

# **PREDIKSI SEKTOR EKONOMI SEBELUM DAN SELAMA PANDEMI COVID-19**

Laporan ini dibuat untuk memenuhi kelulusan  
Matakuliah Program Internship



**Oleh:**

**Heriyanto**

**1184023**

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA  
POLITEKNIK POS INDONESIA  
TAHUN 2022**

# **PREDICT ECONOMY SECTOR BEFORE AND DURING COVID-19 PANDEMIC**

This report was made to fulfil requirement for  
Internship Program



**By:**

**Heriyanto**

**1184023**

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA  
POLITEKNIK POS INDONESIA  
TAHUN 2022**

## ABSTRAK

Pergerakan pasar saham selalu ambigu bagi investor karena berbagai faktor yang berpengaruh, sehingga prediksi yang akurat sulit didapatkan. Faktor yang dapat dilihat sekarang adalah pandemi COVID-19 yang mengakibatkan kenaikan dan penurunan harga saham. Studi ini bertujuan untuk memprediksi harga penutupan saham sebelum dan selama pandemi COVID-19 dengan *Machine Learning* dan algoritma *Deep Learning* untuk mengatasi masalah tersebut. Penelitian ini membandingkan dua model pembelajaran mesin (*Decision Tree* dan *Random Forest*), kemudian dua model pembelajaran mendalam (*Recurrent Neural Network* (RNN) dan *Long short-term memory* (LSTM)) untuk memprediksi harga penutupan saham dari berbagai sumber data, yaitu data saham, data inflasi, dan data *kurs*. Data-data tersebut terdiri dari parameter-parameter *Open*, *High*, *Low*, *Close*, *Adj Close* dan *Volume*. Parameter tersebut diuji korelasinya untuk memilih fitur yang akan digunakan sebagai masukan untuk model atau data latih. Masukan model kemudian dilatih dengan menggunakan algoritma *Principal Components Analysis* (PCA) untuk memperoleh nilai evaluasi informasi data yang ditampung oleh masukan model. Model dievaluasi menggunakan indikator strategis standar: RMSE, MSE, MAE, dan MAPE untuk menilai kinerja prediksi. Berdasarkan hasil evaluasi menunjukkan bahwa model *Random Forest* mengungguli model prediksi lain dengan perbedaan yang cukup besar. Rendahnya nilai indikator tersebut menunjukkan bahwa model tersebut efisien dalam memprediksi harga penutupan saham sebelum dan selama pandemi COVID-19.

Kata Kunci: Prediksi, Saham, COVID-19, *Machine Learning*, *Deep Learning*, *Principal Components Analysis*

## **ABSTRACT**

Stock market movements are always ambiguous for investors due to various influencing factors, so that accurate predictions are difficult to obtain. The factor that can be seen now is the COVID-19 pandemic which has resulted in increases and decreases in stock prices. This study aims to predict the closing price of stocks before and during the COVID-19 pandemic with Machine Learning and Deep Learning algorithms to overcome these problems. This study compares two machine learning models (Decision Tree and Random Forest), then two deep learning models (Recurrent Neural Network (RNN) and Long short-term memory (LSTM)) to predict stock closing prices from various data sources, namely stock data, inflation data, and the exchange rate data. The data consists of parameters Open, High, Low, Close, Adj Close and Volume. These parameters are tested for correlation to select the features that will be used as input for the model or training data. The model input is then trained using the Principal Components Analysis (PCA) algorithm to obtain the evaluation value of the data information that is accommodated by the model input. The model was evaluated using standard strategic indicators: RMSE, MSE, MAE, and MAPE to assess predictive performance. Based on the evaluation results, it is shown that the Random Forest model outperforms other predictive models by quite a large difference. The low value of this indicator shows that the model is efficient in predicting the closing price of shares before and during the COVID-19 pandemic.

**Keywords:** Prediction, Stocks, COVID-19, Machine Learning, Deep Learning, Principal Components Analysis

## **KATA PENGANTAR**

Puji dan syukur saya panjatkan kepada Allah Swt. atas ridanya saya dapat menyelesaikan laporan Internship II ini. Adapun judul Internship II saya adalah “Prediksi Sektor Ekonomi Sebelum dan Selama Pandemi COVID-19”.

Laporan ini diajukan untuk memenuhi syarat kelulusan mata kuliah Internship II di Diploma IV Politeknik Pos Indonesia. Tidak dapat disangkal bahwa butuh usaha yang keras dalam penyelesaian pengerjaan laporan ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling saya yang mendukung dan membantu. Terima kasih saya sampaikan kepada:

1. Kedua orang tua tercinta yang senantiasa memberikan dukungan dan doanya.
2. Dr. Ir. Agus Purnomo, M.T. selaku Direktur Politeknik Pos Indonesia.
3. M.Yusril Helmi Setyawan, S. Kom., M. Kom. selaku Ketua Program Studi DIV Teknik Informatika.
4. Syafrial Fachri Pane, S.T., M.T.I., EBDP. CDSP, SFPC selaku Koordinator Internship II dan Pembimbing Internal yang telah memberikan pengarahan dan membimbing kepada penulis.
5. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

# DAFTAR ISI

<b>ABSTRAK.....</b>	<b>iii</b>
<b>ABSTRACT .....</b>	<b>iv</b>
<b>KATA PENGANTAR .....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR GAMBAR .....</b>	<b>viii</b>
<b>DAFTAR TABEL.....</b>	<b>ix</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah .....	2
1.3 Tujuan dan Manfaat.....	3
1.4 Ruang Lingkup .....	3
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>4</b>
2.1 State of The Art (SOTA) .....	4
2.2 Penelitian Terkait.....	6
2.3 Machine Learning.....	11
2.3.1 <i>Decision Tree</i> .....	13
2.3.2 <i>Random Forest</i> .....	15
2.4 <i>Deep Learning</i> .....	16
2.4.1 <i>Recurrent Neural Network</i> .....	17
2.4.2 <i>Long Short-Term Memory</i> .....	17
2.5 Metode Pra-Pemrosesan Data.....	18
2.5.1 Interpolasi Linear .....	18
2.5.2 <i>Pearson Correlation</i> .....	19
2.5.3 Principal Components Analysis .....	20
2.6 Metode Evaluasi Model <i>Machine Learning</i> .....	20
2.6.1 <i>Root Mean Square Error (RMSE)</i> .....	20
2.6.2 <i>Mean Square Error (MSE)</i> .....	21
2.6.3 <i>Mean Absolute Error (MAE)</i> .....	21
2.6.4 <i>Mean Absolute Percentage Error (MAPE)</i> .....	21
2.6.5 <i>K-Fold Cross Validation</i> .....	22
<b>BAB III ANALISA PERUSAHAAN .....</b>	<b>24</b>

3.1	Sejarah PT Pos Indonesia (Persero).....	24
3.2	Visi dan Misi .....	25
3.3	Tugas dan Fungsi PT. Pos Indonesia (Persero) .....	26
3.4	Struktur Organisasi Perusahaan dan Job Description.....	26
3.5	Deskripsi dan Ruang Lingkup Internship .....	27
<b>BAB IV METODE PENELITIAN.....</b>		<b>29</b>
4.1	Diagram Alur Metodologi Penelitian .....	29
4.2	Indikator Capaian Penelitian .....	29
<b>BAB V HASIL PENELITIAN .....</b>		<b>31</b>
5.1	Hasil Penelitian.....	31
<b>BAB VI PENUTUP .....</b>		<b>45</b>
6.1	Kesimpulan.....	45
<b>DAFTAR PUSTAKA .....</b>		<b>46</b>

## DAFTAR GAMBAR

Gambar 2.3.1 Supervised Learning Classification dan Regression .....	11
Gambar 2.3.2 Random Forest .....	15
Gambar 2.4.1 Recurrent Neural Network .....	17
Gambar 2.4.2 Long Short-Term Memory .....	17
Gambar 2.6.1 Skema 5-Fold Cross Validation .....	22
Gambar 3.4.1 Struktur Organisasi PT Pos Indonesia (Persero) .....	26
Gambar 4.1.1 Alur Metodologi Penelitian .....	29
Gambar 5.1.1 Skrip Pengambilan Data Pemodelan .....	31
Gambar 5.1.2 Skrip Menggabungkan Dataset dan Mengisi Nilai Yang Hilang .....	34
Gambar 5.1.3 Dataset Gabungan.....	34
Gambar 5.1.4 Skrip Korelasi Pearson .....	35
Gambar 5.1.5 Korelasi Pearson Sebelum Pandemi COVID-19.....	35
Gambar 5.1.6 Skrip Memisah Data Fitur dan Prediksi Sebelum Pandemi COVID-19 .....	36
Gambar 5.1.7 Skrip PCA Sebelum Pandemi COVID-19.....	36
Gambar 5.1.8 Korelasi Pearson Selama Pandemi COVID-19 .....	37
Gambar 5.1.9 Skrip Memisah Data Fitur dan Prediksi Selama Pandemi COVID-19 .....	37
Gambar 5.1.10 Skrip PCA Selama Pandemi COVID-19.....	38
Gambar 5.1.11 Skrip Model Random Forest dan Decision Tree .....	38
Gambar 5.1.12 Skrip Model Long Short Term Memory .....	38
Gambar 5.1.13 Skrip Recurrent Neural Network.....	39
Gambar 5.1.14 Skrip Grafik Prediksi Model .....	39
Gambar 5.1.15 Prediksi Saham Sebelum Pandemi COVID-19 .....	39
Gambar 5.1.16 Prediksi Saham Selama Pandemi COVID-19 .....	40
Gambar 5.1.17 Skrip Evaluasi Model .....	40
Gambar 5.1.18 K-Fold Cross Validation Random Forest Sebelum Pandemi COVID-19.....	42
Gambar 5.1.19 K-Fold Cross Validation Decision Tree Sebelum Pandemi COVID-19.....	42
Gambar 5.1.20 K-Fold Cross Validation Random Forest Selama Pandemi COVID-19 .....	42
Gambar 5.1.21 K-Fold Cross Validation Decision Tree Selama Pandemi COVID-19 .....	42
Gambar 5.1.22 Hasil Loss Sebelum Pandemi COVID-19 .....	43
Gambar 5.1.23 Hasil Loss Selama Pandemi COVID-19 .....	43



## DAFTAR TABEL

Tabel 2.2.1 Penelitian Terkait .....	6
Tabel 4.2.1 Indikator Capaian .....	29
Tabel 5.1.1 Dataset Saham .....	32
Tabel 5.1.2 Dataset Kurs .....	32
Tabel 5.1.3 Dataset Inflasi .....	32
Tabel 5.1.4 Jumlah Data .....	33
Tabel 5.1.5 Atribut Dataset .....	33
Tabel 5.1.6 Hasil Evaluasi Prediksi .....	41



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Coronavirus atau SARS-CoV-2 merupakan jenis virus baru yang ditemukan di Wuhan, Tiongkok, virus ini juga dikenal dengan virus COVID-19. Virus ini dapat menular dengan cepat dan menyebabkan gangguan pada system pernapasan manusia, infeksi paru-paru, bahkan kematian(Artha & Paramita, 2021). Karena penularan COVID-19 yang sangat cepat, Organisasi Kesehatan Dunia (WHO) menetapkan COVID-19 sebagai pandemi atau epidemi secara global yang menunjukkan bahwa penyebaran COVID-19 tak dapat dihindarkan di negara manapun(Rusyida & Pratama, 2020).

COVID-19 pertama kali diumumkan masuk ke Indonesia pada tanggal 2 Maret 2020(Nurmasari, 2020). Pemerintah Indonesia pun melakukan upaya untuk mencegah penyebaran COVID-19, yaitu dengan mengeluarkan kebijakan, antara lain melakukan *physical distancing*, menggunakan masker, meliburkan sekolah, *work from home*, dan lain-lain(Lathifah et al., 2021). Ancaman pandemi COVID-19 di Indonesia tidak hanya pada aspek kesehatan saja, akan tetapi juga berdampak pada aspek ekonomi. Akibatnya pertumbuhan ekonomi mulai melambat, terutama pada sektor saham, nilai tukar rupiah, dan inflasi(Irianto et al., 2021)(Haryanto, 2020)(Amalin & Panorama, 2020).

Berdasarkan data BEI, pada tanggal 8 April 2020 catatan jual bersih (*net sell*) asing IHSG turun 26.44% sebesar Rp. 15,01 triliun di pasar reguler, dan di pasar non reguler (tunai dan negosiasi) terjadi aksi beli (*net buy*) asing sebesar Rp. 2,94 triliun(Irianto et al., 2021). Sementara nilai tukar rupiah terhadap US\$ terkoreksi (melemah) sebesar -12,4% sejak awal pandemi 2 Maret sampai dengan 16 April 2020. Sebelum naiknya dollar akibat COVID-19, nilai tukar rupiah per dollar AS berkisar Rp. 14.000/US\$ dan pada saat pandemi COVID-19, nilai tukar rupiah mengalami depresiasi hingga Rp. 16.000/US\$(Haryanto, 2020). Sedangkan tingkat inflasi tahunan menurut data Badan Pusat Statistik (BPS) sebelum pandemi COVID-19 tercatat 2.72% dan pada saat pandemi COVID-19 tingkat inflasi tercatat mengalami penurunan dengan nilai 1.68%(Amalin & Panorama, 2020).

Oleh karena itu, akan sangat menarik jika dilakukan kajian yang lebih dalam terhadap sektor-sektor tersebut. Pada penelitian sebelumnya, peneliti menggunakan *dataset* harga saham, inflasi, dan *kurs* yang terdiri dari tujuh parameter, yaitu *date*, *open*, *high*, *low*, *close*, *adj close*,

dan *volume*. Peneliti menggabungkan ketiga *dataset* tersebut menjadi satu *dataset* Indeks Gabungan, *dataset* ini memiliki delapan parameter, yaitu *Date*, *IHK*, *Kurs*, *Stock Open Price*, *Stock High Price*, *Stock Low Price*, *Stock Close Price*, dan *Stock Adjusted Close Price*.

Parameter-parameter Indeks Gabungan tersebut selanjutnya akan diuji korelasinya menggunakan pendekatan *Deep Learning* untuk mengetahui pengaruh tingkat inflasi dan nilai tukar rupiah terhadap volatilitas harga saham sebelum dan selama pandemi COVID-19 dan melakukan prediksi terhadap harga saham dengan model *Deep Learning* yang diusulkan, yaitu *Bayesian Neural Network*. Hasil penelitian tersebut, menunjukkan hasil yang kurang memuaskan dengan mendapatkan nilai akurasi paling tinggi sebesar 81%. Pada penelitian ini, peneliti akan melakukan pengembangan pada penelitian sebelumnya dengan menggunakan pendekatan *Machine Learning* untuk membandingkan dan mendapatkan hasil prediksi yang lebih akurat dengan menggunakan *dataset* yang sama. Hasil prediksi tersebut, akan diukur kinerjanya berdasarkan nilai *Root Mean Squared Error* (RMSE), *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) serta mengevaluasi *overfitting* dari setiap model yang dibandingkan dengan menggunakan metode algoritma *k-fold cross validation*. *Overfitting* terjadi karena model yang dibuat terlalu fokus pada data *train* tertentu. Oleh karena itu, jika ada kumpulan data yang serupa, data tersebut tidak akan dapat membuat prediksi yang baik. *Overfitting* biasanya menangkap data *noise* yang diabaikan. Dengan demikian penelitian ini akan memprediksi harga saham sebelum dan selama pandemi COVID-19 dengan menggunakan model *machine learning* dengan tujuan untuk memperoleh hasil prediksi yang lebih akurat dan baik berdasarkan evaluasi nilai tolak ukur dan *overfitting* dari setiap model.

