=True)
2 import pandas as pd
3 hasil_akhir = pd.merge(saham, hasil_twitter, how='inner', left_index=True, right_index=True)
4 hasil_akhir.head(20)

```

Gambar 3.3 Penggabungan data *history* saham dan data Twitter

Berdasarkan Gambar 3.3, didapatkan keterangan bahwa pada baris 1 mengimport *library pandas* untuk *dataframe*. Baris ke 3 menggabungkan data saham dan data Twitter berdasarkan tanggal yang sama dengan menggunakan *inner*. Baris 4 menampilkan data penggabungan sebanyak 20 data.

Polarisasi, subjektivitas dan *compound score* dilakukan untuk pelabelan data. Polarisasi digunakan untuk melihat seberapa positif atau negatif suatu kalimat atau kata. Sedangkan *compound score* adalah jumlah keseluruhan *lexicon* yang telah dinormalisasi antara -1 dan +1. Hasilnya adalah data history saham (*close*) dan *sentiment* Twitter (*compound score*) digunakan untuk *input* data *training Bidirectional LSTM* yang dapat dilihat pada Gambar 3.4.

```

1 #Labeling Menggunakan algoritma vader
2 import nltk
3 from nltk.sentiment.vader import SentimentIntensityAnalyzer
4 sid = SentimentIntensityAnalyzer()
5
6 hasil_akhir['scores'] = hasil_akhir['Text'].apply(lambda Text: sid.polarity_scores(Text))
7 hasil_akhir['neg'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['neg'])
8 hasil_akhir['neu'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['neu'])
9 hasil_akhir['pos'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['pos'])
10 hasil_akhir['compound'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['compound'])
11
12 hasil_akhir.head(20)

```

	Close	Username	Text	Date	scores	neg	neu	pos	compound
0	6615	0	Hate fraud	2020-11-19	{'neg': 1.0, 'neu': 0.0, 'pos': 0.0, 'compound...'	1.000	0.000	0.000	-0.8176
1	6600	ogiffology ; kanankiri_	The picture is very ugly	2020-11-20	{'neg': 0.473, 'neu': 0.527, 'pos': 0.0, 'comp...'	0.473	0.527	0.000	-0.5563
2	6600	nonayukee ; AndreasMarbun_	The water is dirty	2020-11-23	{'neg': 0.492, 'neu': 0.508, 'pos': 0.0, 'comp...'	0.492	0.508	0.000	-0.4404
3	6565	wijaryarputu ; medicrickey ; fitulted ; Rese...	the soil is less clean infertile	2020-11-24	{'neg': 0.0, 'neu': 0.675, 'pos': 0.325, 'comp...'	0.000	0.675	0.325	0.3415
4	6410	SwingTradeBot ; febianardr ; vian_sands	The forest caught fire, causing air pollution	2020-11-25	{'neg': 0.286, 'neu': 0.714, 'pos': 0.0, 'comp...'	0.286	0.714	0.000	-0.3400
5	6480	Khus_Indra	the room becomes dark because the electricity ...	2020-11-26	{'neg': 0.35, 'neu': 0.65, 'pos': 0.0, 'compou...'	0.350	0.650	0.000	-0.6486
6	6385	garamdanmicin ; Khus_Indra ; Truthspiller5	toilet clogged to make you dizzy	2020-11-27	{'neg': 0.275, 'neu': 0.725, 'pos': 0.0, 'comp...'	0.275	0.725	0.000	-0.2263
7	6205	itsmaritaaa ; barsnap	my head really dizzy so lazy	2020-11-30	{'neg': 0.59, 'neu': 0.41, 'pos': 0.0, 'compou...'	0.590	0.410	0.000	-0.6969

Gambar 3.4 Pemilihan *input* untuk *Bidirectional Long Short-Term Memory*

Dari Gambar 3.4, didapatkan hasil proses pelabellan dengan algoritma vader. Baris 2 adalah *import library* nltk yang digunakan untuk analisis teks. Baris 3 – 4 merepresentasikan memanggil kelas *SentimentIntensityAnalyzer* dari modul *vaderSentiment.vaderSentiment* dan dimasukkan kedalam objek *sid*. Baris 6 – 11 menggunakan *sid.polarity_scores* untuk menghasilkan polaritas sentimen negatif, netral, positif dan *compound*. Algoritma vader adalah algoritma yang menggunakan *lexicon* untuk mengklasifikasikan sentimen berdasarkan intensitas sentimen (Harish Rao M , Shashikumar D.R, 2017). Perhitungan sentimen dikelompokkan menjadi empat kelas yaitu positif, negatif netral dan *compound score*. *Compound score* adalah jumlah keseluruhan *lexicon* yang telah dinormalisasi antara -1 dan +1. Jika nilai *compound score*

bernilai ≥ 0.5 maka sentimen bernilai positif. Bila nilai *compound score* > -0.05 dan *compound score* < 0.05 , maka sentimen bernilai netral. Jika nilai *compound score* ≤ -0.05 maka sentimen bernilai negatif (Karim & Das, 2018). Python dapat menggunakan *library vaderSentiment* dan *SentimentIntensityAnalyzer* untuk mendapatkan skor polaritas dari sebuah kalimat. Berdasarkan Gambar 3.5, diketahui bahwa *compound score* sebesar -0.4767 sehingga kalimat tersebut bernilai negatif.

```

1 #Algoritma Vader
2 from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
3 analyzer = SentimentIntensityAnalyzer()
4 sentence = 'I am verry sad today'
5 vs = analyzer.polarity_scores(sentence)
6 print(vs)

```

{'neg': 0.437, 'neu': 0.563, 'pos': 0.0, 'compound': -0.4767}

Gambar 3.5 *Sentiment score* menggunakan VADER

Berdasarkan Gambar 3.5, didapatkan sentimen skor dengan menggunakan algoritma Vader. Baris 1 memanggil kelas *SentimentIntensityAnalyzer* dari modul *vaderSentiment.vaderSentiment* dan dimasukkan kedalam objek *analyzer*. Baris 4 adalah contoh kalimat. Baris 5 adalah objek yang berisi fungsi *polarity_score* berguna untuk menghasilkan skor sentimen negatif, netral, positif dan *compound*.

3.2.4 Implementasi

Implementasi adalah tahap penerapan dan sekaligus pengujian bagi sistem berdasarkan hasil analisis dan perancangan yang telah dilakukan. Implementasi dapat berguna untuk mengetahui tentang gambaran yang jelas mengenai penelitian yang dilakukan.

1. Analisis kebutuhan perangkat keras

Spesifikasi perangkat keras yang digunakan dalam penelitian ini, diantaranya yaitu:

- a. Processor : Intel(R) Core (TM) i5-8250U CPU @ 1.60GHz 1.80 GHz
- b. Memory RAM : 12.0 GB
- c. Harddisk : 1 TB
- d. System Type : 64-bit operating system, x64-based processor

2. Analisis kebutuhan perangkat lunak.

Spesifikasi perangkat lunak yang digunakan dalam penelitian ini, diantaranya yaitu:

- 1.2 Windows 10 Home Single Language.
- 2.2 Python 3.8.1

3.2 Numpy 1.19.5, Pandas 1.0.1, yahoo-historical 0.3.2, matplotlib 3.1.3, nltk 3.4.5,
4.2 Scikit-learn 0.24.2, Tensorflow 2.4.0.

3.1 Evaluasi

Tahap terakhir dari tahapan penelitian ini adalah evaluasi. Hasil prediksi yang telah didapatkan selanjutnya memasuki tahapan pengujian untuk memastikan keakurasiannya. Hasil yang didapatkan kemudian dievaluasi menggunakan *mean squared error* (MSE), *root mean squared error* (RMSE), dan *mean absolute percentage error* (MAPE). Matrix evaluasi MSE, RMSE dan MAPE digunakan karena hal tersebut yang paling umum digunakan dalam permasalahan prediksi harga saham menggunakan berbagai macam metode *deep learning* (Hu et al., 2021).

1. Mean squared error (MSE)

MSE merupakan rata-rata kesalahan kuadrat antara nilai data aktual dan peramalan. Pada perhitungan MSE dilakukan pengurangan antara nilai data aktual dengan data peramalan. Nilai tersebut kemudian dikuadratkan serta dijumlahkan secara keseluruhan dan dibagi dengan jumlah data yang ada. Rumus MSE dirumuskan pada persamaan 3.2.

$$MSE = \frac{\sum(Aktual - Prediksi)^2}{n} \quad (3.2)$$

Dimana MSE adalah *mean square error*, aktual adalah data sebenarnya, prediksi adalah nilai prediksi dari variabel aktual dan n banyaknya observasi.

2. Root Mean Squared Error (RMSE).

RMSE adalah perhitungan dengan mengkuadratkan *error* (data aktual - data prediksi). kemudian dibagi dengan jumlah data dan diakarkan. RMSE dirumuskan pada persamaan 3.3.

$$RMSE = \sqrt{\frac{\sum(Aktual - Prediski)^2}{n}} \quad (3.3)$$

Dimana RMSE adalah *root mean square error*, *aktual* adalah data sebenarnya, *prediksi* adalah nilai prediksi dari variabel *aktual* dan n banyaknya obsevasi.