## 1.2 Identifikasi Masalah

Adapun identifikasi masalah pada penelitian ini sebagai berikut,

1. Bagaimana menganalisis *overfitting* pada setiap model *machine learning* yang digunakan?
2. Bagaimana mengevaluasi kinerja prediksi pada setiap model *machine learning* yang digunakan?

### 1.3 Tujuan dan Manfaat

Adapun tujuan pada penelitian ini sebagai berikut,

1. Melakukan analisis *overfitting* pada setiap model *machine learning* yang digunakan dengan menggunakan metode algoritma *k-fold cross validation*?
2. Mengevaluasi kinerja prediksi harga saham berdasarkan nilai tolak ukur yang digunakan.

Adapun manfaat pada penelitian ini sebagai berikut,

1. Merekomendasikan model prediksi yang lebih akurat dan memiliki kinerja yang lebih baik.
2. Mengetahui perbandingan kinerja akurasi prediksi harga saham menggunakan pendekatan *Machine Learning* dengan penelitian sebelumnya sebagai bahan evaluasi untuk pengembangan penelitian berikutnya.

### 1.4 Ruang Lingkup

Adapun ruang lingkup pada penelitian ini sebagai berikut,

1. Data yang digunakan bersumber dari penelitian sebelumnya.
2. Data yang digunakan merupakan data *time series* harian.
3. Periode data yang digunakan dari tahun 2019 hingga 2020.
4. Pendekatan yang digunakan adalah *Machine Learning*.
5. Bahasa pemrograman menggunakan python.
6. Software yang digunakan Jupyter.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 State of The Art (SOTA)**

Indonesia diterpa banyak masalah ekonomi akibat COVID-19. Perekonomian Indonesia diperkirakan akan memburuk pada tahun 2020, dengan meningkatnya pengangguran dan meningkatnya kemiskinan. Berdasarkan *Year on Year* pertumbuhan ekonomi kuartal pertama tahun 2020, adanya pelemahan hanya 2,97%, dibandingkan dengan capaian 5,07% pada kuartal pertama 2019 (BPS, 2020a). Data dari triwulan II juga kurang bersahabat, dengan kemunduran -5,32% (BPS, 2020b). Data triwulan III menunjukkan kenaikan inflasi sebesar 3,49% (BPS, 2020c), sedangkan data triwulan IV menunjukkan kenaikan inflasi sebesar 2,19% (BPS, 2021). Dampak dari peningkatan angka pengangguran dan penduduk miskin yang disebabkan karena PHK selama pandemi Covid-19.

Dampak ekonomi Indonesia yang menurun juga terasa pada sektor saham Indonesia, Pada penutupan perdagangan 30 Desember 2019 Indeks Harga Saham Gabungan (IHSG) masih menunjukkan kinerja yang baik pada level 6.299. Namun, saat penemuan kasus pertama COVID-19 di Indonesia pada tanggal 2 Maret 2020, IHSG menurun ke level 5.361 dan terus mengalami penurunan dan pada tanggal 23 Maret 2020 IHSG berada pada level terendah 3.937. Bank Dunia memperkirakan ekonomi Indonesia akan terus menurun pada tingkat 2,1%. Hal ini disebabkan oleh terus memburuknya kinerja COVID-19 baik di pasar domestik maupun internasional. Selain itu, berpotensi memperparah krisis ekonomi di sejumlah negara jika tidak ditangani secara cepat dan tuntas. Khususnya nilai tukar rupiah terhadap dolar (AS) yang semakin hari semakin fluktuatif. Nilai rupiah selama pandemi telah menyentuh angka Rp16.000 dan menjadi yang terlemah dalam kurun waktu 5 tahun terakhir (Hastuti, 2020).

Berdasarkan hal tersebut, peneliti mencoba melakukan prediksi terhadap harga saham dengan menggunakan data harga saham, nilai tukar rupiah, dan tingkat inflasi sebagai parameter untuk model prediksi. Untuk itu diperlukan pendekatan yang tepat untuk memodelkan parameter tersebut. Pendekatan *Machine Learning* merupakan salah satu metode untuk membuat model prediksi dan dapat diterapkan di berbagai bidang. Studi tentang prediksi harga saham telah dilakukan sebelumnya. Pada penelitian (Kilimci & Duvar, 2020), peneliti memprediksi pergerakan saham yang signifikan, nilai tukar dan arah pasar saham. Studi ini mengusulkan penyisipan kata dan prediksi berbasis *Deep Learning* untuk Bursa Saham Istanbul

(BIST 100) dengan menganalisis saham bank BIST 100 dalam jumlah besar. Peneliti menggunakan model *Long Short-Term Memory*, *Recurrent Neural Networks*, *Convolutional Neural Networks* sebagai algoritma *Deep Learning*. Hasil penelitian CNN, RNN, dan LSTM menunjukkan kinerja klasifikasi masing-masing 81,23%, 82,95%, 87,44%, sementara kombinasi *FastText* dan CNN, RNN, LSTM, masing-masing menunjukkan keberhasilan klasifikasi 80,37%, 80,13% dan 80,23%. Kemudian, studi (Pinelis & Ruppert, 2022) memprediksi *return price* dan volatilitas harga saham dengan menggunakan data dari *market return (Mkt)* dan *risk-free asset return (Rf)*. Peneliti menggunakan model *Random Forest* dan *Elastic Net*. *Random Forest* memberikan hasil yang lebih dengan akurasi sebesar 80,91% dibandingkan dengan model *Elastic Net* dengan akurasi sebesar 80,65%. Dalam studi ini, peneliti mengusulkan pendekatan *machine learning* untuk memprediksi harga saham dengan menggunakan tiga data, yaitu data saham, data inflasi, dan data *kurs*. Studi ini menampilkan model *machine learning* yang dapat memprediksi harga saham dengan akurasi tinggi dan *overfitting* yang rendah.

## 2.2 Penelitian Terkait

Tabel 2.2.1 Penelitian Terkait

No.	Area Penelitian	Tahun	Karakteristik Data	Metode	Model	Hasil Penelitian	Evaluasi				
							RMSE	MSE	MAE	MAPE	ACC
1.	Memprediksi harga saham (Chen & Zhou, 2021)	2021	Dataset kontruksi bank dan dataset pasar saham CSI 300, China	<i>Deep Learning</i>	<i>Long Short Term Memory</i>	Model LSTM yang diusulkan dengan <i>Genetic Algorithm</i> dapat secara efektif meningkatkan akurasi prediksi.	-	0.0039	-	-	-
2.	Prediksi tren pasar saham (Nabipour et al., 2020)	2020	Dataset bursa saham, Teheran	<i>Machine Learning, Deep Learning</i>	<i>s (Recurrent Neural Network (RNN), Long short-term memory (LSTM).</i>	Model LSTM dan RNN menunjukkan adanya peningkatan yang signifikan dalam kinerja model ketika menggunakan data	-	-	-	-	89%, 90%



						biner, bukan data kontinu.					
3.	Memprediksi harga saham berbasis <i>time series</i> dan model baru portofolio investasi (Rather, 2021).	2021	Data saham NIFTY-50 yang diperoleh dari bursa saham Nasional India.	<i>Deep Learning</i>	<i>Long-Short Term Memory Based Deep Neural Network</i>	Hasilnya menunjukkan bahwa model LSTM lebih unggul dari berbagai model standar lainnya.	187.22	-	-	-	-
4.	Memprediksi nilai pasar saham di masa depan (Moghar & Hamiche, 2020).	2020	Data terdiri dari <i>opening prices</i> harian dua saham di <i>New York Stock Exchange</i> NYSE (GOOGL dan NKE), Polandia	<i>Deep Learning</i>	<i>Recurrent Neural Networks (RNN), Long-Short Term Memory model (LSTM)</i>	Hasil pengujian menunjukkan bahwa model RNN dan LSTM dapat memprediksi perubahan <i>opening prices</i> dari kedua data saham.	-	-	-	-	-

5.	Memprediksi harga saham di masa depan (Saud & Shakya, 2020).	2020	Data stok historis dari tanggal 8 Agustus 2007 sampai dengan 5 November 2018, Nepal.	<i>Deep Learning</i>	<i>Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM)</i>	Hasil penelitian menyatakan bahwa model GRU dan LSTM memprediksi lebih baik.	-	-	-	4,71	-
6.	Memprediksi likuiditas pasar saham (Khang et al., 2021).	2021	Data pasar saham Vietnam dari tahun 2011 hingga 2019, Polandia.	<i>Machine Learning, Deep Learning</i>	<i>Long-Short Term Memory</i>	Hasil penelitian menunjukkan bahwa model LSTM lebih baik dengan nilai MSE rendah.	-	0,0169	-	-	-
7.	Prediksi harga penutupan saham (Vijh et al., 2020)	2020	Dataset harga saham lima perusahaan, India.	<i>Machine Learning</i>	<i>Artificial Neural Network, Random Forest</i>	Model ANN dan RF yang diusulkan menunjukkan bahwa model tersebut efisien dalam memprediksi.	0,42	-	0,70%	-	-

8.	Prediksi tren pasar saham (Wen et al., 2019).	2019	Dataset S&P 500, China.	<i>Deep Learning</i>	<i>Long Short Term Memory</i>	Hasil penelitian menemukan pola makroskopik dalam rangkaian waktu keuangan dan memberikan solusi baru untuk prediksi harga.	-	-	-	-	84,21%
9.	Memprediksi pergerakan harga saham (Akhtar et al., 2022).	2022	Dataset Kaggle, Arab Saudi	<i>Machine Learning</i>	<i>Random Forest, Support Vector Machine</i>	Hasil penelitian menunjukkan bahwa skor akurasi model SVM adalah 78,7% sedangkan skor akurasi <i>Random Forest</i> adalah 80,8%.	-	0.2433	-	1.02	80,3%
10.	Prediksi harga saham (Subasi et al., 2021).	2021	Dataset saham National Association of Securities	<i>Machine Learning</i>	<i>Random Forests (RF), Bagging</i>	Hasil menyimpulkan bahwa <i>Random Forest</i> dan <i>Bagging</i>	-	-	-	-	93%

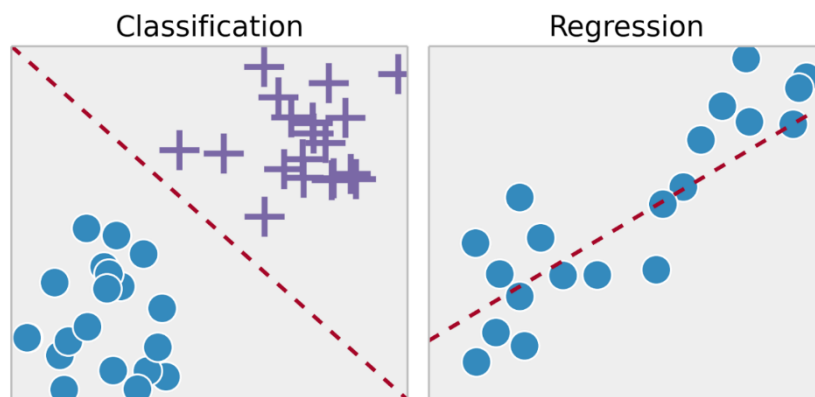
			Dealers Automated Quotations System (NASDAQ), New York Stock Exchange (NYSE), Nikkei, and Financial Times Stock Exchange (FTSE), Arab Saudi			memberikan kinerja lebih baik dibandingkan dengan model lainnya.					
--	--	--	---	--	--	--	--	--	--	--	--

## 2.3 Machine Learning

*Machine Learning* merupakan studi tentang pendekatan algoritma komputasi untuk membuat prediksi dan keputusan berdasarkan pengalaman dan data, tanpa secara eksplisit diprogram untuk melakukannya. Model *machine learning*, pada umumnya dibagi menjadi tiga kategori, yaitu sebagai berikut:

1. *Supervised learning*
2. *Unsupervised learning*
3. *Reinforcement learning*

Dalam penelitian ini, model yang digunakan termasuk dalam kategori *supervised learning regression*. *Supervised learning* merupakan metode *machine learning* yang paling sering digunakan. *Supervised learning* membutuhkan pembelajaran fungsi yang sesuai dengan pasangan *input* nilai dengan *output*. *Supervised learning* mengekstrak pengetahuan dari data *training* berlabel dan setiap pasangan *input* sesuai dengan nilai berlabel (Al-Azzam & Shatnawi, 2021). Algoritma *supervised learning* membutuhkan data berlabel untuk membangun sebuah model. Dalam algoritma *supervised learning*, ada dua variabel, yaitu variabel *input* yang biasa disebut variabel X dan variabel *output* yang biasa disebut variabel Y. Algoritma *supervised learning* bertujuan untuk mempelajari fungsi pemetaan dari variabel X ke variabel Y. Pada umumnya, rumus pemetaan variabel X dan Y adalah  $Y = f(X)$ . Rumus algoritma *supervised learning* ini digunakan untuk memperkirakan fungsi pemetaan ( $f$ ) agar dapat memprediksi variabel Y ketika memiliki data *input* (variabel X) yang baru.



Gambar 2.3.1 *Supervised Learning Classification dan Regression*

*Supervised Learning* dibagi menjadi dua jenis masalah, yaitu sebagai berikut:

1. Klasifikasi (*Classification*) : menggunakan algoritma *supervised learning* untuk menetapkan *train data* ke dalam kategori tertentu secara akurat. Jenis ini dapat mengenali entitas tertentu dalam data dan mencoba untuk menarik beberapa kesimpulan bagaimana entitas tersebut harus diberi label atau didefinisikan.
2. Regresi (*Regression*) : algoritma *supervised learning* digunakan untuk memahami hubungan antara variabel terikat dan variabel bebas. Biasanya jenis *supervised learning* regresi ini digunakan untuk membuat proyeksi.

Algoritma *supervised learning* mendeteksi pola dalam data *training* dan menghasilkan fungsi yang dapat memprediksi pasangan *input* baru atau pengamatan yang tidak pernah terlihat. Algoritma tersebut dapat menggeneralisasikan fungsi untuk memprediksi secara akurat. Algoritma *supervised learning* menerapkan langkah-langkah sebagai berikut:

1. Memperoleh kumpulan data: merupakan langkah pertama dalam metode *supervised learning*, yaitu dengan mengumpulkan data dan mengumpulkan sumber data yang relevan. Data harus cukup dan memiliki jumlah baris dan kolom yang cukup, karena ukuran dataset tergantung pada masalah yang diselesaikan.
2. Pemrosesan data: langkah kedua adalah pemrosesan data dengan mengisi nilai-nilai yang hilang, menormalkan data, dan menghapus *outlier*. Langkah ini merupakan langkah yang paling penting dalam proses *machine learning*, karena mempengaruhi keakuratan prediksi *machine learning*.
3. Menentukan jenis variabel target: jenis variabel target dalam kumpulan data akan menentukan algoritma *supervised learning* yang akan diterapkan. Jika tipe variabelnya kontinu, berarti itu adalah *regression problem*. Jika tipe variabelnya kategoris, berarti itu adalah *classification problem*.
4. Memisahkan data: langkah ini merupakan langkah untuk memisahkan data menjadi *train data* dan *test data*.
5. Melatih model: langkah selanjutnya adalah melatih model, artinya melatih *train data* dan *test data* dengan algoritma *machine learning*.
6. Prediksi: langkah terakhir adalah prediksi. *Input data* diterapkan pada model untuk memprediksi labelnya kemudian hasilnya dievaluasi berdasarkan *output* model.