3. Mean Absolute Percentage Error (MAPE)

MAPE dihitung sebagai perbedaan rata-rata yang absolut antara nilai yang diramalkan dengan aktualnya, dan dicerminkan sebagai persentase nilai aktual. Dalam hal ini, jika kita memiliki nilai yang diramalkan dan aktual untuk periode n , maka MAPE dapat dirumuskan dengan persamaan 3.4.

$$MAPE = \frac{1}{N} \sum_{i=1}^n \frac{y - \hat{y}}{y} \quad (3.4)$$

Dimana MAPE adalah *mean absolute percentage error*, N adalah banyaknya jumlah data, y adalah nilai aktual pada waktu ke- i , dan \hat{y} adalah nilai hasil prediksi pada waktu ke- i .



BAB 4

Hasil dan Pembahasan

4.1 Data Pre-processing

Pada penelitian ini, data saham dikumpulkan selama 1 tahun 7 bulan mulai dari bulan Februari 2020 sampai September 2021. Data sentimen diambil dari Twitter mulai dari bulan Februari 2020 hingga September 2021. *Pre-processing* data bertujuan untuk mentransformasi data agar sesuai dengan penelitian. *Pre-processing* data saham yaitu menghilangkan *missing value*. Beberapa proses yang dilakukan pada tahap *preprocessing* data Twitter, antara lain *case folding*, *tokenisasi (tokenization)*, *normalisasi (normalization)*, *filtering stopword removal*, dan *lemmalization*. Proses *pre-processing* menggunakan bahasa Python. Gambar 4.1 memperlihatkan hasil proses *pre-processing* data.

```
1 # Cleansing data
2 import string
3 import re
4
5 from nltk.tokenize import word_tokenize
6 from nltk.probability import FreqDist
7 from nltk.corpus import stopwords
8 from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
9
10 #membersihkan data
11 def cleaning(text):
12     # remove tab, new line, and back slice
13     text = text.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\ ', "")
14     # remove non ASCII (emoticon, chinese word, .etc)
15     text = text.encode('ascii', 'replace').decode('ascii')
16     # remove mention, link, hashtag
17     text = ' '.join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\/\/\w+)", " ", text).split())
18     # remove incomplete URL
19     return text.replace("http://", " ").replace("https://", " ")
20
21
22 def removeStopword(text):
23     stop_words = set(stopwords.words('stopwords_id'))
24     word_tokens = word_tokenize(text)
25     filtered_sentence = [w for w in word_tokens if not w in stop_words]
26     return ' '.join(filtered_sentence)
27
28 #
29 def stemming(text):
30     factory = StemmerFactory()
31     stemmer = factory.create_stemmer()
32     return stemmer.stem(text)
33
```

Gambar 4.1. Kode pre-processing data

Berdasarkan Gambar 4.1, didapatkan keterangan bahwa baris ke 1-8 merepresentasikan modul dan *library*. NLTK adalah sebuah *library* yang digunakan untuk membantu kita dalam bekerja dengan teks. Baris 5 merepresentasikan *word_tokenize* untuk melakukan tokenisasi dan membuang *stopwords* dari teks. Baris ke 7-8 merepresentasikan

proses pembuangan *stopwords*. *Stopwords* sendiri merupakan kata umum dalam sebuah teks yang sebetulnya tidak memiliki makna. *Library* sastrawi merupakan *stopword* dalam bahasa Indonesia. Baris ke 11-19 adalah fungsi yang berguna untuk *cleansing* data. Fungsi *cleaning* digunakan untuk menghapus *tab*, baris baru, *emoticon*, URL, *hashtag*, *link* dan bilangan non-ASCII. Baris ke 22-23 merupakan fungsi dari *removeStopword*. Baris ke 29-32 merepresentasikan fungsi dalam proses *stemming*. Tokenisasi merupakan proses untuk memecah isi dari teks menjadi kata-kata. *Case folding* yaitu proses mengubah semua huruf pada dokumen menjadi huruf kecil. Tokenisasi adalah proses membagi teks yang berupa kalimat menjadi token-token. Normalisasi merupakan proses menormalkan kata menjadi kata baku yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). *Filtering stopwords removal* adalah proses pengambilan kata penting dari hasil token dengan menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *Lemmalization* yakni proses untuk mengubah sebuah kata menjadi bentuk dasarnya.

Data Twitter yang telah melalui tahap *pre-processing* data selanjutnya di beri label. *Tools SentimentIntensityAnalyzer* digunakan untuk proses pelabelan data. *SentimentIntensityAnalyzer* adalah *library vader lexicon* yang berasal dari *package NLTK*. Data diproses dengan menerapkan algoritma *sentiment analysis Vader*. Vader berfungsi untuk menghitung skor dari tiap kalimat (ulasan). Skor yang dihasilkan yaitu positif, negatif, dan netral. Skor yang dihasilkan akan digabungkan dan menghasilkan nilai *compound score*. *Compound score* adalah matriks yang menghitung semua skor yang telah dinormalisasi antara -1 dan +1. Nilai *compound score* dibawah 0 (*compound score* < 0) bernilai negatif, sedangkan nilai *compound score* diatas 0 (*compound score* ≥ 0) bernilai positif.. Data yang sudah diberikan label kemudian divalidasi secara manual sehingga kemungkinan kecil terjadi kesalahan. Data yang sudah tervalidasi lalu dibagi menjadi data, yaitu latih dan data uji. Total data yang tervalidasi adalah sebanyak 390 data. Gambar 4.2 menunjukkan proses pelabelan data dengan menggunakan Vader.

```

1 #Labeliing Menggunakan algoritma vader
2 import nltk
3 from nltk.sentiment.vader import SentimentIntensityAnalyzer
4 sid = SentimentIntensityAnalyzer()
5
6 hasil_akhir['scores'] = hasil_akhir['Text'].apply(lambda Text: sid.polarity_scores(Text))
7
8 hasil_akhir['neg'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['neg'])
9 hasil_akhir['neu'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['neu'])
10 hasil_akhir['pos'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['pos'])
11 hasil_akhir['compound'] = hasil_akhir['Text'].apply(lambda x:sid.polarity_scores(x)['compound'])
12
13
14 #hasil_akhir['neg'] = hasil_akhir['Text'].apply(lambda Text: sid.polarity_scores(['neg']))
15 hasil_akhir.head()

```

Gambar 4.2. Kode pelabelan menggunakan Vader.

Berdasarkan Gambar 4.2, selain berfungsi untuk mengkategorikan kalimat menjadi positif dan negatif, kode pelabelan tersebut juga berfungsi untuk mendapatkan nilai *compound score*. Kode tersebut dapat digunakan sebagai fitur pada proses *training* karena memberikan nilai *compound score*. Data yang sudah diberikan label kemudian divalidasi agar sesuai dengan kebutuhan penelitian. Dalam proses validasi, penulis memeriksa hasil pelabelan secara manual. Data saham dan *sentiment analysis* kemudian digabungkan berdasarkan tanggal. Data tersebut selanjutnya dibagi menjadi dua, yaitu data latih dan data uji.

4.2 Pelatihan Model dan Evaluasi

Data yang sudah tervalidasi dibagi menjadi data latih dan data uji. Pada proses ini, data saham dan nilai *compound score* digabungkan berdasarkan tanggal. kedua data tersebut akan digunakan untuk proses pelatihan. Dalam penelitian ini, dibuat empat skenario. Skenario dibuat sejauh mana model dapat memprediksi harga saham. Skenario pertama memakai data *training* selama 1 bulan dan data *testing* selama 1 bulan. Skenario kedua memakai data *training* 3 bulan dan data *testing* 1 bulan. Skenario ketiga menggunakan data *training* 6 bulan dan data *testing* 1 bulan. Skenario keempat menggunakan data *training* 1 tahun dan data *testing* 1 bulan. *Script* yang digunakan untuk proses ini dapat dilihat pada Gambar 4.3.

```

5 #MinMaxScaler()
6 from sklearn.preprocessing import MinMaxScaler
7 tc = MinMaxScaler()
8 trainingset = tc.fit_transform(trainingset)
9 testset     = tc.fit_transform(testset)
10 print(trainingset)
11 print("\n")
12
13 tc1 = MinMaxScaler()
14 trainingset1 = tc1.fit_transform(trainingset1)
15 testset1     = tc1.fit_transform(testset1)
16 print(trainingset1)
17 print("\n")
18
19
20 def createDataset1(dataku, dataku2, window1):
21     dataX1, dataY1 = [], []
22     for i in range(len(dataku)-window1):
23         temp1 = []
24         for j in range(i, i+window1):
25             if j==i+window1-1:
26                 temp1.append(dataku2[j-(window1-1),0])
27             else:
28                 temp1.append(dataku[j,0])
29         dataX1.append(temp1)
30         dataY1.append(dataku[i+window1, 0])
31     return np.array(dataX1), np.array(dataY1)
32
33 window1 = 4
34 trainX1, trainY1 = createDataset1(trainingset, trainingset1, window1)
35 print((trainX1))
36 print('\n')
37 trainX1= trainX1.reshape(len(trainingset1)-window1, window1, 1)
38 print((trainX1))

```

Gambar 4.3 Kode normalisasi data, penggabungan data saham, dan nilai *compound score*

Gambar 4.3 merupakan proses normalisasi dan pembentukan data latih. Data dibagi menjadi data latih dan data uji. Baris ke 6-17 merupakan proses normalisasi data *MinMaxScaler()* yang dilakukan pada data *training* dan data *testing*. Metode *minmaxscaler* berfungsi untuk mengubah data pada rentang 0 sampai 1. Normalisasi bertujuan agar tidak rentang terhadap pencilan data. Baris ke 20-31 merupakan fungsi yang berguna untuk membentuk matriks *dataset* sebelum dimasukkan kedalam proses *training* dan *testing*. Baris ke 33-37 merepresentasikan data hasil penggabungan yang dimasukkan kedalam fungsi *createDataset1* dalam bentuk *array*. Potongan kode diatas menjelaskan bahwa data saham dan nilai *compound score* digabungkan dalam fungsi *createDataset1* sebelum dimasukkan ke dalam proses pelatihan. Data yang sudah siap dilatih selanjutnya dimasukkan kedalam proses pelatihan sebagaimana ditunjukkan pada Gambar 4.4.

```

1 # Model
2 regressor=keras.Sequential()
3 # hidden layer (4)
4 regressor.add(Bidirectional(LSTM(units=2000,activation='tanh',return_sequences=True,input_shape=(trainX1.sha
5 regressor.add(Dropout(0.2))
6
7 regressor.add(Bidirectional(LSTM(units=2000, return_sequences=True)))
8 regressor.add(Dropout(0.2))
9
10 regressor.add(Bidirectional(LSTM(units=2000, return_sequences=True)))
11 regressor.add(Dropout(0.2))
12
13 regressor.add(Bidirectional(LSTM(units=20)))
14 regressor.add(Dropout(0.2))
15 # Adding the output layer
16 regressor.add(Dense(units = 1))
17 regressor.compile(optimizer='adam',loss='mean_squared_error') #adam
18 #regressor.fit(trainX1,trainY1,epochs=500,batch_size=32)
19 history = regressor.fit(trainX1,trainY1,epochs=205,batch_size=32, validation_split=0.2)

```

Gambar 4.4 Potongan kode proses *training* data

Gambar 4.4 menunjukkan kombinasi parameter untuk setiap skenario pemodelan. Baris ke 4-14 merepresentasikan jumlah *hidden layer*. Penelitian ini menggunakan *hidden layer* sebanyak empat. *Hidden layer* adalah jumlah lapisan tersembunyi diantara *input layer* dan *output layer*. *Hidden layer* dapat diartikan sebagai lapisan antara *input layer* dan *output layer*. *Artificial neuron* memiliki sekumpulan *input* pembobot (*weight*) dan prosedur untuk menghasilkan *output neuron* melalui *activation function*.

Pada baris ke 4-14 terdapat jumlah *neuron* sebanyak 2000. *Neuron* menunjukkan jumlah *neuron* yang terdapat di dalam *hidden layer*. *Neuron* akan mentransformasikan informasi yang diterima melalui sambungan keluarnya menuju ke *neuron* lain. Hubungan antar *neuron* dapat disebut dengan bobot.

Jumlah *epoch* adalah banyaknya siklus penuh *forward* dan *backward* untuk semua data latih. Satu *epoch* dapat diartikan bahwa setiap sampel dalam *dataset training* memiliki kesempatan untuk memperbaharui parameter model internal. Satu *epoch* terdiri dari satu atau lebih *batch*.

Batch size merupakan banyaknya sampel data yang masuk pada setiap iterasi *batch* nya. Ukuran *batch* adalah *hyperparameter* yang menentukan jumlah sampel untuk dikerjakan sebelum memperbaharui parameter model internal. Jumlah *epoch* dan *batch size* yang digunakan dalam penelitian ini mengacu pada penelitian (A. Kumar et al., 2021). Masing masing model dilatih menggunakan optimisasi Adam. Gambar 4.5 menunjukkan pelatihan model. Pada gambar tersebut, model dilatih dengan menggunakan laptop.

```

To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
1/1 [=====] - 216s 216s/step - loss: 0.0462 - val_loss: 0.2071
Epoch 2/205
1/1 [=====] - 9s 9s/step - loss: 0.0158 - val_loss: 0.0965
Epoch 3/205
1/1 [=====] - 6s 6s/step - loss: 0.1961 - val_loss: 0.5928
Epoch 4/205
1/1 [=====] - 8s 8s/step - loss: 0.0648 - val_loss: 0.5232
Epoch 5/205
1/1 [=====] - 8s 8s/step - loss: 0.0519 - val_loss: 0.3792
Epoch 6/205
1/1 [=====] - 7s 7s/step - loss: 0.0316 - val_loss: 0.3057
Epoch 7/205
1/1 [=====] - 8s 8s/step - loss: 0.0243 - val_loss: 0.2565
Epoch 8/205
1/1 [=====] - 8s 8s/step - loss: 0.0192 - val_loss: 0.1980
Epoch 9/205
1/1 [=====] - 8s 8s/step - loss: 0.0160 - val_loss: 0.1335
Epoch 10/205
1/1 [=====] - 8s 8s/step - loss: 0.0150 - val_loss: 0.1450

```

Gambar 4.5 Tangkapan layar saat pelatihan model.

Hasil pelatihan model dapat dilihat Tabel 4.1 dan Tabel 4.2. Tabel 4.1 menjelaskan *loss* pada saat pelatihan dengan menggabungkan data sentimen. Penelitian ini menggunakan jumlah hidden layer 4, neuron 2000, epoch 205, dan batch size 32. Skenario keempat pada Tabel 4.1 memiliki nilai *loss* terendah yang diikuti oleh skenario ketiga, pertama, dan kedua, dengan nilai *loss* masing-masing sebesar 0,0033, 0,0118, 0,0141, dan 0,0187.

Tabel 4.1 Parameter Dan *Loss* Pemodelan Dengan Menggunakan Sentimen

No	Skenario	Hidden Layer	Neuron	Epoch	Batch size	Loss
1.	1	4	2000	205	32	0.0141
2.	2	4	2000	205	32	0.0187
3.	3	4	2000	205	32	0.0118
4.	4	4	2000	205	32	0.0033

Tabel 4.2 menjelaskan *loss* pada saat pelatihan tanpa menggabungkan data sentimen. Skenario keempat pada Tabel 4.2 memiliki nilai *loss* rendah diikuti dengan skenario ketiga, pertama, dan kedua, dengan nilai *loss* masing-masing sebesar 0,0037, 0,0129, 0,0159 dan 0,0162.

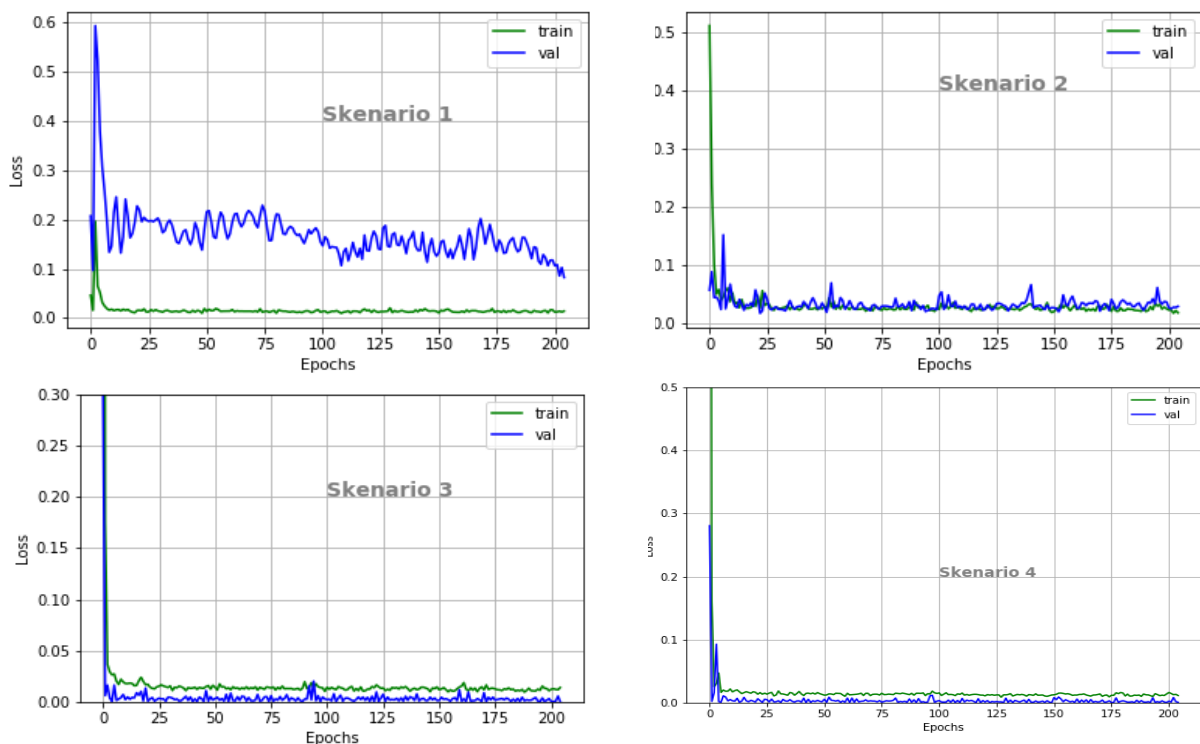
Tabel 4.2 Parameter Dan *Loss* Pemodelan Tanpa Menggunakan Sentimen

No	Skenario	Hidden Layer	Neuron	Epoch	Batch size	Loss
1.	1	4	2000	205	32	0.0159
2.	2	4	2000	205	32	0.0162
3.	3	4	2000	205	32	0.0129
4.	4	4	2000	205	32	0.0037