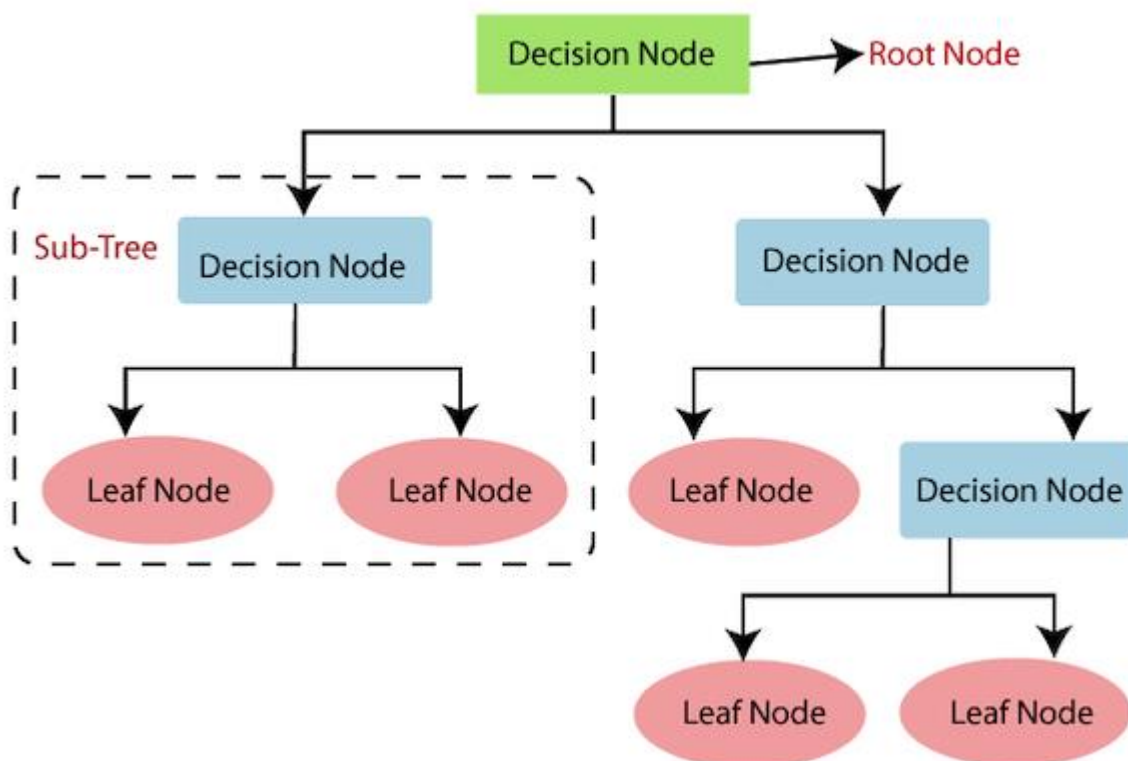
Berikut beberapa contoh algoritma yang digunakan pada *supervised learning*, yaitu:

1. *Clasification (Categorical) and Regression (Numerical)*

2. *Logistic Regression*
3. *Model Ensemble*
4. *Time series*

Adapun model *machine learning* yang digunakan pada penelitian ini adalah sebagai berikut:

### 2.3.1 *Decision Tree*



Konsep dari pohon keputusan adalah mengubah data menjadi *decision tree* dan aturan-aturan keputusan. Manfaat utama dari penggunaan *decision tree* adalah kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Nama lain dari *decision tree* adalah CART (*Classification and Regression Tree*). Dimana metode ini merupakan gabungan dari dua jenis pohon, yaitu *classification tree* dan juga *regression tree*. Untuk memudahkan, berikut ilustrasi dari keduanya. *Decision tree* juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. *Decision tree* memadukan antara eksplorasi data dan pemodelan,

sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

Kelebihan lain dari metode ini adalah mampu mengeliminasi perhitungan atau data-data yang kiranya tidak diperlukan. Sebab, sampel yang ada biasanya hanya diuji berdasarkan kriteria atau kelas tertentu saja. Meski memiliki banyak kelebihan, namun bukan berarti metode ini tidak memiliki kekurangan. *Decision tree* ini bisa terjadi overlap, terutama ketika kelas dan kriteria yang digunakan sangat banyak tentu saja dapat meningkatkan waktu pengambilan keputusan sesuai dengan jumlah memori yang dibutuhkan. Dalam hal akumulasi, *decision tree* juga seringkali mengalami kendala eror terutama dalam jumlah besar. Selain itu, terdapat pula kesulitan dalam mendesain *decision tree* yang optimal. Apalagi mengingat kualitas keputusan yang didapatkan dari metode *decision tree* sangat tergantung pada bagaimana pohon tersebut didesain.

Berikut adalah rumus perhitungan *decision tree*:

Perhitungan Gain:

$$GAIN(S, A) = Entropy(S) - \sum_{i=1}^n \times Entropy(S_i) \quad (1)$$

Keterangan:

S : himpunan  
A : atribut  
N : jumlah partisi atribut A  
| Si | : jumlah kasus pada partisi ke-i  
| S | : jumlah kasus dalam S

Menghitung nilai *entropy*:

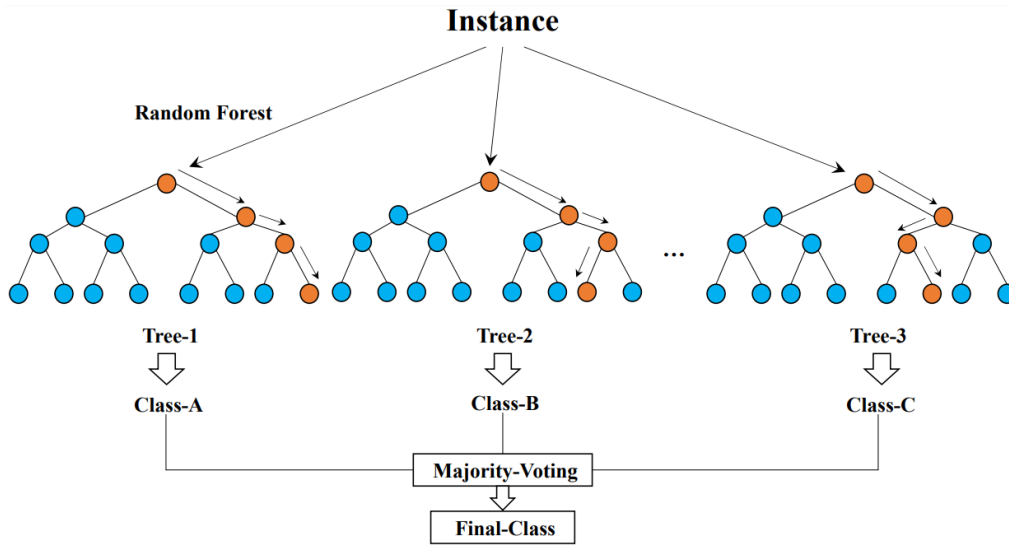
$$Entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (2)$$

Keterangan:

S : himpunan kasus  
A : fitur  
n : jumlah partisi S  
pi : proporsi dari Si terhadap S



### 2.3.2 Random Forest



Gambar 2.3.2 Random Forest

*Random Forest* adalah algoritma pembelajaran ensemble berdasarkan *Decision Tree*. RF menggunakan teknologi bootstrap untuk mengekstrak sampel acak dari sampel asli dan membangun pohon keputusan tunggal. Pada setiap *node* pohon keputusan, pilih titik pengurutan menggunakan *subspace* fitur acak. Terakhir, gabungkan pohon keputusan ini untuk mendapatkan hasil prediksi akhir dengan suara terbanyak (Zhu et al., 2021).

*Decision Tree* banyak digunakan sebagai pengklasifikasi seperti pohon. Dimulai pada *node* akar pohon, memilih atribut kolom terbaik, selanjutnya terurai, dan membangun *node* pohon satu per satu sampai kondisi berhenti untuk membangun pohon tercapai. Saat membuat prediksi, *Decision Tree* dimulai dari *node* akar dan menentukan satu-satunya jalur ke *node* daun. Sebagian besar kategori sampel *node* daun merupakan nilai prediksi untuk *Decision Tree*. Menurut berbagai kriteria yang diterapkan saat memilih atribut klasifikasi, algoritma *Decision Tree* yang umum adalah ID3 dan C4.5 berdasarkan pembagian entropi Shannon, dan CART berdasarkan klasifikasi koefisien Gini. Ketika digunakan untuk tugas klasifikasi, algoritma C4.5 adalah algoritma yang paling umum digunakan. Membagi *Decision Tree* menggunakan rasio perolehan informasi sebagai kriteria pemisahan. Dirumuskan sebagai berikut.

$$g_r(X, Y) = \frac{g(X, Y)}{H(Y)} \quad (3)$$

$$g = (X, Y) = H(Y) - H(Y|X) \quad (4)$$

Di mana  $g(X, Y)$  mengacu dalam perolehan informasi dan rasio ketidakpastian peristiwa, yang dihitung menggunakan mengurangi entropi bersyarat dari entropi. Rasio perolehan

informasi menambahkan kata penalty  $H(Y|X)$  menurut perolehan informasi. Rumus perhitungan penalti adalah:

$$H(Y|X) = \sum_{i=1}^n p_i H(Y|X = x_i) \quad (5)$$

Rumus perhitungan entropi adalah:

$$H(p) = - \sum_{i=1}^n p_i \log_2 p_i \quad (6)$$

Semakin besar entropi, semakin besar jumlah informasi yang terkandung dan semakin besar ketidakpastian. Pembentukan *Decision Tree* pada dasarnya memisahkan data lapis demi lapis dengan merumuskan aturan sehingga ketidakpastian dalam data terus berkurang setelah klasifikasi.

## 2.4 Deep Learning

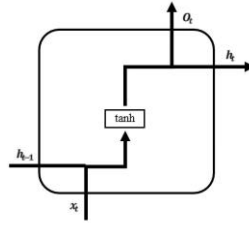
*Deep learning* merupakan bidang dari *machine learning* dan yang paling sering digunakan karena dapat memecahkan masalah metode tradisional. Teknik *deep learning*, arsitekturnya adalah *unsupervised learning*, *supervised learning*, dan sebagainya (Gao & Lima, 2022). *Deep learning* yang merupakan cabang dari *machine learning* terdiri dari operasi berurutan yang secara progresif mengekstraksi fitur kompleks dengan memanfaatkan hasil operasi sebelumnya sebagai *input*. *Deep learning model* memiliki keunggulan yang berbeda berdasarkan paradigma berbasis data, dan sebagai hasilnya, dapat lebih fleksibel dalam secara otomatis memperoleh informasi tingkat tinggi dari *big data* (Ma & Mei, 2021).

*Deep learning* merupakan kelas pendekatan *data modeling* tingkat lanjut yang dikembangkan dalam beberapa tahun terakhir. Karena kemampuannya yang kuat dalam *non-linear modeling* dan *high-dimensional information modeling* (Wang et al., 2022). Pada penelitian ini, penulis menggunakan metode *deep learning* untuk membangun model prediksi harga penutupan dan harga penutupan yang disesuaikan. Berikut beberapa manfaat dari penerapan *deep learning*:

1. Dapat memproses *unstructured data* seperti teks dan gambar.
2. Dapat mengotomatisasi proses ekstraksi fitur tanpa perlu melakukan proses pelabelan secara manual.
3. Memberikan hasil akhir yang berkualitas.
4. Dapat mengurangi biaya operasional.
5. Dapat melakukan manipulasi data dengan lebih efektif.

Adapun model *deep learning* yang digunakan pada penelitian ini adalah sebagai berikut:

### 2.4.1 Recurrent Neural Network



Gambar 2.4.1 Recurrent Neural Network

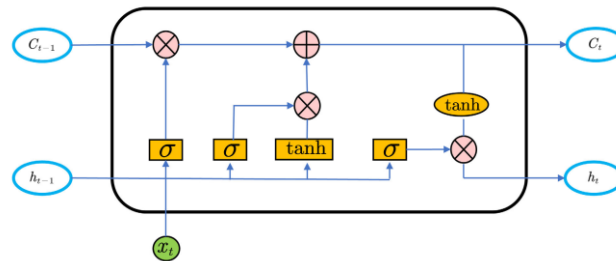
*Recurrent Neural Network* (RNN) adalah jenis ANN dengan "memori" yang menangkap informasi masa lalu dan cocok untuk urutan data dengan panjang berapa pun (Song et al., 2021). "Memori" ini disebut *hidden state* dan merupakan fungsi utama dan terpenting dari RNN. Sebuah RNN terdiri dari *input layer*, *hidden layer*, dan *output layer*. *Hidden state*  $h_t$  dan *output* dari satu *hidden layer*  $o_t$  yang diberikan urutan *input* pada langkah waktu  $t$   $x_t$  dapat diturunkan sebagai berikut.

$$h_t = H(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (7)$$

$$o_t = O(W_{ho}h_t + b_o) \quad (8)$$

Di mana  $W$  menunjukkan *weights* koneksi antara lapisan yang dibagi di semua langkah dan  $b$  mewakili vektor bias. Persamaan (1) dan (2) dan Gambar menunjukkan bahwa *hidden state*  $h_t$  dihitung berdasarkan *hidden state* sebelumnya  $h_{t-1}$  dan *input* pada langkah saat ini  $x_t$ .  $H(\cdot)$  dan  $O(\cdot)$  adalah fungsi aktivasi, yang dalam banyak kasus dinyatakan sebagai *tanh*.

### 2.4.2 Long Short-Term Memory



Gambar 2.4.2 Long Short-Term Memory

Hochreiter dan Schmidhuber (Song et al., 2021) mengusulkan varian dari RNN yang disebut LSTM yang dirancang untuk menghindari masalah ketergantungan jangka panjang. Blok

memori yang terdiri dari beberapa gerbang dan sel memori adalah fitur terpenting dari sistem LSTM. Gambar 2.3.2 menunjukkan blok memori dari model LSTM.

LSTM memiliki tiga gerbang utama: *input gate* ( $i_t$ ), *forget gate* ( $f_t$ ), dan *output gate* ( $o_t$ ). *Input gate* mengontrol sinyal *input* yang mengubah keadaan sel memori. *Forget gate* mengatur nomor status sel ( $h_{t-1}$ ) sebelum dapat dilewati. *Output gate* menentukan apakah keadaan sel memori memungkinkannya mempengaruhi unit lain. Perhitungan keadaan setiap gerbang dan sel dinyatakan sebagai berikut.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (9)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (10)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (11)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (12)$$

$$\tilde{c}_t = \tanh(W_{co}x_t + W_{co}h_{t-1} + b_c) \quad (13)$$

$$h_t = o_t \odot \tanh(c_t) \quad (14)$$

Di mana  $c_t$  adalah sel memori;  $\tilde{c}_t$  adalah *hidden state* internal;  $\sigma(\cdot)$  adalah fungsi sigmoid;  $\tanh(\cdot)$  adalah fungsi tangen hiperbolik; dan  $\odot$  adalah produk vektor *elementwise*.