Gambar 4.6 menunjukkan grafik *loss* dalam proses pelatihan keempat skenario. Berdasarkan Gambar 4.6, dapat dilihat bahwa pada skenario pertama, dihasilkan nilai *loss* yang besar dengan jumlah *epoch* 205. Skenario kedua menghasilkan nilai *loss* yang rendah

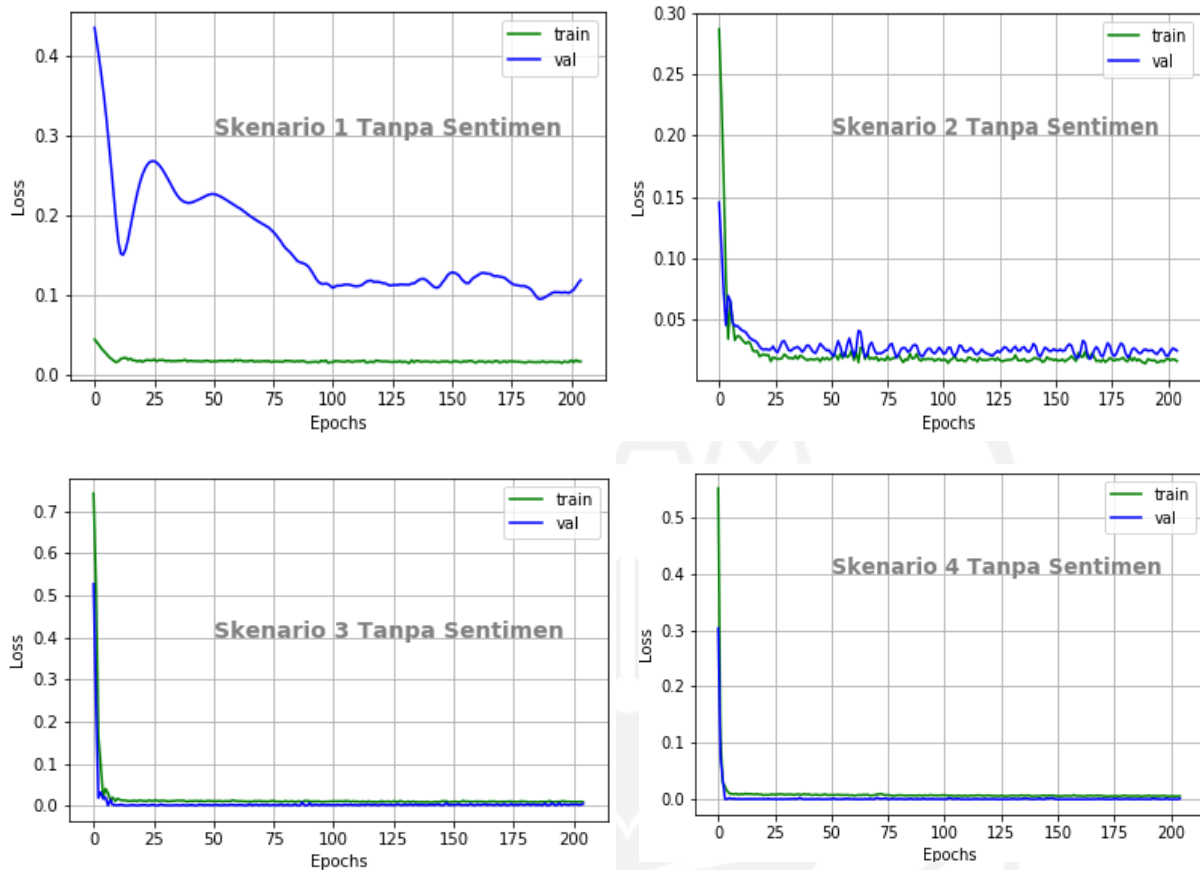
dengan jumlah *epoch* 205. Skenario ketiga dan keempat menghasilkan nilai *loss* yang lebih rendah dengan menggunakan jumlah *epoch* 205.

Gambar 4.6 menunjukkan grafik *loss* dalam proses pelatihan keempat skenario. Pada Gambar 4.6 dapat dilihat skenario pertama menghasilkan nilai *loss* yang besar dengan jumlah *epoch* 205. Skenario kedua menghasilkan nilai *loss* yang rendah dengan jumlah *epoch* 205. Skenario ketiga dan keempat menghasilkan nilai *loss* yang lebih rendah dengan menggunakan jumlah *epoch* 205.



Gambar 4.6. Nilai *loss* pada proses pelatihan

Gambar 4.7 menunjukkan grafik *loss* dalam proses pelatihan keempat skenario tanpa data sentimen. Berdasarkan Gambar 4.7, dapat dilihat skenario pertama, dihasilkan nilai *loss* yang besar dengan jumlah *epoch* 205. Skenario kedua menghasilkan nilai *loss* yang rendah dengan jumlah *epoch* 205. Skenario ketiga dan keempat menghasilkan nilai *loss* yang lebih rendah dengan jumlah *epoch* 205.



Gambar 4.7. Nilai *loss* pada proses pelatihan model tanpa data sentimen

4.3. Pengujian

Setelah melakukan proses *training*, tahap berikutnya yang telah dilakukan adalah pengujian. Tahap ini bertujuan untuk mengetahui model yang dikembangkan dapat memprediksi harga saham. Pengujian dalam pembuatan laporan ini menggunakan data mulai dari tanggal 18-09-2021 sampai dengan 29-09-2021. Tabel 4.3 menunjukkan hasil pengujian model dengan empat skenario.

Tabel 4.3. Hasil Pengujian Empat Skenario Model

No	Skenario	MSE	RMSE	Dstat	MAPE
1.	1	0.133	0.364	68	27.90
2.	2	0.123	0.351	64	27.57
3.	3	0.107	0.327	44	23.83
4.	4	0.094	0.306	55.55	23.85

Data yang digunakan adalah data uji. Data uji yang dipakai data dari tanggal 18-09-2021 sampai dengan 29-09-2021. Masing-masing skenario pelatihan diterapkan pada proses pengujian. Pada skenario pertama, data latih yang digunakan adalah dari tanggal 05-07-2021

sampai dengan tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,133. Nilai RMSE yang diperoleh sebesar 0,364. Nilai Dstat yang didapatkan sebesar 68.

Pada percobaan kedua, skenario menggunakan data mulai dari tanggal 05-07-2021 hingga tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,123. Nilai RMSE yang diperoleh sebesar 0,351. Nilai Dstat yang didapatkan sebesar 64.

Pada percobaan ketiga, skenario menggunakan data mulai dari tanggal 19-02-2020 hingga tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,107. Nilai RMSE yang diperoleh sebesar 0,327. Nilai Dstat didapatkan sebesar 44.

Pada percobaan keempat, skenario yang digunakan adalah data mulai dari tanggal 19-02-2020 hingga tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,094. Nilai RMSE yang didapatkan sebesar 0,306. Nilai Dstat yang didapatkan sebesar 55,55. Grafik yang disajikan dalam Gambar 4.3 merupakan gambar hasil prediksi. Garis biru merupakan garis nilai prediksi, sedangkan garis merah merupakan garis nilai sebenarnya.

Pada tahap pengujian peneliti menguji, telah dilakukan pengujian empat skenario tanpa menggunakan data sentimen. Tabel 4.4 memperlihatkan tabel hasil pengujian model dengan empat skenario tanpa data sentimen.

Tabel 4.4. Hasil Pengujian Empat Skenario Model Tanpa Sentimen

No	Skenario	MSE	RMSE	Dstat	MAPE
1.	1	0.148	0.385	65.38	31.66
2.	2	0.150	0.38	61.53	30.95
3.	3	0.151	0.389	57.69	30.57
4.	4	0.171	0.413	50.0	34.43

Pada skenario pertama, digunakan data latih dari tanggal 05-07-2021 sampai dengan tanggal 16-08-2021. Nilai MSE yang didapatkan yaitu sebesar 0,148. Nilai RMSE yang didapatkan sebesar 0,385. Nilai Dstat yang didapatkan sebesar 65,38 dan nilai MAPE sebesar 31,66.

Pada percobaan kedua, skenario menggunakan data dari tanggal 05-07-2021 sampai dengan tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,150. Nilai RMSE yang didapatkan yakni sebesar 0,380. Nilai Dstat yang didapatkan sebesar 61,53 dan nilai MAPE sebesar 30,95.

Pada percobaan ketiga, skenario menggunakan data mulai dari tanggal 19-02-2020 hingga tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,151. Nilai RMSE yang diperoleh sebesar 0,389. Nilai Dstat didapatkan sebesar 57,69 dan nilai MAPE sebesar 30,57.

Pada percobaan keempat, skenario yang digunakan adalah data mulai dari tanggal 19-02-2020 hingga tanggal 16-08-2021. Nilai MSE yang didapatkan sebesar 0,171. Nilai RMSE yang diperoleh sebesar 0,413. Nilai Dstat yang didapatkan sebesar 50,0 dan nilai MAPE sebesar 34,43.

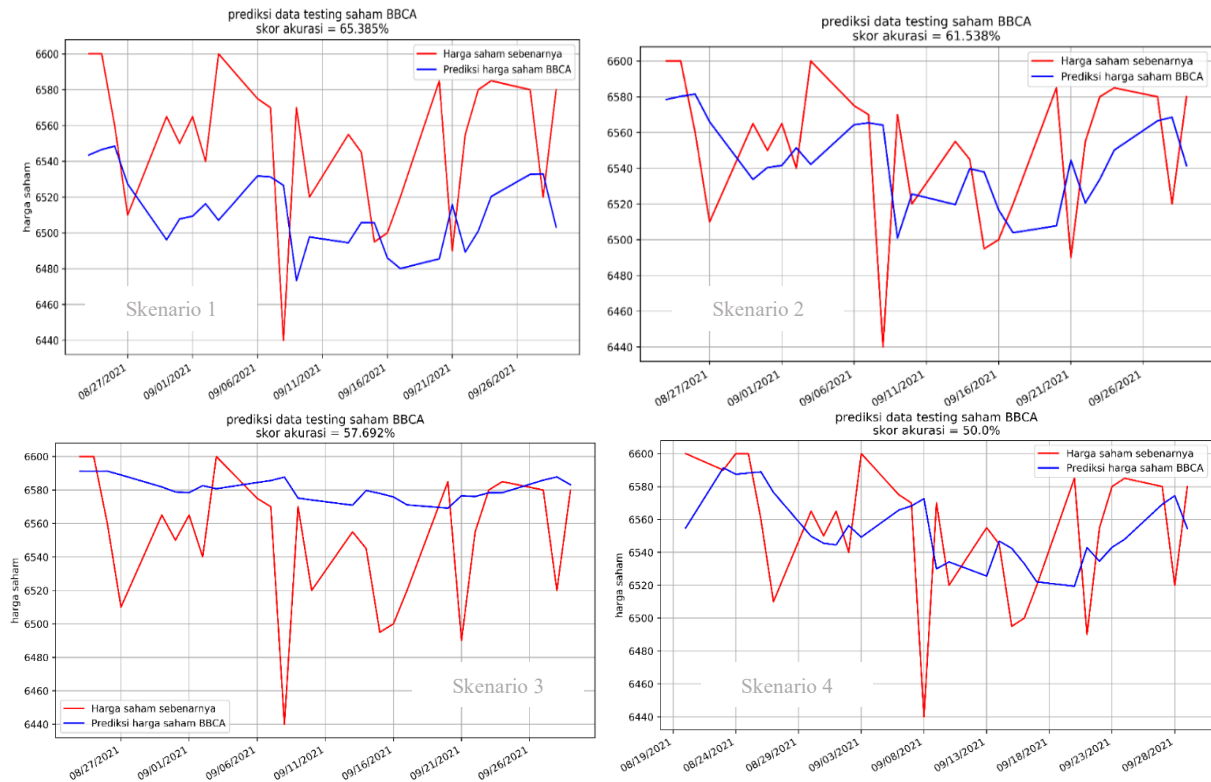
Grafik hasil prediksi harga saham pada keempat skenario model dapat dilihat pada Gambar 4.8. Garis berwarna biru merepresentasikan hasil prediksi dan garis berwarna merah merepresentasikan nilai saham riil. Dari gambar tersebut, terlihat bahwa pola prediksi dari harga saham yang paling menyerupai pola harga riil nya adalah model dari skenario 1 dan 2. Hal ini dapat menjadi justifikasi tingginya nilai *Dstat* yang didapatkan pada skenario 1 dan 2 dibandingkan dengan skenario yang lainnya.



Gambar 4.8 Gambar hasil prediksi

Dengan melihat hasil yang didapatkan pada Tabel 4.4 dan Gambar 4.8 serta nilai MSE, RMSE, dan MAPE yang tidak terlalu jauh dari model yang dibangun pada skenario kedua merupakan alternatif yang lebih baik dengan memberikan gambaran perubahan nilai yang merupakan nilai riil harga saham.

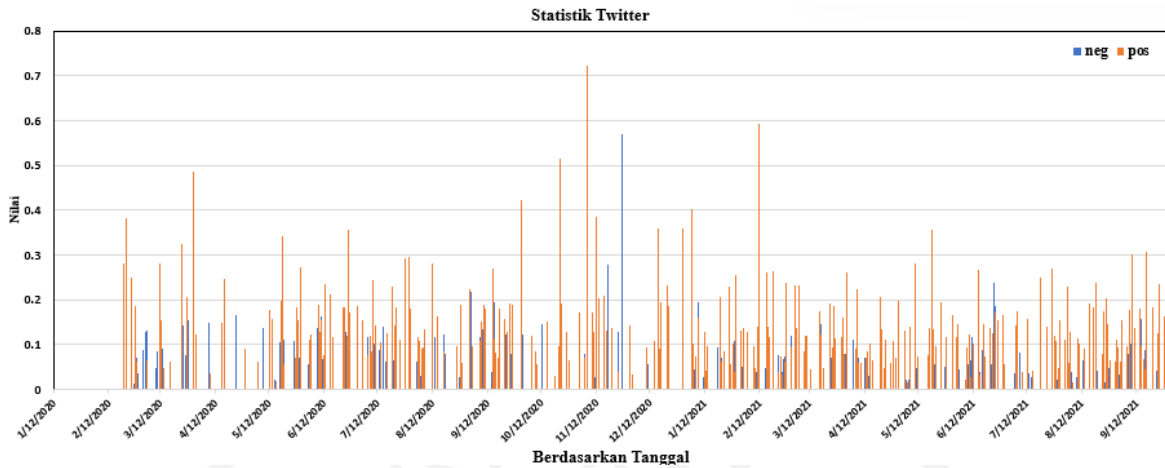
Gambar 4.9 adalah grafik hasil prediksi tanpa menggunakan data sentimen. Garis merah merepresentasikan hasil prediksi, sedangkan garis biru merepresentasikan harga saham sebenarnya. Pada skenario 1 dan 2 pola garis prediksi menyerupai pola harga saham sebenarnya.



Gambar 4.9 Gambar hasil prediksi tanpa data sentimen

Dengan melihat hasil yang didapatkan pada tabel 4.4 dan gambar 4.9 serta nilai MSE dan RMSE yang tidak terlalu jauh model yang dibangun pada skenario kedua merupakan alternatif yang lebih baik dengan memberikan gambaran perubahan nilai yang merupakan nilai riil harga saham.

Gambar 4.10 merupakan statistik sentimen positif dan negatif. Warna oranye menandakan sentimen positif. Warna biru menandakan sentimen negatif. Sentimen positif memiliki nilai *mean* sebesar 0,031, nilai *median* sebesar 0,0, nilai *variance data* sebesar 0,0029 dan nilai standar deviasi sebesar 0,00296. Sentimen negatif memiliki nilai *mean* sebesar 0,1006, nilai *median* sebesar 0,088, nilai *variance data* sebesar 0,0029 dan standar deviasi sebesar 0,0112.



Gambar 4.10 Statistik *tweet*

4.3.1 Evaluasi Kualitatif

Evaluasi kualitatif adalah evaluasi terhadap proses pembelajaran tentang interaksi yang terjadi pada suatu kejadian.

1. Pada tanggal 25 Agustus 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,9812 dan harga saham diprediksi naik dari harga 6600,0 menjadi 6650,84 rupiah. *Tweet* berisi rencana *stock split* dan regulasi baru terkait dengan persyaratan menjadi bank digital.
2. Pada tanggal 26 Agustus 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,8779 dan harga saham diprediksi naik dari harga 6560,0 menjadi 6577,3867 rupiah. *Tweet* berisi aplikasi BLU dari BCA sangat bagus dan lebih konsumen lebih mengandalkan bank konvensional.
3. Pada tanggal 27 Agustus 2021, sentimen saham BBCA bernilai negatif yaitu nilai *compound score* -0,0094 dan harga saham diprediksi naik dari harga 6510,0 menjadi 6624,68 rupiah. *Tweet* berisi berupa rencana *stock split* saham BBCA dan secara teknikal saham BBCA naik.
4. Pada tanggal 30 Agustus 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,3612 dan harga saham diprediksi naik dari harga 6565,0 menjadi 6619,91 rupiah. *Tweet* berisi saham BBCA secara teknikal terindikasi mengalami kenaikan.
5. Pada tanggal 31 Agustus, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,3818 dan harga saham diprediksi naik dari harga 6550,0 menjadi 6571,60 rupiah. *Tweet* berisi secara teknikal saham bank BCA mengalami kenaikan dan artikel berita terkait investasi bisnis BBCA.

6. Pada tanggal 01 September 2021, sentimen saham BBCA bernilai positif dan harga saham diprediksi turun dari harga 6565,0 menjadi 6511,54 rupiah. *Tweet* berisi harga saham BBCA berkisar di angka 29.850-33.000.
7. Pada tanggal 02 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,3182 dan harga saham diprediksi turun dari harga 6540,0 menjadi 6521,45 rupiah. *Tweet* berisi pergerakan menjelang penutup harga saham BBCA mengalami kenaikan dan secara teknikal terdapat indikasi beli.
8. Pada tanggal 03 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,6908 dan harga saham diprediksi turun dari harga 6600,0 menjadi 6554,42 rupiah. *Tweet* berisi saham BBCA mengalami *rebound* dan diakumulasi oleh investor asing. Penjualan saham BBCA cukup banyak.