## 2.5 Metode Pra-Pemrosesan Data

Pra-pemrosesan data adalah metode untuk menghilangkan beberapa masalah pada data, seperti nilai yang hilang, *noise* pada data dan lain-lain. Pada penelitian ini, ada tiga metode yang digunakan, yaitu interpolasi linear untuk mengisi nilai yang hilang. Kemudian, metode korelasi pearson untuk mengetahui keterkaitan antara variabel independent dengan variabel dependen untuk memilih fitur yang akan digunakan pada pemodelan. Terakhir, metode *principal components analysis* untuk mengevaluasi persentase fitur data pelatihan sebelum melakukan pemodelan. Adapun rincian lengkap dari metode-metode yang telah disebutkan adalah sebagai berikut:

### 2.5.1 Interpolasi Linear

Di dunia nyata, data seringkali tidak disajikan secara utuh. Seringkali ada nilai data yang hilang. Ada banyak alasan untuk ini, mulai dari kesalahan manusia hingga keterbatasan kemampuan alat ukur. Kondisi lain yang muncul dari data yang kita miliki adalah adanya outlier atau nilai yang sangat berbeda dengan sebagian besar data yang kita miliki. Nilai ini akan

menentukan hasil analisis atau uji statistik yang kita lakukan, terutama jika kita menggunakan metode parametrik. Interpolasi adalah teknik untuk menemukan nilai variabel yang hilang dalam rentang data yang diketahui.

Interpolasi *linear* adalah teknik penentuan nilai suatu fungsi pada setiap titik perantara ketika nilai dua titik yang berdekatan diketahui. Interpolasi *linear* pada dasarnya adalah perkiraan nilai yang tidak diketahui dalam dua nilai yang diketahui. Dengan menggunakan rumus berikut, kita dapat menginterpolasi secara linier titik-titik data yang diberikan.

$$\text{Linear Interpolation : } y(x) = y_1 + (x - x_1) \frac{(y_2 - y_1)}{(x_2 - x_1)} \quad (15)$$

Berikut  $(x_1, y_1)$  adalah koordinat titik data pertama. Dan  $(x_2, y_2)$  adalah koordinat titik data kedua, di mana  $x$  adalah titik di mana dilakukan interpolasi dan  $y$  adalah nilai interpolasi.

### **2.5.2 Pearson Correlation**

Koefisien korelasi Pearson (dinamakan untuk Karl Pearson) dapat digunakan untuk meringkas kekuatan hubungan linier antara dua sampel data. Koefisien korelasi Pearson dihitung sebagai kovarians dari dua variabel dibagi dengan produk dari standar deviasi setiap sampel data. Ini adalah normalisasi kovarians antara dua variabel untuk memberikan skor yang dapat ditafsirkan. Penggunaan mean dan standar deviasi dalam perhitungan menunjukkan perlunya dua sampel data memiliki distribusi Gaussian atau seperti Gaussian. Hasil perhitungan, koefisien korelasi dapat diinterpretasikan untuk memahami hubungan. Koefisien mengembalikan nilai antara -1 dan 1 yang mewakili batas korelasi dari korelasi negatif penuh ke korelasi positif penuh. Nilai 0 berarti tidak ada korelasi.

Koefisien korelasi Pearson mengukur hubungan linier antar variabel. Nilainya dapat diartikan seperti ini:

1. +1 - Korelasi positif sangat kuat
2. +0,8 - Korelasi positif yang kuat
3. +0,6 - Korelasi positif sedang
4. 0 - tidak ada korelasi sama sekali
5. -0,6 - Korelasi negatif sedang
6. -0,8 - Korelasi negatif yang kuat
7. -1 - Korelasi negatif sangat

Persamaan untuk mendapatkan nilai *pearson correlation* dinyatakan pada persamaan di bawah berikut:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}} \quad (16)$$

Dimana x dan y adalah dua vektor dengan panjang n. m, x dan m, y masing-masing sesuai dengan rata-rata x dan y.

### 2.5.3 Principal Components Analysis

PCA merupakan teknik multivariat yang menganalisis tabel data di mana observasi dijelaskan oleh beberapa variabel dependen kuantitatif yang saling berkorelasi. Tujuannya untuk mengekstrak informasi penting dari tabel, untuk mewakilinya sebagai satu set variabel ortogonal baru yang disebut komponen utama, dan untuk menampilkan pola kesamaan pengamatan dan variabel sebagai titik dalam peta. Dengan demikian, PCA adalah teknik yang dapat digunakan untuk mereduksi kumpulan data multidimensi ke dimensi yang lebih rendah untuk analisis (Kamari & Peter Leslie Schultz, 2022). PCA adalah teknik statistik multivariat yang mengekstrak informasi dari data dimensi tinggi kemudian memproyeksikannya ke dimensi yang lebih rendah (Aidoo et al., 2021).

## 2.6 Metode Evaluasi Model *Machine Learning*

Evaluasi model *machine learning* dilakukan untuk menilai seberapa baik model tersebut dalam memprediksi. Model-model tersebut dievaluasi dengan melihat nilai *error* pada prediksi yang diperoleh. Ada beberapa perhitungan untuk mendapatkan nilai *error* tersebut, yaitu dengan perhitungan *root mean square error* (RMSE), *mean square error* (MSE), *mean absolute error* (MAE), dan *mean absolute percentage error* (MAPE). Terakhir, metode evaluasi dengan *k-fold cross validation* untuk mengecek *overfitting* pada model yang digunakan. Adapun rincian metode-metode evaluasi adalah sebagai berikut:

### 2.6.1 Root Mean Square Error (RMSE)

*Root Mean Square Error* (RMSE), adalah jumlah dari kesalahan kuadrat atau selisih antara nilai sebenarnya dengan nilai prediksi yang telah ditentukan (Khalimi, 2022a). Metode estimasi *Root Mean Square Error* (RMSE) lebih kecil dari 0,5 dikatakan lebih akurat (Khoiri, 2020c). Rumus formula RMSE adalah sebagai berikut:

$$RMSE = \sqrt{\sum \frac{(Y' - Y)^2}{n}} \quad (17)$$

Keterangan:

Y' : Nilai Prediksi

Y : Nilai Sebenarnya

n : Jumlah Data

### 2.6.2 Mean Square Error (MSE)

Untuk menghitung nilai MSE sama halnya dengan RMSE. Hanya saja tidak menggunakan proses akar. Nilai *Mean Squared Error* yang rendah atau mendekati nol menunjukkan bahwa hasil peramalan sesuai dengan data aktual dan dapat dijadikan untuk perhitungan peramalan di periode mendatang (Khoiri, 2020b).

$$MSE = \sum \frac{(Y' - Y)^2}{n} \quad (18)$$

Keterangan:

Y' : Nilai Prediksi

Y : Nilai Sebenarnya

n : Jumlah Data

### 2.6.3 Mean Absolute Error (MAE)

MAE atau *Mean Absolute Error* menunjukkan nilai kesalahan rata-rata yang *error* dari nilai sebenarnya dengan nilai prediksi. MAE sendiri secara umum digunakan untuk pengukuran prediksi error pada analisis *time series*. Nilai MAE yang lebih kecil dari 0,2 maka semakin baik model tersebut dalam melakukan prediksi (Khalimi, 2022a). Rumus dari MAE sendiri didefinisikan sebagai berikut:

$$MAE = \sum \frac{|Y' - Y|}{n} \quad (19)$$

Keterangan:

Y' : Nilai Prediksi

Y : Nilai Sebenarnya

n : Jumlah Data

### 2.6.4 Mean Absolute Percentage Error (MAPE)

*Mean Absolut Percentage error* (MAPE) adalah persentase kesalahan rata-rata secara multak (absolut). Pengertian *Mean Absolute Percentage Error* adalah Pengukuran statistik tentang akurasi perkiraan (prediksi) pada metode peramalan. Semakin kecil nilai persentasi kesalahan (*percentage error*) pada MAPE maka semakin akurat hasil peramalan tersebut (Khoiri, 2020a). Nilai MAPE dapat diinterpretasikan atau ditafsirkan ke dalam 4 kategori yaitu:

1.  $< 10\%$  = "Sangat Baik"
2.  $10-20\%$  = "Baik"
3.  $20-50\%$  = "Cukup Baik"
4.  $> 50\%$  = "Tidak Akurat"

Rumus *Mean Absolute Percentage Error* (MAPE) sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (20)$$

Keterangan:

$Y_t$  : nilai actual pada periode t

$\hat{Y}_t$  : nilai prediksi pada periode t

$n$  : Jumlah data

### 2.6.5 K-Fold Cross Validation

*Cross validation* adalah suatu metode tambahan dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal. Metode ini sering juga disebut dengan *k-fold cross validation* dimana percobaan sebanyak  $k$  kali untuk satu model dengan parameter yang sama (Khalimi, 2022b). *Cross-validation* (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi / evaluasi (Sitorus, 2020). Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Selanjutnya pemilihan jenis CV dapat didasarkan pada ukuran dataset. Biasanya CV K-fold digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi.

Percobaan 1	Test	Train	Train	Train	Train
Percobaan 2	Train	Test	Train	Train	Train
Percobaan 3	Train	Train	Test	Train	Train
Percobaan 4	Train	Train	Train	Test	Train
Percobaan 5	Train	Train	Train	Train	Test

Gambar 2.6.1 Skema 5-Fold Cross Validation

Berikut langkah-langkah dalam *k fold cross validation*, yaitu:

1. Total data dibagi menjadi  $k$  bagian.
2. *Fold* ke-1 menjadi data uji dan sisanya menjadi data latih, kemudian hitung akurasi suatu pengukuran dengan angka atau data sesuai dengan porsi data tersebut. Perhitungan akurasi menggunakan persamaan sebagai berikut:



$$Akurasi = \frac{\Sigma data\ uji\ benar\ klasifikasi}{\Sigma total\ data\ uji} \times 100 \quad (21)$$

3. *Fold* ke-2 menjadi data uji dan sisanya menjadi data latih, kemudian hitung akurasi berdasarkan porsi data tersebut.
4. Demikian seterusnya sampai *fold* ke- $k$ . hitung rata-rata akurasi dari  $k$ , rata-rata inilah hasil akurasi akhir.

## **BAB III**

### **ANALISA PERUSAHAAN**

#### **3.1 Sejarah PT Pos Indonesia (Persero)**

Diawali pada tahun 1602 saat perhubungan pos hanya di beberapa kota yang berada di Pulau Jawa dan Luar Pulau Jawa, surat – surat hanya diletakan di Gedung penginapan kota sehingga pengecekan harus dilakukan sendiri oleh orang – orang yang merasa mempunyai surat atau paket yang dikirimkan untuknya. Dalam rangka meningkatkan dan menjamin keamanan surat dan paket para penduduk, Pada tanggal 26 Agustus 1746, Gubernur Jenderal G.W Baron van Imhoff mendirikan kantor pos pertama di Batavia. Pelayanan pos sejak itu lahir dan mengemban peran serta fungsinya terhadap masyarakat.

Seiring berjalannya waktu peranan kantor pos semakin berkembang pesat, pos indonesia beberapa kali mengalami perubahan status yang dimulai dari Jawatan PTT (Post, Telegraph dan Telephone) berdasarkan statland nomor 395 tahun 1906. Dengan dikeluarkannya undang – undang perusahaan Negara Hindia Belanda (Indische bedrijvenwet). Sejak tahun 1907 Jawatan Pos Telegram dipegang oleh departemen perusahaan pemerintah (Departement Van Government Bedrijvenment).

Tibanya Jepang mengambil alih kekuasaan Belanda di Indonesia, Jawatan Pos Telegram Jawa dan Jawatan Pos Telegram Sulawesi. Jawatan Pos Telegram Republik Indonesia ini mulai berdiri secara resmi pada tanggal 27 September 1945 setelah dilakukan pengambilan alih kantor pos di Bandung oleh Angkatan Muda Pos Telegraf dan telepon yang disingkat menjadi AMPTT dari pemerintah militer Jepang. Dalam peristiwa ini gugur sekelompok pemuda anggota sehingga pada tanggal tersebut menjadi tolak awal berdirinya Jawatan Pos Telegraf & Telepon Republik Indonesia dan diperingati setiap tahunnya sebagai bakti PTT, yang kemudian menjadi hari bakti pariwisata, pos dan telekomunikasi (PARPOSTEL).

Mengamati perkembangan zaman dimana sektor pos dan telekomunikasi berkembang sangat pesat, maka pada tahun 1965 berganti menjadi Perusahaan Negara Pos dan Giro (PN Pos dan Giro), dan pada tahun 1978 berubah menjadi Perum Pos dan Giro yang sejak ini ditegaskan sebagai badan usaha tunggal 9 dalam menyelenggarakan dinas pos dan giropos baik untuk hubungan dalam maupun luar negeri. Selama 17 tahun berstatus Perum, maka pada 1995 berubah menjadi Perseroan Terbatas dengan nama PT Pos Indonesia (Persero) (persero)

dilaksanakan berdasarkan PP No. 5 Tahun 1995 tanggal 27 Februari 1995, dan perubahan tersebut secara efektif mulai berlaku pada tanggal 20 Juni 1995.

Seiring berjalannya waktu, Pos Indonesia kini telah mampu menunjukkan kreatifitasnya dalam pengembangan bidang perposan Indonesia. Dengan memanfaatkan infrastruktur jejaring yang dimilikinya yang mencapai sekitar 24 ribu titik layanan yang menjangkau 100 persen kota/kabupaten, hampir 100 persen kecamatan dan 42 persen kelurahan/desa, dan 940 lokasi transmigrasi terpencil di Indonesia.

### **3.2 Visi dan Misi**

#### **1. VISI**

Menjadi Postal Operator, Penyedia Jasa Kurir, Logistik dan Keuangan Paling Kompetitif.

#### **2. MISI**

- a. Berkomitmen kepada pelanggan untuk menyediakan layanan yang selalu tepat waktu dan nilai terbaik.
- b. Berkomitmen kepada karyawan untuk memberikan iklim kerja yang aman, nyaman dan menghargai kontribusi.
- c. Berkomitmen kepada pemegang saham untuk memberikan hasil usaha yang menguntungkan dan terus bertambah.
- d. Berkomitmen untuk berperilaku transparan dan terpercaya kepada seluruh pemangku kepentingan.

Untuk mewujudkan Visi dan Misi, persyaratan utama (key words) yang perlu dilakukan adalah:

- 1 Memberikan produk yang relevan sesuai dengan kebutuhan pasar;
- 2 Memberikan jasa layanan yang prima;
- 3 Menjalankan proses bisnis secara efisien;
- 4 Membangun solusi teknologi informasi yang prima dan human capital yang andal;
- 5 Memperkuat sistem pengendalian internal, governance, dan manajemen risiko untuk mencapai tingkat kematangan yang memadai untuk mengamankan pencapaian tujuan Perusahaan.

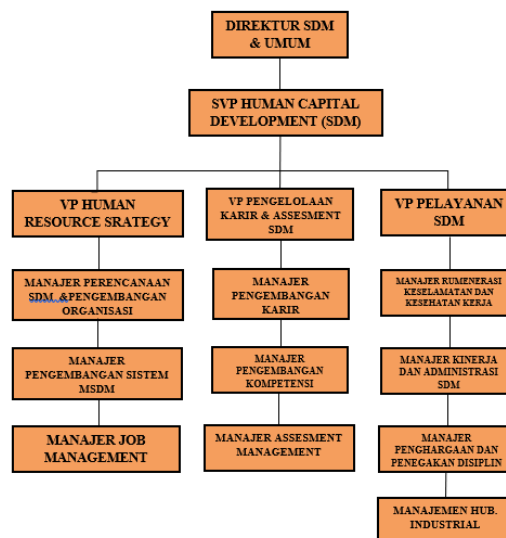
### 3.3 Tugas dan Fungsi PT. Pos Indonesia (Persero)

PT. Pos Indonesia mempunyai tugas pokok menyelenggarakan jasa pos untuk umum dalam dan luar negeri yang meliputi jasa pos, jasa keuangan, dan jasa keagenan serta usaha-usaha lain yang menunjang jasa pos sesuai dengan peraturan perundang-undangan yang berlaku. Untuk menjalankan tugas tersebut PT Pos Indonesia mempunyai fungsi diantaranya:

- 1 Pengembangan usaha dan pemasaran jasa pos dan giro pos serta jasa keagenan.
- 2 Pengelolaan SDM dan sarana.
- 3 Pembangunan sarana dan fasilitas
- 4 Pengelolaan keuangan perusahaan.
- 5 Pelaksanaan penelitian, pembangunan, dan penyusunan rencana kegiatan perusahaan.
- 6 Pengelolaan pendidikan dan pelatihan.
- 7 Penyelenggaraan usaha-usaha lain yang diperlukan sesuai dengan lapangan usahanya.
- 8 Penyelenggaraan kerjasama dengan badan atau pihak lain baik yang dianggap perlu untuk menunjang usaha perseroan.
- 9 Pendirian anak perusahaan.
- 10 Menghilangkan isolasi daerah terpencil.
- 11 Alat terdepan dalam usaha menghimpun dan menyalurkan dana dari dan kepada pemerintah.

### 3.4 Struktur Organisasi Perusahaan dan Job Description

#### 1. Struktur Organisasi



Gambar 3.4.1 Struktur Organisasi PT Pos Indonesia (Persero)

## 2. Job Description

### a. Human Resource Strategy

Tugas Pokok Human Resource Strategy adalah bertanggung jawab atas ketersediaan kebijakan dan peraturan manajemen sumber daya manusia yang bersifat strategis. Untuk melaksanakan tugas pokok tersebut Human Resource Strategy melakukan aktivitas utama sebagai berikut:

- Perencanaan strategis bidang manajemen sumber daya manusia meliputi; peraturan tata tertib karyawan, kebijakan rekrutasi, pengangkatan karyawan, sistem karir dan kompetensi karyawan, kenaikan gaji dan grade, penghargaan dan hukuman disiplin, sistem balas jasa, kebutuhan training karyawan, job management, pemutusan hubungan kerja, dan kebijakan sumber daya manusia strategis lainnya.
- Perencanaan dan pengembangan organisasi perusahaan.
- Mengevaluasi implementasi job management.

### b. Pengelolaan Karir dan Assemen SDM

Tugas pokok pengelolaan karir dan assesmen SDM adalah bertanggung jawab atas pemenuhan karyawan yang seimbang dan kompeten antara ketersediaan dan potensi karyawan, dan meningkatkan kompetensi karyawan, pengembangan karir karyawan, serta assesmen karyawan untuk mencapai sumber daya manusia yang handal sesuai dengan kebutuhan perusahaan.

### c. Pelayanan SDM

Tugas pokok Pelayanan SDM adalah bertanggung jawab atas pengelolaan remunerasi dan K3, pengelolaan kinerja, administrasi SDM, SIM SDM, dan reward and punishment guna mendukung operasional dan bisnis perusahaan, serta melaksanakan fungsi hubungan industrial.

## 3.5 Deskripsi dan Ruang Lingkup Internship

Kegiatan Praktek Kerja Lapangan yang dilaksanakan sesuai dengan ketentuan civitas academia oleh penulis di PT Pos Indonesia Assessment & Learning Center. Penulis diberi

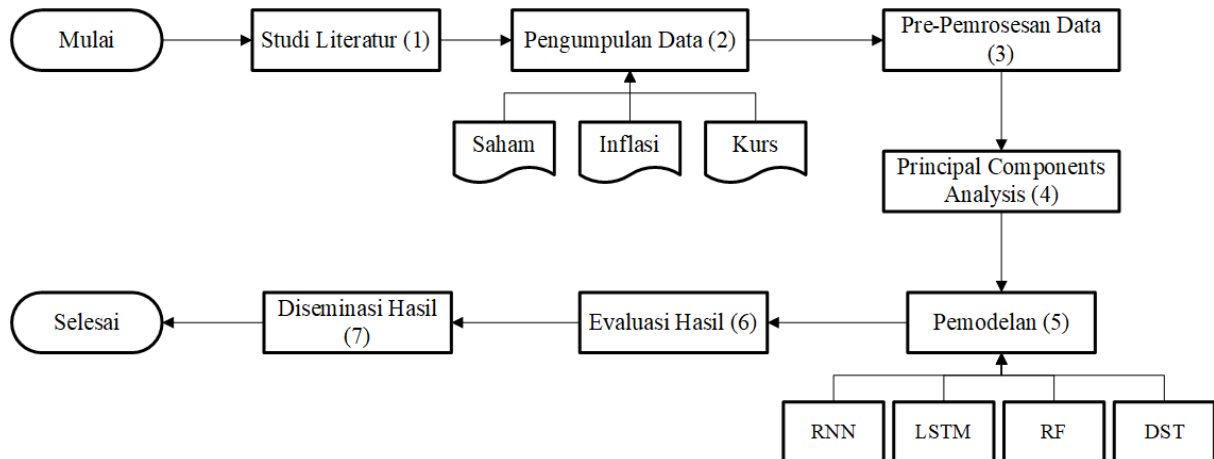
kesempatan untuk mengenal dan mempelajari deskripsi pekerjaan divisi Human Capital Development yang sebagian besar mempunyai tanggung jawab atas pengembangan sumber daya manusia di PT. POS Indonesia. Kegiatan yang dilaksanakan diantaranya ada yang berupa kegiatan rutin dan kegiatan insidental. Penugasan penulis yang ditempatkan pada divisi Human Capital Development oleh pembimbing Agus Suhendar di dasari dan di sesuaikan dengan kebutuhan yang ada di lapangan.

Selama kegiatan berlangsung penulis bertugas untuk membuat soal, mereview materi untuk pelatihan, menginput soal pelatihan dan menguji coba teknis pengejerjaan soal pada Learning Management System PT. Pos Indonesia. Setelah proses tersebut selesai penulis ditugaskan untuk merekap data dan nilai peserta pelatihan yang mana adalah karyawan PT. Pos Indonesia yang tersebar diseluruh Indonesia sebanyak 6 Kantor Regional.

## BAB IV

### METODE PENELITIAN

#### 4.1 Diagram Alur Metodologi Penelitian



Gambar 4.1.1 Alur Metodologi Penelitian

#### 4.2 Indikator Capaian Penelitian

Berdasarkan diagram alur metodologi penelitian diatas, terdapat indikator capaian sebagai berikut.

Tabel 4.2.1 Indikator Capaian

No.	Tahapan		Indikator capaian
1.	Studi Literatur	→	1. Deskripsi teori, temuan dan bahan penelitian terkait prediksi saham.
2.	Pengumpulan Data	→	2. Data mentah yang terdiri dari data saham, data inflasi, dan data <i>kurs</i> .
3.	Pre-Pemrosesan Data	→	3. Penggabungan data, dan penanganan nilai yang hilang (interpolasi linear), <i>features selection</i> (pearson correlation), <i>split</i> data fitur dan data target.

4.	<i>Principal Components Analysis</i>	→	4. Mengevaluasi persentase data pelatihan dengan algoritma PCA.
5.	Pemodelan	→	5. Model pendekatan <i>machine learning</i> dan <i>deep learning</i> digunakan untuk prediksi saham.
6.	Evaluasi Hasil	→	6. Menilai kinerja model.
7.	Diseminasi Hasil	→	7. Laporan internship dan jurnal nasional terakreditasi SINTA 3.



## BAB V

### HASIL PENELITIAN

#### 5.1 Hasil Penelitian

Kajian tentang prediksi harga saham telah banyak dilakukan, penelitian (Kilimci & Duvar, 2020), peneliti memprediksi pergerakan saham yang signifikan, nilai tukar dan arah pasar saham. Studi ini mengusulkan penyisipan kata dan prediksi berbasis *Deep Learning* untuk Bursa Saham Istanbul (BIST 100) dengan menganalisis saham bank BIST 100 dalam jumlah besar. Peneliti menggunakan model *Long Short-Term Memory*, *Recurrent Neural Networks*, *Convolutional Neural Networks* sebagai algoritma *Deep Learning*. Hasil penelitian CNN, RNN, dan LSTM menunjukkan kinerja klasifikasi masing-masing 81,23%, 82,95%, 87,44%. Kemudian penelitian yang dilakukan oleh (Subasi et al., 2021) memprediksi harga saham dengan dataset saham *National Association of Securities Dealers Automated Quotations System* (NASDAQ), *New York Stock Exchange* (NYSE), Nikkei, dan *Financial Times Stock Exchange* (FTSE). Model yang digunakan peneliti adalah *random forest* dan *bagging* dengan hasil akurasi 93%.

Pada penelitian ini, peneliti melakukan hal sama dengan memprediksi harga penutupan saham sebelum dan selama pandemi COVID-19 dengan menggunakan pendekatan *machine learning* dan *deep learning*. Adapun model yang diusulkan adalah model *random forest* (RF), *decision tree* (DST), *long short-term memory* (LSTM), dan *recurrent neural network* (RNN). Data yang digunakan pada penelitian ini adalah dataset saham, dataset *kurs*, dan dataset inflasi sebelum dan selama pandemi COVID-19 yang diperoleh dari *Yahoo Finance*. Tabel 5.1.1 - 5.1.3 menunjukkan dataset yang digunakan.

```
# Data Kurs
df = pd.read_csv('IDR=X.csv')
df = df.set_index('Date')
df = df.rename({'Open': 'Kurs Open', 'High': 'Kurs High',
               'Low': 'Kurs Low', 'Close': 'Kurs Close',
               'Adj Close': 'Kurs Adj Close', 'Volume': 'Kurs Volume'}, axis=1)

# Data Inflasi
df2 = pd.read_csv('^JKSE.csv')
df2 = df2.set_index('Date')
df2 = df2.rename({'Open': 'Ihk Open', 'High': 'Ihk High',
                  'Low': 'Ihk Low', 'Close': 'Ihk Close',
                  'Adj Close': 'Ihk Adj Close', 'Volume': 'Ihk Volume'}, axis=1)

# Data Saham
df3 = pd.read_csv('IMAS.JK.csv')
df3 = df3.set_index('Date')
df3 = df3.rename({'Open': 'Saham Open', 'High': 'Saham High',
                  'Low': 'Saham Low', 'Close': 'Saham Close',
                  'Adj Close': 'Saham Adj Close', 'Volume': 'Saham Volume'}, axis=1)
```

Gambar 5.1.1 Skrip Pengambilan Data Pemodelan

Tabel 5.1.1 Dataset Saham

Date	Open	High	Low	Close	Adj Close	Volume
01/01/2019	2160	2160	2160	2160	2129.737793	0
02/01/2019	2150	2180	2140	2150	2119.87793	446900
03/01/2019	2140	2190	2130	2150	2119.87793	576700
04/01/2019	2150	2220	2110	2140	2110.018066	2755000
07/01/2019	2140	2190	2130	2130	2100.158203	1773100
08/01/2019	2130	2140	2110	2110	2080.438477	578900
09/01/2019	2120	2230	2120	2230	2198.75708	4708800
10/01/2019	2230	2440	2230	2410	2376.235107	10715700
11/01/2019	2430	2640	2430	2600	2563.573486	15397800
14/01/2019	2590	2610	2490	2560	2524.133789	4563500
15/01/2019	2550	2870	2540	2840	2800.210938	11259000
16/01/2019	2840	3040	2750	2930	2888.949951	11253900
17/01/2019	2950	3010	2840	2900	2859.370361	3262500
18/01/2019	2900	2970	2870	2930	2888.949951	2697100
21/01/2019	2930	2950	2870	2900	2859.370361	1287500

Tabel 5.1.2 Dataset *Kurs*

Date	Open	High	Low	Close	Adj Close	Volume
01/01/2019	14460	14460	14460	14460	14460	0
02/01/2019	14460	14513	14373	14460	14460	0
03/01/2019	14409.7002	14502.5	14367.9004	14409.7002	14409.7002	0
04/01/2019	14375.4004	14403	14211	14375.4004	14375.4004	0
07/01/2019	14263	14263	14008	14267	14267	0
08/01/2019	14094	14164	13988	14083	14083	0
09/01/2019	14145	14192.5	14058	14138	14138	0
10/01/2019	14130	14130	14008	14130	14130	0
11/01/2019	14118	14118	13991	14048	14048	0
14/01/2019	14038	14176	14018	14038	14038	0
15/01/2019	14117	14166	14054	14117	14117	0
16/01/2019	14139	14185	14104	14111	14111	0
17/01/2019	14157	14302	14118	14140	14140	0
18/01/2019	14195.5	14236.5	14159	14173	14173	0
21/01/2019	14168	14263	14156	14168	14168	0

Tabel 5.1.3 Dataset Inflasi

Date	Open	High	Low	Close	Adj Close	Volume
02/01/2019	6197.871094	6205.89502	6164.833984	6181.174805	6181.174805	52797800
03/01/2019	6176.151855	6221.009766	6176.151855	6221.009766	6221.009766	72166700
04/01/2019	6211.096191	6274.540039	6200.854004	6274.540039	6274.540039	80858100
07/01/2019	6317.625977	6354.757813	6287.224121	6287.224121	6287.224121	90278300
08/01/2019	6292.263184	6316.240234	6251.375977	6262.847168	6262.847168	90537400

09/01/2019	6296.115234	6311.579102	6265.326172	6272.237793	6272.237793	105604200
10/01/2019	6293.958008	6328.713867	6268.507813	6328.713867	6328.713867	122462300
11/01/2019	6347.973145	6361.464844	6337.547852	6361.464844	6361.464844	78112000
14/01/2019	6351.329102	6359.875	6302.395996	6336.116211	6336.116211	79731000
15/01/2019	6347.101074	6408.78418	6333.644043	6408.78418	6408.78418	95139900
16/01/2019	6417.130859	6425.940918	6385.589844	6413.359863	6413.359863	104252800
17/01/2019	6421.407227	6457.734863	6410.950195	6423.779785	6423.779785	99214400
18/01/2019	6444.212891	6452.558105	6409.095215	6448.155762	6448.155762	77604300
21/01/2019	6448.536133	6472.149902	6438.900879	6450.833984	6450.833984	101769000
22/01/2019	6452.604004	6468.562012	6416.592773	6468.562012	6468.562012	100776200

Adapun Tabel 5.1.4 menunjukkan jumlah data saham, inflasi, dan *kurs* sebelum dan selama pandemi yang digunakan adalah sebagai berikut:

Tabel 5.1.4 Jumlah Data

Dataset	Jumlah Data	
	Sebelum Pandemi COVID-19	Selama Pandemi COVID-19
Saham	258	242
Inflasi	257	242
Kurs	261	262

Kemudian pada Tabel 5.1.5 menjelaskan tentang attribute data-data tersebut adalah sebagai berikut:

Tabel 5.1.5 Attribut Dataset

Attribut	Keterangan	Tipe Data
<i>Date</i>	Tanggal (Format Hari – Bulan – Tahun)	Tanggal
<i>Open</i>	Harga pembukaan pada hari tersebut	Numerik
<i>High</i>	Harga tertinggi pada hari tersebut	Numerik
<i>Low</i>	Harga terendah pada hari tersebut	Numerik
<i>Close</i>	Harga penutupan pada hari tersebut	Numerik
<i>Adj Close</i>	Harga penutupan yang disesuaikan dengan aksi korporasi seperti <i>right issue</i> , <i>split</i> atau <i>reverse</i> .	Numerik
<i>Volume</i>	Volume transaksi biasanya dalam jumlah lembar	Numerik