9. Pada tanggal 06 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,7003 dan harga saham diprediksi turun dari harga 6575,0 menjadi 6547,28 rupiah. *Tweet* berisi penjualan saham BBCA cukup tinggi.
10. Pada tanggal 07 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,6597 dan harga saham diprediksi naik dari harga 6570,0 menjadi 6592,84 rupiah. *Tweet* berisi *public expose* 2021 yang diselenggarakan BBCA.
11. Pada tanggal 08 September 2021, sentimen saham BBCA bernilai negatif yaitu nilai *compound score* -0,1531 dan harga saham diprediksi naik dari harga 6440,0 menjadi 6590,9 rupiah. *Tweet* berisi saham bank BCA memiliki *market cap* terbesar di Indonesia dan harga saham mengalami kenaikan jelang penutupan harga saham.
12. Pada tanggal 09 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,9805 dan harga saham diprediksi naik dari harga 6570,0 menjadi 6605,5 rupiah. *Tweet* berisi *market cap* saham BCA salah satu yang terbesar di Indonesia.
13. Pada tanggal 10 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,5106 dan harga saham prediksi turun dari harga 6520,0 menjadi 6515,9 rupiah. *Tweet* berisi Rencana pembelian saham BBCA oleh investor setelah melakukan *stocksplit*.
14. Pada tanggal 13 September 2021 sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,8428 dan prediksi turun dari harga 6555,0 menjadi 6499,94 rupiah. *Tweet* berisi rencana pembelian saham BBCA dengan seluruh dana.

15. Pada tanggal 14 September 2021, sentimen saham BBCA bernilai negatif yaitu nilai *compound score* -0.34 dan harga saham prediksi turun dari harga 6545,0 menjadi 6488,70 rupiah.
16. Pada tanggal 15 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,296 dan harga saham diprediksi naik dari harga 6495,0 menjadi 6563,8 rupiah. *Tweet* berisi pembelian saham BBCA dengan tujuan untuk tabungan pernikahan.
17. Pada tanggal 16 September 2021, sentimen saham BBCA bernilai turun yaitu nilai *compound score* -0.4019 dan harga saham diprediksi naik dari harga 6500,0 menjadi 6543,86 rupiah.
18. Pada tanggal 17 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,296 dan harga saham diprediksi naik dari harga 6520,0 menjadi 6533,01 rupiah.
19. Pada tanggal 20 September 2021, sentimen saham BBCA bernilai netral dan harga saham diprediksi turun dari harga 6585,0 menjadi 6456,65 rupiah. *Tweet* tidak ada.
20. Pada tanggal 21 September 2021, sentimen saham BBCA bernilai netral dan harga saham diprediksi turun dari harga 6490,0 menjadi 6481,26 rupiah. *Tweet* tidak ada.
21. Pada tanggal 22 September 2021, sentimen saham BBCA bernilai positif dan harga saham diprediksi naik dari harga 6555,0 menjadi 6488,29 rupiah. *Tweet* berisi pemegang saham BBCA menyetujui rencana *stocksplit* bank BCA.
22. Pada tanggal 23 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0.6597 dan harga saham diprediksi turun dari harga 6580,0 menjadi 6519,72 rupiah. *Tweet* berisi rencana *stocksplit* saham BBCA.
23. Pada tanggal 24 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,92 dan harga saham diprediksi turun dari harga 6585,0 menjadi 6534,44 rupiah. *Tweet* berisi harga saham BBCA dipermainkan bila dilihat dari grafik.
24. Pada tanggal 27 September 2021, sentimen saham BBCA bernilai positif dan harga saham diprediksi turun dari harga 6580,0 menjadi 6528,62 rupiah.
25. Pada tanggal 28 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0.4019 dan harga saham diprediksi naik dari harga 6520,0 menjadi 6566,04 rupiah. *Tweet* berisi keinginan investor membeli saham BBCA sebesar 1000 lot.

26. Pada tanggal 29 September 2021, sentimen saham BBCA bernilai positif yaitu nilai *compound score* 0,926 dan harga saham diprediksi naik dari harga 6580,0 menjadi 6613,18 rupiah. *Tweet* berisi keinginan investor membeli saham BBCA sebesar 15 lot.

4.4 Pembahasan

Berdasarkan Gambar 4.3, diketahui bahwa prediksi dengan menambahkan data sentimen memiliki nilai *error* yang kecil. Hal ini, didukung oleh penelitian (Renault, 2017) yang menyatakan bahwa sentimen investor mampu berkontribusi pada perkiraan pola indeks saham. Setiap investor yang berinvestasi pada portofolio saham tentunya memiliki sentimen yang berbeda-beda. Sentimen tersebut didasari oleh perilaku dan sifat masing-masing investor. Ada investor yang memandang suatu saham itu akan mencapai profit di masa yang akan datang, namun ada juga yang tidak. Senada dengan hal tersebut, sentimen investor dan volatilitas *return* saham Bursa Efek Johannesburg menunjukkan hubungan yang signifikan (Rupande et al., 2019). Sentimen investor dapat menjadi suatu kecenderungan sebagai bahan spekulasi atau perasaan optimis atau pesimis dalam suatu aset yang diberikan. Sebagai contoh, dampak kasus Covid-19 menyebabkan saham mengalami penurunan yang ditunjukkan dengan nilai signifikansi $0.00 < 0.05$ (Nurmasari, 2020).

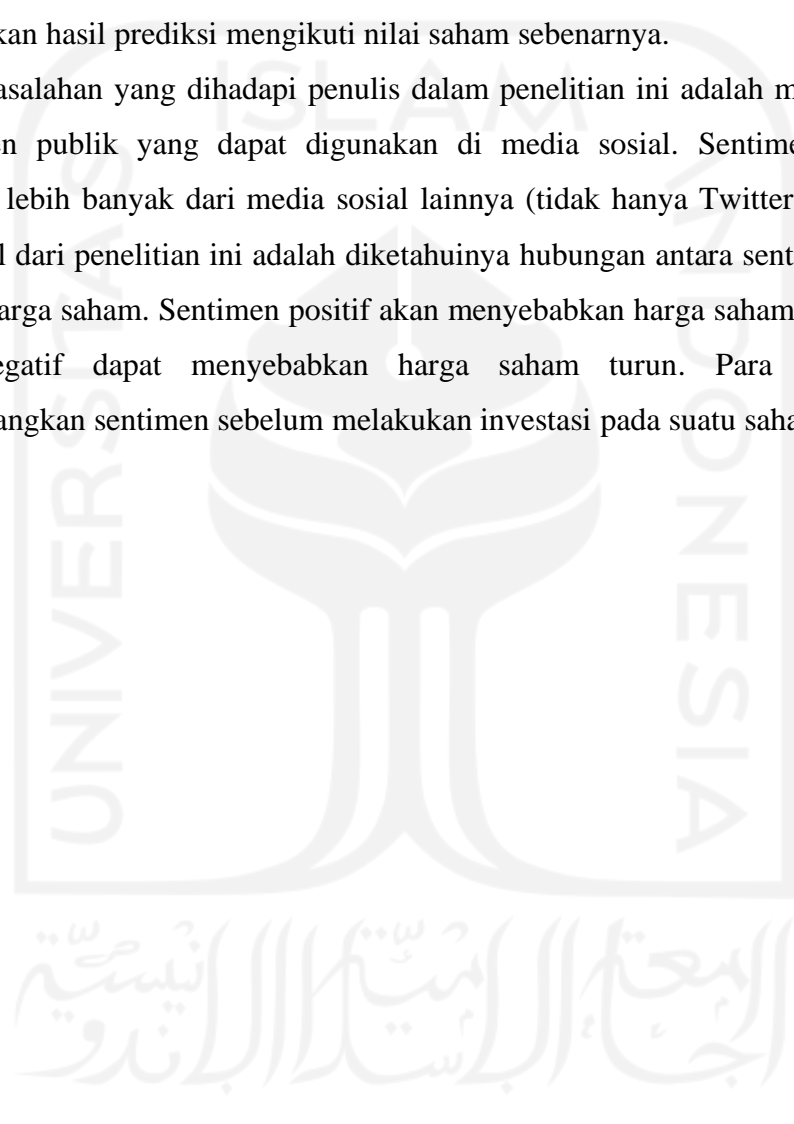
Karno (2020) menjelaskan prediksi harga saham yang dilakukan menggunakan metode LSTM dengan data bank BRI dan didapatkan nilai MSE sebesar 227,4703. Namun, penelitian tersebut masih menggunakan nilai *epoch* sebanyak sembilan. Pemilihan jumlah *epoch* pada suatu penelitian harus dilakukan secara hati-hati agar mendapatkan nilai akurasi yang baik dan program dapat berjalan dalam waktu yang lebih cepat. Senada dengan penelitian tersebut, penelitian lain Julian & Pribadi (2021) menggunakan metode LSTM dengan data saham TINS dan mendapatkan nilai RMSE sebesar 31,71. Nilai *epoch* yang digunakan sebanyak 200. Pengkajian ulang kembali terhadap penentuan jumlah parameter dilakukan karena kemungkinan besar jumlah parameter berdampak pada hasil prediksi. Pada penelitian tersebut, hanya dilakukan pengkajian pada variasi *epoch* saja. Penelitian terdahulu menyarankan agar menggunakan variance lain LSTM.