Ketiga dataset tersebut digabungkan menjadi satu dataset. Proses selanjutnya adalah mengisi nilai yang hilang pada dataset gabungan tersebut, metode yang digunakan interpolasi linear.

```
# Menggabungkan Dataset
frames = [df, df2, df3]
data = pd.concat(frames, axis = 1)

# Mengisi Nilai Yang Hilang
data1 = data.interpolate(method='linear', limit_direction = 'forward')
data1 = data1.interpolate(method='linear', limit_direction = 'backward')
```

Gambar 5.1.2 Skrip Menggabungkan Dataset dan Mengisi Nilai Yang Hilang

	Kurs Open	Kurs High	Kurs Low	Kurs Close \	
Date					
2019-01-01	14460.000000	14460.0	14460.000000	14460.000000	
2019-01-02	14460.000000	14513.0	14373.000000	14460.000000	
2019-01-03	14409.700195	14502.5	14367.900391	14409.700195	
2019-01-04	14375.400391	14403.0	14211.000000	14375.400391	
2019-01-07	14263.000000	14263.0	14008.000000	14267.000000	

	Kurs Adj Close	Kurs Volume	Ihk Open	Ihk High \	
Date					
2019-01-01	14460.000000	0.0	NaN	NaN	
2019-01-02	14460.000000	0.0	6197.871094	6205.895020	
2019-01-03	14409.700195	0.0	6176.151855	6221.009766	
2019-01-04	14375.400391	0.0	6211.096191	6274.540039	
2019-01-07	14267.000000	0.0	6317.625977	6354.757813	

	Ihk Low	Ihk Close	Ihk Adj Close	Ihk Volume	Saham Open \
Date					
2019-01-01	NaN	NaN	NaN	NaN	2160.0
2019-01-02	6164.833984	6181.174805	6181.174805	52797800.0	2150.0
2019-01-03	6176.151855	6221.009766	6221.009766	72166700.0	2140.0
2019-01-04	6200.854004	6274.540039	6274.540039	80858100.0	2150.0
2019-01-07	6287.224121	6287.224121	6287.224121	90278300.0	2140.0

	Saham High	Saham Low	Saham Close	Saham Adj Close	Saham Volume
...					
Date					
2019-01-01	2160.0	2160.0	2160.0	2129.737793	0.0
2019-01-02	2180.0	2140.0	2150.0	2119.877930	446900.0
2019-01-03	2190.0	2130.0	2150.0	2119.877930	576700.0
2019-01-04	2220.0	2110.0	2140.0	2110.018066	2755000.0

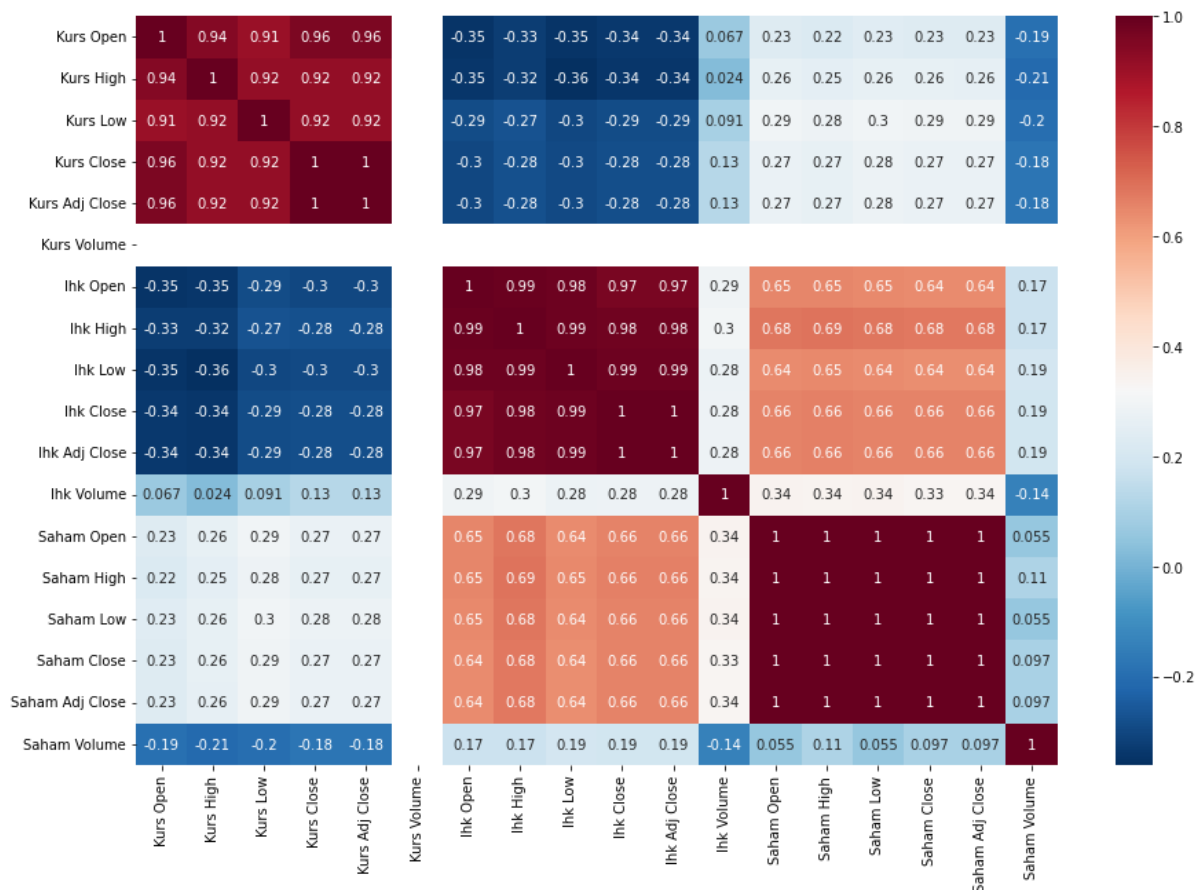
Gambar 5.1.3 Dataset Gabungan

Kemudian dilakukan langkah pemilihan fitur dengan menggunakan metode *pearson correlation* untuk memilih fitur atau variabel independen yang berhubungan erat dengan variabel dependen model, yaitu “Saham Close” sebelum dan selama pandemi COVID-19. Pemilihan variabel independen dilihat berdasarkan nilai korelasi yang lebih dari nol (0), hasil

korelasi sebelum pandemi COVID-19 dapat dilihat pada gambar 5.1.1 dan selama pandemi COVID-19 pada gambar 5.1.2.

```
pearsoncorr = data.corr(method='pearson')
fig, ax = plt.subplots(figsize=(15, 10))
sns.heatmap(pearsoncorr,
             xticklabels=pearsoncorr.columns,
             yticklabels=pearsoncorr.columns,
             cmap='RdBu_r',
             annot=True)
```

Gambar 5.1.4 Skrip Korelasi Pearson



Gambar 5.1.5 Korelasi Pearson Sebelum Pandemi COVID-19

Berdasarkan nilai korelasi sebelum pandemi COVID-19 pada gambar 5.1.5 di atas dapat diperhatikan bahwa variabel independen *kurs open*, *kurs high*, *kurs low*, *kurs close*, dan *kurs adj close* memiliki hubungan positif lemah terhadap variabel dependen *saham close* dengan

rata-rata nilai 0,2. Sedangkan, variabel *ihk open*, *ihk high*, *ihk low*, *ihk close*, dan *ihk adj close* memiliki hubungan positif sedang dengan rata-rata nilai 0,6, tetapi variabel *ihk volume* memiliki hubungan positif lemah dengan nilai 0,3. Terakhir, hubungan antara variabel saham *open*, saham *high*, saham *low*, dan saham *adj close* memiliki hubungan positif yang sangat kuat dengan nilai korelasi satu (1), tetapi hubungan variabel saham *volume* sangat lemah dengan variabel dependen saham *close*.

```
# Data Fitur
sc1 = MinMaxScaler(feature_range=(-1,1))
Xs = sc1.fit_transform(data1[['Kurs Open', 'Kurs High', 'Kurs Low', 'Kurs Close', 'Kurs Adj Close',
                              'Ihk Open', 'Ihk High', 'Ihk Low', 'Ihk Close', 'Ihk Adj Close', 'Ihk Volume',
                              'Saham Open', 'Saham High', 'Saham Low', 'Saham Adj Close', 'Saham Volume']])

# Data Prediksi
sc2 = MinMaxScaler(feature_range=(-1,1))
ys = sc2.fit_transform(data1[['Saham Close']])
```

Gambar 5.1.6 Skrip Memisah Data Fitur dan Prediksi Sebelum Pandemi COVID-19

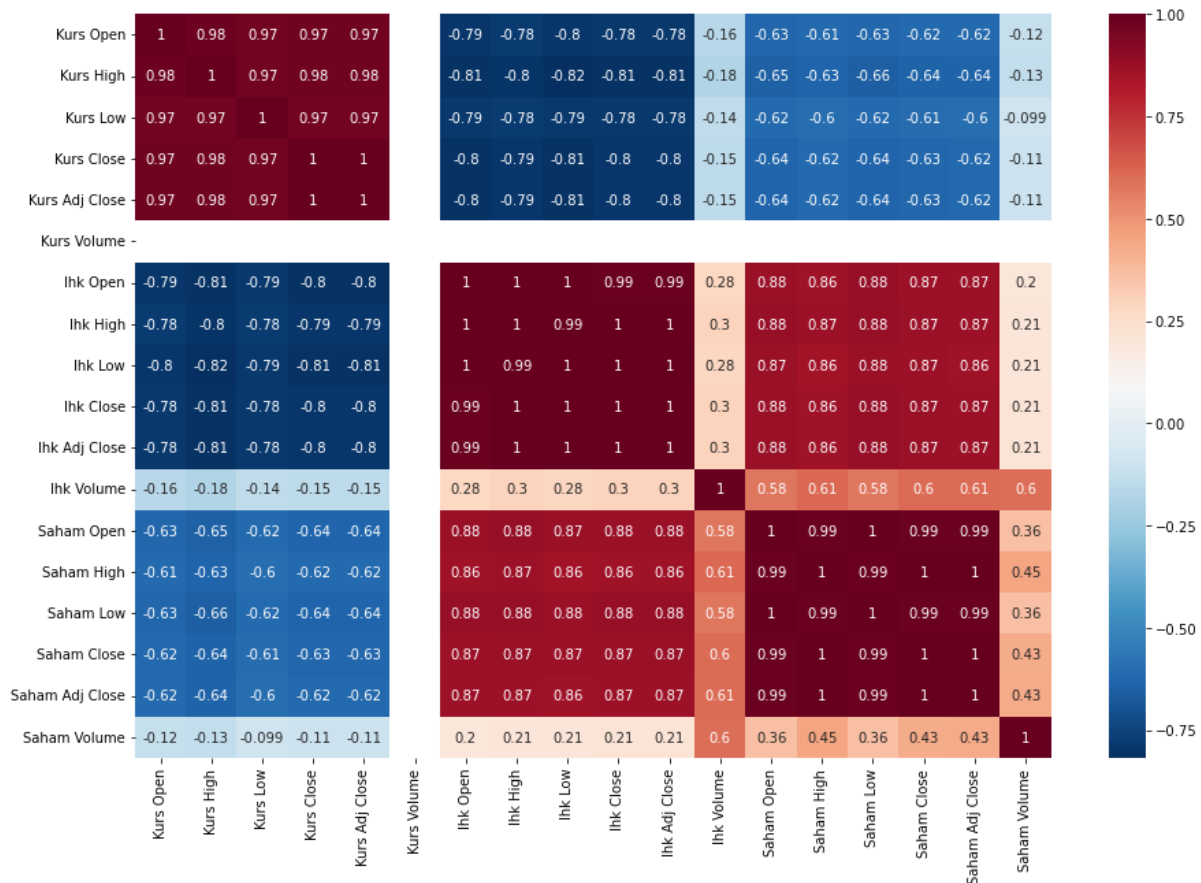
Selanjutnya dataset dipisah ke data fitur dan data prediksi. Data *output* atau prediksi, yaitu variabel dependen saham *close*. Data fitur terdiri dari *kurs open*, *kurs high*, *kurs low*, *kurs close*, *kurs adj close*, *ihk open*, *ihk high*, *ihk low*, *ihk close*, *ihk adj close*, *ihk volume*, saham *open*, saham *high*, saham *low*, saham *adj close*, dan saham *volume*.

```
pca = PCA(n_components = 2)
Xs = pca.fit_transform(Xs)
print('Explained variation per principal component: {}'.format(pca.explained_variance_ratio_ * 100))

Explained variation per principal component: [59.28917823 29.48234962]
```

Gambar 5.1.7 Skrip PCA Sebelum Pandemi COVID-19

Data fitur kemudian dilatih menggunakan metode *principal components analysis* (PCA) dengan dua komponen untuk melihat persentase informasi pada data latih, hasilnya komponen pertama menampung 59,29% informasi dan komponen kedua menampung 29,49%, jika ditotalkan jumlahnya adalah 88,78% informasi yang ada pada data latih, sisanya 11,22% data atau informasi hilang.



Gambar 5.1.8 Korelasi Pearson Selama Pandemi COVID-19

Berdasarkan nilai korelasi selama pandemi COVID-19 pada gambar 5.1.8 di atas dapat diperhatikan bahwa variabel independen *kurs open*, *kurs high*, *kurs low*, *kurs close*, dan *kurs adj close* memiliki hubungan negatif sedang terhadap variabel dependen *saham close* dengan rata-rata nilai -0,6. Sedangkan, variabel *ihk open*, *ihk high*, *ihk low*, *ihk close*, dan *ihk adj close* memiliki hubungan positif kuat dengan rata-rata nilai 0,8, tetapi variabel *ihk volume* memiliki hubungan positif sedang dengan nilai 0,6. Terakhir, hubungan antara variabel *saham open*, *saham high*, *saham low*, dan *saham adj close* memiliki hubungan positif yang sangat kuat dengan nilai korelasi satu (1), tetapi hubungan variabel *saham volume* sangat lemah dengan variabel dependen *saham close*.

```
# Data Fitur
sc1 = MinMaxScaler(feature_range=(-1,1))
Xs = sc1.fit_transform(data1[['Ihk Open', 'Ihk High', 'Ihk Low', 'Ihk Close', 'Ihk Adj Close', 'Ihk Volume',
                               'Saham Open', 'Saham High', 'Saham Low', 'Saham Adj Close', 'Saham Volume']])

# Data Prediksi
sc2 = MinMaxScaler(feature_range=(-1,1))
ys = sc2.fit_transform(data1[['Saham Close']])
```

Gambar 5.1.9 Skrip Memisah Data Fitur dan Prediksi Selama Pandemi COVID-19

```
pca = PCA(n_components = 2)
Xs = pca.fit_transform(Xs)
print('Explained variation per principal component: {}'.format(pca.explained_variance_ratio_ * 100))
```

Explained variation per principal component: [86.99218928 9.56678307]

Gambar 5.1.10 Skrip PCA Selama Pandemi COVID-19

Data fitur terdiri dari *ihk open*, *ihk high*, *ihk low*, *ihk close*, *ihk adj close*, *ihk volume*, *saham open*, *saham high*, *saham low*, *saham adj close*, dan *saham volume*. Hasil evaluasi PCA komponen pertama menampung 86,99% informasi dan komponen kedua menampung 9,57%, totalnya adalah 96,56% informasi yang ada pada data latih, sisanya 3,44% data atau informasi hilang. Setelah mendapatkan evaluasi data latih, langkah berikut melakukan pemodelan dan prediksi dengan menggunakan model yang diusulkan, hasil prediksi dapat dilihat pada gambar dan evaluasi model dapat dilihat pada tabel 5.1.6 berikut.

```
# Random Forest
regressor = RandomForestRegressor(max_depth = 2, random_state = 500)
regressor.fit(Xs, ys)
rf_pred = regressor.predict(Xs)

# Decision Tree
dst = DecisionTreeRegressor(max_depth = 2, random_state = 500)
dst.fit(Xs, ys)
dst_pred = dst.predict(Xs)
```

Gambar 5.1.11 Skrip Model Random Forest dan Decision Tree

```
window = 60
X = []
y = []
for i in range(window, len(Xs)):
    X.append(Xs[i-window:i,:])
    y.append(ys[i])

X, y = np.array(X), np.array(y)

# Long Short Term Memory
lstm = Sequential()
lstm.add(LSTM(units = 50, return_sequences = True, \
            input_shape = (X.shape[1], X.shape[2])))
lstm.add(Dropout(0.2))
lstm.add(LSTM(units = 50, return_sequences = True))
lstm.add(Dropout(0.2))
lstm.add(LSTM(units = 50))
lstm.add(Dropout(0.2))
lstm.add(Dense(units = 1))

lstm.compile(optimizer = 'adam', loss = 'mean_squared_error')
result = lstm.fit(X, y, epochs = 5, validation_split = 0.33, batch_size = 32, verbose = 0)
lstm_pred = lstm.predict(X)
```

Gambar 5.1.12 Skrip Model Long Short Term Memory



```

rnn = Sequential()
rnn.add(LSTM(units = 50, return_sequences = True, input_shape = (X.shape[1], X.shape[2])))
rnn.add(LSTM(units = 50, return_sequences = True))
rnn.add(LSTM(units = 50))
rnn.add(Dense(units = 1))

rnn.compile(optimizer = 'adam', loss = 'mean_squared_error')
result2 = rnn.fit(X, y, epochs = 5, validation_split = 0.33, batch_size = 32, verbose = 2)
rnn_pred = rnn.predict(X)

```

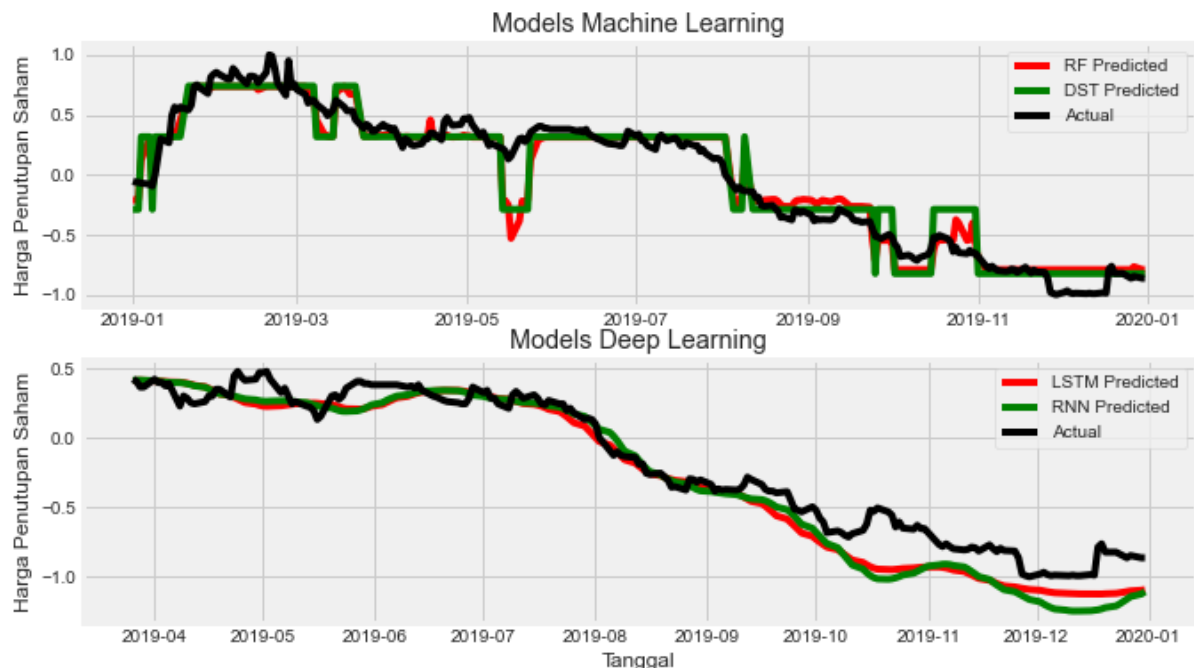
Gambar 5.1.13 Skrip Recurrent Neural Network

```

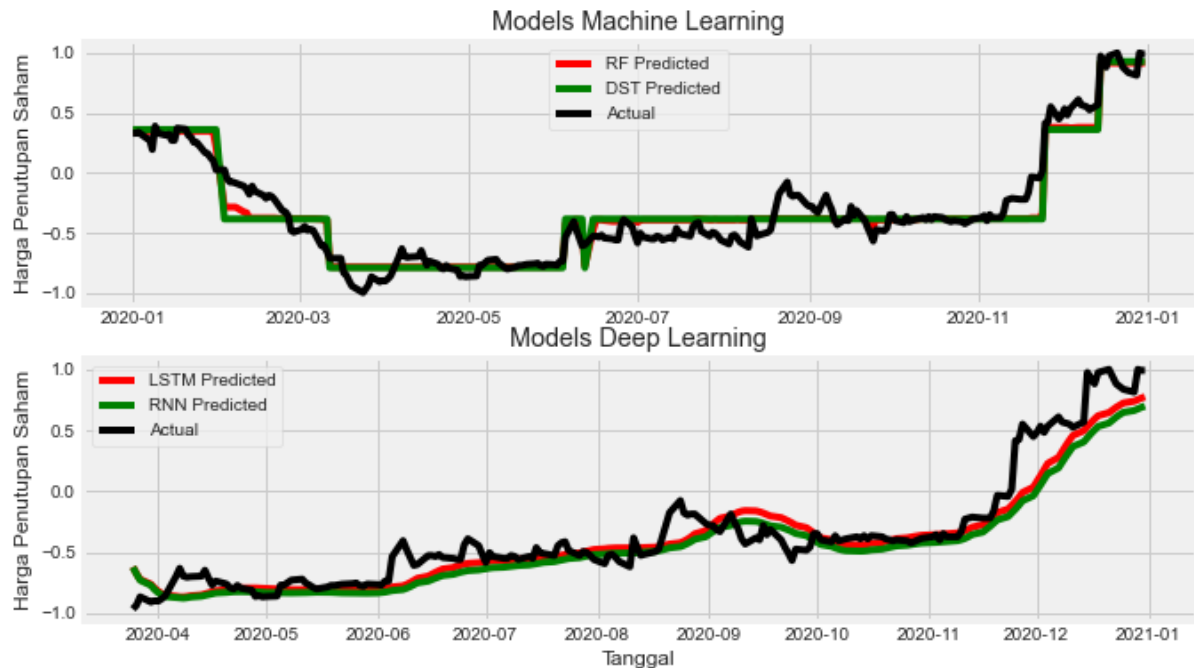
plt.figure(figsize=(10,6))
plt.subplot(2,1,1)
plt.title('Models Machine Learning')
plt.plot(data1['Date'], rf_pred,'r',label='RF Predicted')
plt.plot(data1['Date'], dst_pred,'g',label='DST Predicted')
plt.plot(data1['Date'], ys,'k',label='Actual')
plt.legend()
plt.ylabel('Harga Penutupan Saham')
plt.subplot(2,1,2)
plt.title('Models Deep Learning')
plt.plot(data1['Date'][window:], lstm_pred,'r',label='LSTM Predicted')
plt.plot(data1['Date'][window:], rnn_pred,'g',label='RNN Predicted')
plt.plot(data1['Date'][window:], y,'k',label='Actual')
plt.xlabel('Tanggal');
plt.ylabel('Harga Penutupan Saham')
plt.legend()
plt.show()

```

Gambar 5.1.14 Skrip Grafik Prediksi Model



Gambar 5.1.15 Prediksi Saham Sebelum Pandemi COVID-19



Gambar 5.1.16 Prediksi Saham Selama Pandemi COVID-19

Gambar 5.1.15 – 5.1.16 menunjukkan grafik prediksi harga penutupan saham, harga penutupan saham adalah harga yang menutup penjualan saham pada hari tersebut. Sebelum dan selama pandemi COVID-19, dapat dilihat model *random forest* dan *recurrent neural network* dapat melakukan prediksi lebih baik.

```
# Random Forest
mape1 = metrics.mean_absolute_percentage_error(ys, rf_pred)
print('Mean Absolute Error:', metrics.mean_absolute_error(ys, rf_pred))
print('Mean Absolute Percentage Error:', round(mape1, 2), '%.')
print('Mean Squared Error:', metrics.mean_squared_error(ys, rf_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(ys, rf_pred)))
print('Akurasi Model Random Forest', regressor.score(Xs, ys))

# Decision Tree
mape2 = metrics.mean_absolute_percentage_error(ys, dst_pred)
print('Mean Absolute Error:', metrics.mean_absolute_error(ys, dst_pred))
print('Mean Absolute Percentage Error:', round(mape2, 2), '%.')
print('Mean Squared Error:', metrics.mean_squared_error(ys, dst_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(ys, dst_pred)))
print('Akurasi Model Decision Tree', dst.score(Xs, ys))

# Long Short Term Memory
mape3 = metrics.mean_absolute_percentage_error(y, lstm_pred)
accuracy = 100 - np.mean(mape3)
print('Accuracy:', round(accuracy, 2), '%.')
print('Mean Absolute Error:', metrics.mean_absolute_error(y, lstm_pred))
print('Mean Absolute Percentage Error:', round(mape3, 2), '%.')
print('Mean Squared Error:', metrics.mean_squared_error(y, lstm_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y, lstm_pred)))

# Recurrent Neural Network
mape4 = metrics.mean_absolute_percentage_error(y, rnn_pred)
accuracy1 = 100 - np.mean(mape4)
print('Accuracy:', round(accuracy1, 2), '%.')
print('Mean Absolute Error:', metrics.mean_absolute_error(y, rnn_pred))
print('Mean Absolute Percentage Error:', round(mape4, 2), '%.')
print('Mean Squared Error:', metrics.mean_squared_error(y, rnn_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y, rnn_pred)))
```

Gambar 5.1.17 Skrip Evaluasi Model

Tabel 5.1.6 Hasil Evaluasi Prediksi

Periode	Model	Akurasi	RMSE	MSE	MAE	MAPE
1 Januari 2019 – 31 Desember 2019	<b>RF</b>	<b>93%</b>	<b>0,1480</b>	<b>0,0219</b>	<b>0,1074</b>	<b>1,52</b>
	DST	91%	0,1687	0,0284	0,1234	1,62
	LSTM	99,7%	0,1746	0,0305	0,1340	0,3
	RNN	99,62%	0,1877	0,0352	0,1414	0,38
1 Januari 2020 – 31 Desember 2020	<b>RF</b>	<b>93%</b>	<b>0,1221</b>	<b>0,0149</b>	<b>0,0967</b>	<b>0,54</b>
	DST	92%	0,1342	0,0180	0,1050	0,63
	LSTM	99,67%	0,1657	0,0275	0,1201	0,33
	RNN	99,6%	0,1757	0,0309	0,1339	0,4

Data tabel 5.1.6 di atas dapat dilihat hasil evaluasi sebelum pandemi COVID-19 (1 Januari 2019 – 31 Desember 2019), LSTM memperoleh skor akurasi 99,7%, tetapi hasil evaluasi RMSE 0,1746, MSE 0,0305, MAE 0,1340, dan MAPE 0,3, lebih tinggi dari DST. Sebaliknya, DST yang memperoleh akurasi 91%, tetapi hasil evaluasi RMSE 0,1687, MSE 0,0284, MAE 0,1234, dan MAPE 1,62 lebih rendah dari LSTM. Kemudian RNN memperoleh akurasi tinggi 99,6%, tetapi hasil evaluasi RMSE 0,1877, MSE 0,0352, MAE 0,1414, dan MAPE 0,38 lebih tinggi dari RF. Sedangkan, RF akurasi 93% tetapi hasil evaluasi RMSE 0,1480, MSE 0,0219, MAE 0,1074, dan MAPE 1,52 paling rendah diantara semua model.

Kemudian hasil evaluasi selama pandemi COVID-19 (1 Januari 2020 – 31 Desember 2020), LSTM memperoleh skor akurasi 99,67%, tetapi hasil evaluasi RMSE 0,1657, MSE 0,0275, MAE 0,1201, dan MAPE 0,33 lebih tinggi dari DST. Sebaliknya, DST yang memperoleh akurasi 92%, tetapi hasil evaluasi RMSE 0,1342, MSE 0,0180, MAE 0,1050, dan MAPE 0,63 lebih rendah daripada LSTM. Kemudian RNN memperoleh akurasi 99,6% tetapi hasil evaluasinya RMSE 0,1757, MSE 0,0309, MAE 0,1339, dan MAPE 0,4 lebih tinggi dibandingkan model lainnya. Sedangkan, RF akurasinya 93%, tetapi hasil evaluasinya RMSE 0,1221, MSE 0,0149, MAE 0,0967, dan MAPE 0,54 paling rendah diantara semua model.

Berdasarkan hasil evaluasi bahwa model *Random Forest* merupakan model prediksi yang mengungguli model lainnya baik itu sebelum pandemi COVID-19 dengan nilai RMSE 0,1489, MSE 0,0219, MAE 0,1074, dan MAPE 1,52 maupun selama pandemi COVID-19 dengan nilai RMSE 0,1221, MSE 0,0149, MAE 0,0967, dan MAPE 0,54.

```

sc = cross_val_score(regressor, Xs, ys, cv = 5)
print(sc)
print(sc.mean())
print("AUC: %0.2f (+/- %0.2f)" % (sc.mean(), sc.std() * 2))

[-3.13622048e-01 -8.98144159e+00  8.94834685e-03 -8.67976402e+00
-1.064444964e+01]
-5.72207514386637
AUC: -5.72 (+/- 9.20)

```

Gambar 5.1.18 K-Fold Cross Validation Random Forest Sebelum Pandemi COVID-19

```

sc1 = cross_val_score (dst, Xs, ys, cv = 5)
print(sc1)
print(sc1.mean())
print("AUC: %0.2f (+/- %0.2f)" % (sc1.mean(), sc1.std() * 2))

[-0.08548056 -9.71490047  0.11684804 -9.49802017 -11.59228675]
-6.154767980777143
AUC: -6.15 (+/- 10.18)

```

Gambar 5.1.19 K-Fold Cross Validation Decision Tree Sebelum Pandemi COVID-19

```

sc = cross_val_score(regressor, Xs, ys, cv = 5)
print(sc)
print(sc.mean())
print("AUC: %0.2f (+/- %0.2f)" % (sc.mean(), sc.std() * 2))

[ 0.76713674 -14.0950744 -2.10180567  0.07503561  0.54183429]
-2.962574689015854
AUC: -2.96 (+/- 11.32)