Penulis menggunakan salah satu *variance* dari LSTM yaitu *bidirectional* LSTM. Berdasarkan hasil penelitian, didapatkan nilai terkecil MSE sebesar 0,094, nilai terkecil RMSE sebesar 0,306 dan nilai MAPE sebesar 34,43. Skenario keempat menghasilkan nilai matriks evaluasi terkecil. Jumlah parameter *hidden layer*, *epoch*, dan banyak data menentukan seberapa kecil nilai MSE, RMSE, dan MAPE yang didapat. Tabel 4.3 menunjukkan hasil pengujian keempat skenario model. Gambar 4.8 memperlihatkan data

history saham yang di tambahkan dengan sentimen publik yang mengikuti nilai saham sebenarnya.

Penulis menambahkan eksperimen dengan tanpa menggabungkan data sentimen. Tabel 4.4 menunjukkan hasil pengujian empat skenario model tanpa sentimen. Skenario pertama menghasilkan nilai matriks terkecil. Berdasarkan hasil eksperimen didapatkan nilai MSE terkecil sebesar 0,148, nilai RMSE sebesar 0,385 dan nilai MAPE sebesar 31,66. Gambar 4.9 menunjukkan hasil prediksi tanpa menggunakan data sentimen. Skenario pertama memperlihatkan hasil prediksi mengikuti nilai saham sebenarnya.

Permasalahan yang dihadapi penulis dalam penelitian ini adalah mengenai sumber data sentimen publik yang dapat digunakan di media sosial. Sentimen publik perlu ditambahkan lebih banyak dari media sosial lainnya (tidak hanya Twitter). Manfaat yang dapat diambil dari penelitian ini adalah diketahuinya hubungan antara sentimen publik dan pergerakan harga saham. Sentimen positif akan menyebabkan harga saham naik sedangkan sentimen negatif dapat menyebabkan harga saham turun. Para investor dapat mempertimbangkan sentimen sebelum melakukan investasi pada suatu saham.



BAB 5

Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan data, penulis memperoleh kesimpulan yang dapat diambil dari penelitian mengenai Prediksi Saham dengan *Bidirectional LSTM* dan Analisis Sentimen yaitu:

1. Penelitian ini mempunyai tujuan untuk membangun model Prediksi harga saham dengan *Bidirectional Long-Short Term Memory*.
2. Penelitian menggunakan fitur harga penutup saham (*close*) dan nilai *compound score*. Nilai *compound score* adalah jumlah keseluruhan *lexicon* (Positif, negatif dan netral) yang telah dinormalisasi antara -1 dan +1.
3. Peneliti membangun model dengan menggabungkan data sentimen dan tanpa menggabungkan data sentimen.
4. Skenario 1 memakai data *training* mulai dari tanggal 05-07-2021 hingga tanggal 16-08-2021. Skenario 2 memakai data *training* mulai dari tanggal 05-07-2021 hingga tanggal 16-08-2021. Skenario 3 memakai data *training* mulai dari tanggal 19-02-2020 hingga tanggal 16-08-2021. Skenario keempat memakai data *training* mulai dari tanggal 19-02-2020 hingga tanggal 16-08-2021. Data *testing* untuk setiap skenario yang digunakan adalah mulai dari tanggal 18-09-2021 hingga dengan 29-09-2021.
5. Data *testing* untuk setiap skenario yang digunakan adalah mulai dari tanggal 18-09-2021 hingga dengan 29-09-2021.
6. Evaluasi model menggunakan metode *mean squared error* (MSE), *root mean square error* (RMSE), dan *mean absolute percentage error* (MAPE).
7. Pengujian model dengan menggabungkan data sentimen mendapatkan nilai error terkecil yaitu MSE sebesar 0.094, RMSE sebesar 0.306, dan MAPE 23.85. Pengujian model tanpa menggabungkan data sentimen mendapatkan nilai error terkecil yaitu MSE sebesar 0,150, RMSE sebesar 0,380, dan MAPE sebesar 30,95.

5.2 Saran

Saran untuk penelitian selanjutnya adalah menggunakan data sentimen dari platfrom sosial media lain. Sosial media yang lebih spesifik membahas terkait dengan saham sehingga data yang digunakan lebih akurat.

Daftar Pustaka

- Ahmed, G., Gasaymeh, A. Al, & Mehmood, T. (2017). The Global Financial Crisis and International Trade. *Asian Economic and Financial Review*, 7(6), 600–610.
<https://doi.org/10.18488/journal.aefr.2017.76.600.610>
- Alamsyah, A., Ayu, S. P., & Rikumahu, B. (2019). Exploring relationship between headline news sentiment and stock return. *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*, 1–6.
<https://doi.org/10.1109/ICoICT.2019.8835298>
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107.
<https://doi.org/10.3923/jeasci.2017.4102.4107>
- B. Caluza, L. J. (2017). Deciphering West Philippine Sea: A Plutchik and VADER Algorithm Sentiment Analysis. *Indian Journal of Science and Technology*, 11(47), 1–12. <https://doi.org/10.17485/ijst/2018/v11i47/130980>
- Batra, R., & Daudpota, S. M. (2018). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. *2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, ICoMET 2018 - Proceedings, 2018-Janua*, 1–5. <https://doi.org/10.1109/ICOMET.2018.8346382>
- Bhuriya, D., Kaushal, G., Sharma, A., & Singh, U. (2017). Stock market predication using a linear regression. *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017-Janua*, 510–513.
<https://doi.org/10.1109/ICECA.2017.8212716>
- Campan, A., Atnafu, T., Truta, T. M., & Nolan, J. (2019). Is Data Collection through Twitter Streaming API Useful for Academic Research? *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 3638–3643.
<https://doi.org/10.1109/BigData.2018.8621898>
- Chaturvedi, S., Mishra, V., & Mishra, N. (2018). Sentiment analysis using machine learning for business intelligence. *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017*, 2162–2166.
<https://doi.org/10.1109/ICPCSI.2017.8392100>
- Deeyaf, Q., & Padjadjaran, U. (2019). *Resesi Perekonomian Indonesia yang Berdampak*

Pada Kelesuan Ekonomi Resesi Perekonomian Indonesia yang Berdampak Pada Kelesuan Ekonomi. December.

- Ertugrul, A. M., & Karagoz, P. (2018). Movie Genre Classification from Plot Summaries Using Bidirectional LSTM. *Proceedings - 12th IEEE International Conference on Semantic Computing, ICSC 2018, 2018-Janua*, 248–251.
<https://doi.org/10.1109/ICSC.2018.00043>
- Ghiassi, M., & Lee, S. (2018). A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, 106, 197–216. <https://doi.org/10.1016/j.eswa.2018.04.006>
- Ghosh, A., Bose, S., Maji, G., Debnath, N. C., & Sen, S. (2019). Stock price prediction using lstm on indian share market. *EPiC Series in Computing*, 63, 101–110.
<https://doi.org/10.29007/qgcz>
- Graves, A., Mohamed, A., & Hinton, G. (2013). *SPEECH RECOGNITION WITH DEEP RECURRENT NEURAL NETWORKS* Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton Department of Computer Science, University of Toronto. 3.
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5–6), 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>
- Gu, X., Zhang, H., Zhang, D., & Kim, S. (2017). DeepAM: Migrate APIs with multi-modal sequence to sequence learning. *IJCAI International Joint Conference on Artificial Intelligence*, 0, 3675–3681. <https://doi.org/10.24963/ijcai.2017/514>
- Handayani, L. P. S., & Prastyo, D. D. (2020). Analisis Likuiditas Saham Sektor Perbankan di BEI Menggunakan Analisis Intervensi dan Autoregressive Conditional Duration. *Inferensi*, 3(1), 47. <https://doi.org/10.12962/j27213862.v3i1.6881>
- Harish Rao M , Shashikumar D.R, H. R. M. , S. D. . (2017). Automatic Product Review Sentiment Analysis Using Vader and Feature Visulaization. *International Journal of Computer Science Engineering and Information Technology Research*, 7(4), 53–66.
<https://doi.org/10.24247/ijcseitraug20178>
- Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), 1–30.
<https://doi.org/10.3390/ASI4010009>
- Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, 30(4), 330–338.
<https://doi.org/10.1016/j.jksues.2016.04.002>

- Hutto, C.J. and Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for. *Eighth International AAAI Conference on Weblogs and Social Media*, 18.
<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109>
- Izzah, A., Sari, Y. A. A., Widyastuti, R., & Cinderatama, T. A. (2017). Mobile app for stock prediction using Improved Multiple Linear Regression. *Proceedings - 2017 International Conference on Sustainable Information Engineering and Technology, SIET 2017, 2018-Janua*, 150–154. <https://doi.org/10.1109/SIET.2017.8304126>
- Jahan, I., & Sajal, S. (2018). *Stock Price Prediction using Recurrent Neural Network (RNN) Algorithm on Time-Series Data*.
- Jamshed, H., Khan, M. S. A., Khurram, M., Inayatullah, S., & Athar, S. (2019). Data Preprocessing: A preliminary step for web data mining. *3C Tecnología_Glosas de Innovación Aplicadas a La Pyme, May*, 206–221.