```

Gambar 5.1.20 K-Fold Cross Validation Random Forest Selama Pandemi COVID-19

```

sc1 = cross_val_score (dst, Xs, ys, cv = 5)
print(sc1)
print(sc1.mean())
print("AUC: %0.2f (+/- %0.2f)" % (sc1.mean(), sc1.std() * 2))

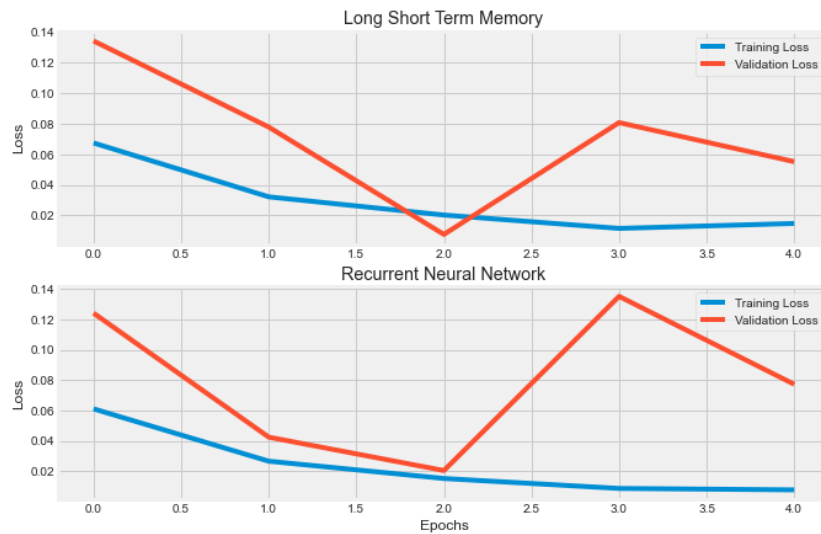
[ 0.48647994 -14.2355856 -3.24229905 -0.0799067  0.51735038]
-3.3107922069188667
AUC: -3.31 (+/- 11.27)

```

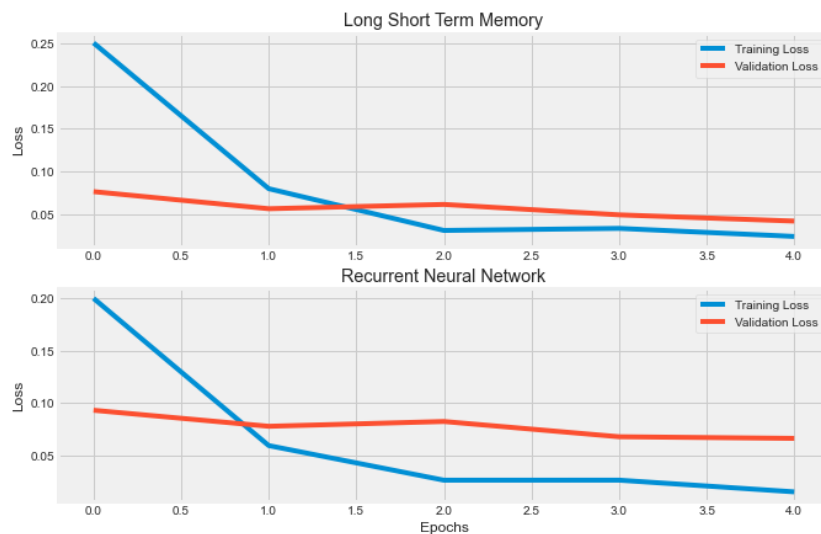
Gambar 5.1.21 K-Fold Cross Validation Decision Tree Selama Pandemi COVID-19

Selanjutnya adalah evaluasi *overfitting* model *machine learning* dengan validasi *k-fold cross validation* dengan jumlah CV sebanyak 5 (lima), artinya melakukan pengujian sebanyak lima kali dengan menggunakan semua bagian yang berbeda pada data latih. Gambar 5.1.18 dan 5.1.19 adalah hasil evaluasi *overfitting* sebelum pandemi COVID-19, model *Random Forest* memperoleh hasil akhir -5,72% dan *Decision Tree* -6,15%, hal ini mengartikan kedua model tersebut mengalami *overfitting*. Begitu pula dengan hasil evaluasi *overfitting* selama pandemi

COVID-19, kedua model tersebut mengalami *overfitting* dengan hasil akhir masing-masing -2,96% dan -3,31%.



Gambar 5.1.22 Hasil Loss Sebelum Pandemi COVID-19



Gambar 5.1.23 Hasil Loss Selama Pandemi COVID-19

Selanjutnya adalah evaluasi *overfitting* pada model *deep learning*, yaitu LSTM dan RNN. Untuk mengevaluasi *overfitting* pada model *deep learning* dapat dilakukan dengan membandingkan nilai *loss* atau nilai kerugiannya, yaitu *training loss* dan *validation loss* berdasarkan setiap *epochs*. *Loss* adalah rata-rata kesalahan untuk setiap sampel (*epochs*). Model *deep learning* dikatakan *overfitting* jika *training loss* lebih kecil daripada *validation loss*, sebaliknya dikatakan *underfitting* dan jika keduanya sama rata dikatakan *perfect fitting*. Pada gambar 5.1.22 menunjukkan hasil *loss* sebelum pandemi COVID-19, model LSTM pada *epoch* 1, 3, dan 4 mengalami *overfitting*, sedangkan *epoch* 2 mengalami *underfitting*. Kemudian,

model RNN mengalami *overfitting* dari *epoch* 1 sampai 4. Untuk hasil *loss* selama pandemi COVID-19 yang ditunjukkan pada gambar 5.1.23, model LSTM mengalami *underfitting* pada *epoch* 1, lalu *epoch* 2 dan seterusnya mengalami *overfitting*. Kemudian, model RNN dari *epochs* 1 sampai 4 mengalami *overfitting*.

## **BAB VI**

### **PENUTUP**

#### **6.1 Kesimpulan**

Berdasarkan hasil penelitian yang dilakukan, didapatkan kesimpulan sebagai berikut:

1. Model *machine learning* dan *deep learning* yang digunakan untuk memprediksi harga penutupan saham menunjukkan bahwa model *Long Short Term Memory* merupakan model terbaik berdasarkan hasil akurasi, baik sebelum dan selama pandemi COVID-19 dengan akurasi masing-masing 99,7% dan 99,67%.
2. Model *Random Forest* merupakan model terbaik berdasarkan hasil evaluasi lebih rendah dengan dengan nilai RMSE 0,1489, MSE 0,0219, MAE 0,1074, dan MAPE 1,52 sebelum pandemi COVID-19 dan selama pandemi COVID-19 dengan nilai RMSE 0,1221, MSE 0,0149, MAE 0,0967, dan MAPE 0,54.
3. Model *machine learning* dan *deep learning* yang digunakan dapat melakukan prediksi dengan akurasi tinggi, namun model-model tersebut terdeteksi mengalami *overfitting*.

## DAFTAR PUSTAKA

- Aidoo, E. N., Appiah, S. K., Awashie, G. E., Boateng, A., & Darko, G. (2021). Geographically weighted principal component analysis for characterising the spatial heterogeneity and connectivity of soil heavy metals in Kumasi, Ghana. *Heliyon*, 7(9), e08039. <https://doi.org/10.1016/j.heliyon.2021.e08039>
- Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. *Journal of King Saud University - Science*, 101940. <https://doi.org/10.1016/j.jksus.2022.101940>
- Al-Azzam, N., & Shatnawi, I. (2021). Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer. *Annals of Medicine and Surgery*, 62(November 2020), 53–64. <https://doi.org/10.1016/j.amsu.2020.12.043>
- Amalin, R. L., & Panorama, M. (2020). *Dampak COVID-19 Terhadap Tingkat Inflasi (Kelompok Pengeluaran Makanan, Minuman, Tembakau dan Kesehatan Periode 2010-2020)*. 1(September 2021), 73–84.
- Artha, A., & Paramita, R. A. S. (2021). Pengaruh Makroekonomi dan Indeks Global terhadap Indeks Harga Saham Gabungan Selama Pandemi COVID-19 di Indonesia. *Jurnal Ilmu Manajemen*, 9(2), 681. <https://doi.org/10.26740/jim.v9n2.p681-697>
- BPS, B. P. S. (2020a). *Ekonomi Indonesia Triwulan I 2020 Tumbuh 2,97 Persen*. <https://www.bps.go.id/pressrelease/2020/05/05/1736/ekonomi-indonesia-triwulan-i-2020-tumbuh-2-97-persen.html>
- BPS, B. P. S. (2020b). *Ekonomi Indonesia Triwulan II 2020 Turun 5,32 Persen*. <https://www.bps.go.id/pressrelease/2020/08/05/1737/-ekonomi-indonesia-triwulan-ii-2020-turun-5-32-persen.html>
- BPS, B. P. S. (2020c). *Ekonomi Indonesia Triwulan III 2020 Tumbuh 5,05 Persen (q-to-q)*. <https://www.bps.go.id/pressrelease/2020/11/05/1738/ekonomi-indonesia-triwulan-iii-2020-tumbuh-5-05-persen--q-to-q-.html>
- BPS, B. P. S. (2021). *Ekonomi Indonesia 2020 Turun sebesar 2,07 Persen (c-to-c)*. <https://www.bps.go.id/pressrelease/2021/02/05/1811/ekonomi-indonesia-2020-turun-sebesar-2-07-persen--c-to-c-.html>
- Chen, S., & Zhou, C. (2021). Stock Prediction Based on Genetic Algorithm Feature Selection



- and Long Short-Term Memory Neural Network. *IEEE Access*, 9, 9066–9072. <https://doi.org/10.1109/ACCESS.2020.3047109>
- Gao, S., & Lima, D. (2022). A review of the application of deep learning in the detection of Alzheimer's disease. *International Journal of Cognitive Computing in Engineering*, 3(October 2021), 1–8. <https://doi.org/10.1016/j.ijcce.2021.12.002>
- Haryanto. (2020). Dampak Covid-19 terhadap Pergerakan Nilai Tukar Rupiah dan Indeks Harga Saham Gabungan (IHSG). *Jurnal Perencanaan Pembangunan: The Indonesian Journal of Development Planning*, 4(2), 151–165. <https://doi.org/10.36574/jpp.v4i2.114>
- Hastuti, P. (2020). Fenomena Kurs Rupiah Sebelum Dan Selama Covid-19. *Niagawan*, 9(3), 197. <https://doi.org/10.24114/niaga.v9i3.18936>
- Irianto, Kisnawati, B., & Istiarto. (2021). Perbandingan Harga Saham Sebelum dan Selama Covid-19. *Jurnal Kompetitif: Media Informasi Ekonomi Pembangunan, Manajemen Dan Akuntansi*, 7(1), 1–16.
- Kamari, A., & Peter Leslie Schultz, C. (2022). A combined principal component analysis and clustering approach for exploring enormous renovation design spaces. *Journal of Building Engineering*, 48(January), 103971. <https://doi.org/10.1016/j.jobbe.2021.103971>
- Khalimi, A. M. (2022a). *Cara Hitung RMSE, MSE, MAPE, dan MAE Dengan Excel*. <https://www.pengalaman-edukasi.com/2021/01/cara-menghitung-rmse-root-mean-square.html>
- Khalimi, A. M. (2022b). *Pengujian Data dengan Cross Validation*. <https://www.pengalaman-edukasi.com/2020/04/apa-itu-k-fold-cross-validation.html>
- Khang, P. Q., Kaczmarczyk, K., Tutak, P., Golec, P., Kuziak, K., Depczynski, R., Hernes, M., & Rot, A. (2021). Machine learning for liquidity prediction on Vietnamese stock market. *Procedia Computer Science*, 192, 3590–3597. <https://doi.org/10.1016/j.procs.2021.09.132>
- Khoiri. (2020a). *Cara Menghitung Mean Absolute Percentage Error (MAPE)*. <https://www.khoiri.com/2020/12/pengertian-dan-cara-menghitung-mean-absolute-percentage-error-mape.html>
- Khoiri. (2020b). *Cara Menghitung Mean Squared Error (MSE)*. <https://www.khoiri.com/2020/12/pengertian-dan-cara-menghitung-mean-squared-error-mse.html>
- Khoiri. (2020c). *Pengertian dan Cara Menghitung Root Mean Square Error (RMSE)*. <https://www.khoiri.com/2020/12/cara-menghitung-root-mean-square-error-rmse.html>

- Kilimci, Z. H., & Duvar, R. (2020). An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using twitter and financial news sites: A case of istanbul stock exchange (BIST 100). *IEEE Access*, 8(Bist 100), 188186–188198. <https://doi.org/10.1109/ACCESS.2020.3029860>
- Lathifah, H. M., Febrianti, D. S., Utami, A. P., Ulhaq, A. A., Tulasmi, T., & Mukti, T. (2021). Dampak Pandemi Covid-19 Terhadap Nilai Harga Saham Syariah di Indonesia. *Jurnal Ilmiah Ekonomi Islam*, 7(1), 223. <https://doi.org/10.29040/jiei.v7i1.1772>
- Ma, Z., & Mei, G. (2021). Deep learning for geological hazards analysis: Data, models, applications, and opportunities. *Earth-Science Reviews*, 223, 103858. <https://doi.org/10.1016/j.earscirev.2021.103858>
- Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science*, 170, 1168–1173. <https://doi.org/10.1016/j.procs.2020.03.049>
- Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., & Mosavi, A. (2020). Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis. *IEEE Access*, 8, 150199–150212. <https://doi.org/10.1109/ACCESS.2020.3015966>
- Nurmasari, I. (2020). Dampak Covid-19 Terhadap Perubahan Harga Saham dan Volume Transaksi (Studi Kasus Pada PT. Ramayana Lestari Sentosa, Tbk.). *Jurnal SEKURITAS (Saham, Ekonomi, Keuangan Dan Investasi)*, 3(3), 230. <https://doi.org/10.32493/skt.v3i3.5022>
- Pinelis, M., & Ruppert, D. (2022). Machine learning portfolio allocation. *Journal of Finance and Data Science*, 8, 35–54. <https://doi.org/10.1016/j.jfds.2021.12.001>
- Rather, A. M. (2021). LSTM-based Deep Learning Model for Stock Prediction and Predictive Optimization Model. *EURO Journal on Decision Processes*, 9(May), 100001. <https://doi.org/10.1016/j.ejdp.2021.100001>
- Rusyida, W. Y., & Pratama, V. Y. (2020). Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA. *Square : Journal of Mathematics and Mathematics Education*, 2(1), 73. <https://doi.org/10.21580/square.2020.2.1.5626>
- Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *Procedia Computer Science*, 167(2019), 788–798. <https://doi.org/10.1016/j.procs.2020.03.419>
- Sitorus, C. M. (2020). *NK-Fold Cross Validation Secara Singkat*.

- <https://cmemorys.medium.com/k-fold-cross-validation-secara-singkat-30f8e5188f46>
- Song, D., Chung Baek, A. M., & Kim, N. (2021). Forecasting Stock Market Indices Using Padding-Based Fourier Transform Denoising and Time Series Deep Learning Models. *IEEE Access*, 9, 83786–83796. <https://doi.org/10.1109/ACCESS.2021.3086537>
- Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock Market Prediction Using Machine Learning. *Procedia Computer Science*, 194, 173–179. <https://doi.org/10.1016/j.procs.2021.10.071>
- Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, 167(2019), 599–606. <https://doi.org/10.1016/j.procs.2020.03.326>
- Wang, Q., Zhou, W., Yang, L., & Huang, K. (2022). Comparison between conventional and deep learning-based surrogate models in predicting convective heat transfer performance of U-bend channels. *Energy and AI*, 8(November 2021), 100140. <https://doi.org/10.1016/j.egyai.2022.100140>
- Wen, M., Li, P., Zhang, L., & Chen, Y. (2019). Stock market trend prediction using high-order information of time series. *IEEE Access*, 7, 28299–28308. <https://doi.org/10.1109/ACCESS.2019.2901842>
- Zhu, L., Zhou, X., & Zhang, C. (2021). Rapid identification of high-quality marine shale gas reservoirs based on the oversampling method and random forest algorithm. *Artificial Intelligence in Geosciences*, 2(November), 76–81. <https://doi.org/10.1016/j.aiig.2021.12.001>