<https://doi.org/10.17993/3ctecno.2019.specialissue2.206-221>
- Joosery, B., & Deepa, G. (2019). Comparative analysis of time-series forecasting algorithms for stock price prediction. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3373477.3373699>
- Joshi, A., Bhattacharyya, P., & Ahire, S. (2017). *A Practical Guide to Sentiment Analysis*.
https://doi.org/10.1007/978-3-319-55394-8_5
- Julian, R., & Pribadi, M. R. (2021). Peramalan Harga Saham Pertambangan Pada Bursa Efek Indonesia (BEI) Menggunakan Long Short Term Memory (LSTM). *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(3), 1570–1580.
<https://doi.org/10.35957/jatisi.v8i3.1159>
- Karim, M., & Das, S. (2018). Sentiment Analysis on Textual Reviews. *IOP Conference Series: Materials Science and Engineering*, 396(1). <https://doi.org/10.1088/1757-899X/396/1/012020>
- Karno, A. S. B. (2020). Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory). *Journal of Informatic and Information Security*, 1(1), 1–8. <https://doi.org/10.31599/jiforty.v1i1.133>
- Khedr, A. E., Salama, S. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, 9(7), 22–30.
<https://doi.org/10.5815/ijisa.2017.07.03>
- Kordonis, J., Symeonidis, S., & Arampatzis, A. (2016). Stock price forecasting via sentiment analysis on Twitter. *ACM International Conference Proceeding Series*.

<https://doi.org/10.1145/3003733.3003787>

- Kumar, A., Purohit, K., & Kumar, K. (2021). Stock Price Prediction Using Recurrent Neural Network and Long Short-Term Memory. *Lecture Notes in Networks and Systems*, 175(February), 153–160. https://doi.org/10.1007/978-3-030-67187-7_17
- Kumar, Y. J., Goh, O. S., Basiron, H., Choon, N. H., & Suppiah, P. C. (2016). A review on automatic text summarization approaches. *Journal of Computer Science*, 12(4), 178–190. <https://doi.org/10.3844/jcssp.2016.178.190>
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 260–270. <https://doi.org/10.18653/v1/n16-1030>
- Maimaiti, M., Wumaier, A., Abiderexiti, K., & Yibulayin, T. (2017). Bidirectional long short-term memory network with a conditional random field layer for Uyghur part-of-speech tagging. *Information (Switzerland)*, 8(4), 1–12. <https://doi.org/10.3390/info8040157>
- Mathur, R., Pathak, V., & Bandil, D. (2019). Emerging Trends in Expert Applications and Security. In *Emerging Trends in Expert Applications and Security* (Vol. 841). <https://doi.org/10.1007/978-981-13-2285-3>
- Mike Schuster and Kuldeep K. Paliwal. (1997). Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing*, 8(2), 1–4. [https://doi.org/10.1016/s1634-6939\(13\)59289-1](https://doi.org/10.1016/s1634-6939(13)59289-1)
- Muklis, F. (2016). Perkembangan Dan Tantangan Pasar Modal Indonesia. *Al Masraf (Jurnal Lembaga Keuangan Dan Perbankan)*, 1(1), 1–12.
- Nafan, M. Z., & Amalia, A. E. (2019). Kecenderungan Tanggapan Masyarakat terhadap Ekonomi Indonesia berbasis Lexicon Based Sentiment Analysis. *Jurnal Media Informatika Budidarma*, 3(4), 268. <https://doi.org/10.30865/mib.v3i4.1283>
- Nasution, Y. S. J. (2015). Peranan Pasar Modal Dalam Perekonomian Negara. *HUMAN FALAH: Jurnal Ekonomi Dan Bisnis Islam*, 2(1), 95–112.
- Nurmasari, I. (2020). Dampak Covid - 19 Terhadap Perubahan Harga Saham dan Volume Transaksi (Studi Kasus Pada PT. Ramayana Lestari Sentosa, Tbk.). *Jurnal Sekuritas*, 3(1), 230–236.
- Osaze BE. (2000). *The Nigeria Capital Market in the African and Global Financial System*. Bofic Consults Group Limited.

- Permata, C. P., & Ghoni, M. A. (2019). Peranan Pasar Modal Dalam Perekonomian Negara Indonesia. *Jurnal Akun Stie*, 5(2), 50–61.
- PT Kustodian Sentral Efek Indonesia. (2019). *Pertumbuhan Single Investor Identification (SID)*. https://www.ksei.co.id/files/uploads/fokuss_bulletins/fokuss_file/id-id/58_edisi_2_20190902150110.pdf
- Putra, J. W. G. (2019). Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. *Computational Linguistics and Natural Language Processing Laboratory*, 4, 1–235. <https://www.researchgate.net/publication/323700644>
- Putri, L. P. (2015). *PENGARUH PROFITABILITAS TERHADAP HARGA SAHAM PADA PERUSAHAAN PERTAMBANGAN BATU BARA DI INDONESIA*. 2015(02), 1–239.
- Rahmah, M. (2019). *Pasar Modal Pasar Modal*. 1–5.
- Ramadani, L. A. (2020). Oversubscribed, undersubscribed, dan Initial Public Offering (IPO) saham Syariah di Indonesia. *Journal of Enterprise and Development*, 2(1), 51–61. <https://doi.org/10.20414/jed.v2i01.2045>
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking and Finance*, 84, 25–40. <https://doi.org/10.1016/j.jbankfin.2017.07.002>
- Rupande, L., Muguto, H. T., & Muzindutsi, P. F. (2019). Investor sentiment and stock return volatility: Evidence from the Johannesburg Stock Exchange. *Cogent Economics and Finance*, 7(1). <https://doi.org/10.1080/23322039.2019.1600233>
- Saputra, D. (2018). Pengaruh Manfaat, Modal, Motivasi, dan Edukasi Terhadap Minat Berinvestasi di Pasar Modal (Studi pada Mahasiswa Sekolah Tinggi Ilmu Ekonomi Balikpapan). *Future Jurnal Manajemen Dan Akuntansi*, 5(2), 178–190.
- Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 3285–3292. <https://doi.org/10.1109/BigData47090.2019.9005997>
- Sihaloho, E. D. (2020). Dampak Covid-19 Terhadap Perekonomian Indonesia Dampak Covid-19 Terhadap Perekonomian Indonesia. *ResearchGate*, April. <https://doi.org/10.13140/RG.2.2.13651.94241/1>
- Suchy, E., & Soltes, M. (2018). Influence of Capital Market on the US Unemployment and GDP. *International Scientific Conference*, 708–714.
- Sumarni, Y. (2020). Pandemi Covid 19: Tantangan Ekonomi dan Bisnis. *Al-Intaj Jurnal Ekonomi Dan Perbankan Syariah*, 2(1), 46–58.

- Suyanto. (2019). *DATA MINING UNTUK KLASIFIKASI DAB KLASTERISASI DATA*. Informatika Bandung.
- Suyanto, Kurniawan Nur Ramadhani, S.T., M. T., & Satria Mandala, P. D. (2019). *DEEP LEARNING MODERNISASI MACHINE LEARNING UNTUK BIG DATA*. Informatika Bandung.
- Trupthi, M., Pabboju, S., & Narasimha, G. (2017). Sentiment analysis on twitter using streaming API. *Proceedings - 7th IEEE International Advanced Computing Conference, IACC 2017*, 915–919. <https://doi.org/10.1109/IACC.2017.0186>
- U, R., B, B. G. H., Kale, M., Mankame, P., & Kulkarni, G. (2018). *Deep Learning for Digital Text Analytics: Sentiment Analysis*. 1–8. <http://arxiv.org/abs/1804.03673>
- UNDANG-UNDANG REPUBLIK INDONESIA. (2016). UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 8 TAHUN 1995 TENTANG PASAR MODAL. *Covering Globalization*, 17–31. <https://doi.org/10.7312/schi13174-003>
- Van Essen, B., Kim, H., Pearce, R., Boakye, K., & Chen, B. (2015). *Lbann*. 1–6. <https://doi.org/10.1145/2834892.2834897>
- Wang, Z., Ho, S. B., & Lin, Z. (2019). Stock market prediction analysis by incorporating social and news opinion and sentiment. *IEEE International Conference on Data Mining Workshops, ICDMW, 2018-Novem*, 1375–1380. <https://doi.org/10.1109/ICDMW.2018.00195>
- Westergaard, D., Stærfeldt, H. H., Tønsberg, C., Jensen, L. J., & Brunak, S. (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS Computational Biology*, 14(2), 1–16. <https://doi.org/10.1371/journal.pcbi.1005962>
- Wibowo, A., & Darmanto, S. (2017). Reaction of Indonesian Capital Market Investors to the Implementation of Tax Amnesty. *Jurnal Keuangan Dan Perbankan*, 21(4), 597–608. <https://doi.org/10.26905/jkdp.v21i4.1607>
- Word Bank. (2020). *Indonesia Economic Prospects The Long Road to Recovery*. <http://documents1.worldbank.org/curated/en/804791594826869284/pdf/Indonesia-Economic-Prospects-The-Long-Road-to-Recovery.pdf>
- Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T. C., & Coimbra, C. F. M. (2018). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy*, 168(October 2017), 60–101. <https://doi.org/10.1016/j.solener.2017.11.023>
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data.

Information Fusion, 42(August 2017), 146–157.

<https://doi.org/10.1016/j.inffus.2017.10.006>

Zhichen Lu, Wen Long, & Ying Guo. (2018). Extreme Market Prediction for Trading Signal with Deep Recurrent Neural Network. In *Iccs* (Vol. 1). School of Economics and Management, University of Chinese Academy of Sciences.

<https://doi.org/10.1007/978-3-319-93701-4>

