

**ANALISA PREDIKSI NILAI SAHAM BANK DIGITAL INDONESIA  
DENGAN MENGGUNAKAN ALGORITMA *SUPPORT VECTOR*  
*MACHINE***

**SKRIPSI**

Disusun Untuk Memenuhi Persyaratan Guna Memperoleh Sarjana Komputer

Oleh:

**RAAFI HAYKAL BIMANTORO**

**1914000017**



**PERBANAS  
INSTITUTE**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**INSTITUT KEUANGAN PERBANKAN DAN INFORMATIKA ASIA**

**(ASIAN BANKING FINANCE AND INFORMATICS INSTITUTE)**

**PERBANAS**

**JAKARTA**

**2023**

**INSTITUT KEUANGAN PERBANKAN DAN  
INFORMATIKA ASIA PERBANAS  
JAKARTA  
PROGRAM STUDI INFORMATIKA**

**LEMBAR PERSETUJUAN**

Skripsi yang berjudul

**ANALISA PREDIKSI NILAI SAHAM BANK DIGITAL INDONESIA  
DENGAN MENGGUNAKAN ALGORITMA *SUPPORT VECTOR  
MACHINE***



Oleh

Nama : RAAFI HAYKAL BIMANTORO

NIM : 191400017

Program Studi : Teknik Informatika

Telah disetujui dan diujikan

Jakarta, 27 Juli 2023	
Mengetahui,	
Ketua Program Studi Teknik Informatika	Dosen Pembimbing Skripsi
	
Winny Purbaratri S.Kom., M.Kom.	Prof. Dr. Harya Damar Widiputra, S.T., M.Kom.



**INSTITUT KEUANGAN PERBANKAN DAN  
INFORMATIKA ASIA PERBANAS  
JAKARTA  
PROGRAM STUDI INFORMATIKA**

**LEMBAR PERNYATAAN**

Seluruh isi dan materi skripsi ini menjadi tanggung jawab penyusun sepenuhnya.

Jakarta, 27 Juli 2023

Penulis

A handwritten signature in black ink, appearing to read "Raafi Haykal Bimantoro", is positioned above the printed name.

Raafi Haykal Bimantoro  
(1914000017)



**INSTITUT KEUANGAN PERBANKAN DAN  
INFORMATIKA ASIA PERBANAS  
JAKARTA  
PROGRAM STUDI INFORMATIKA**

**LEMBAR PERNYATAAN BEBAS PLAGIAT**

Yang bertanda tangan dibawah ini,

Nama : Raafi Haykal Bimantoro

NIM : 1914000017

Program Studi : Teknik Informatika

Judul Skripsi : **ANALISA PREDIKSI NILAI SAHAM  
BANK DIGITAL INDONESIA  
DENGAN MENGGUNAKAN  
ALGORITMA *SUPPORT VECTOR  
MACHINE***

Menyatakan bahwa hasil penulisan Skripsi yang telah saya buat merupakan hasil karya sendiri dan benar keasliannya. Apabila ternyata dikemudian hari penulisan skripsi ini merupakan hasil plagiat atau penjiplakan terhadap karya orang lain, saya bersedia mempertanggung jawabkan sekaligus menerima sanksi berdasarkan aturan tata tertib di ABFI Institute Perbanas Jakarta.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak ada unsur paksaan

Jakarta, 27 Juli 2023

Penulis

A handwritten signature in black ink, appearing to be "Raafi Haykal Bimantoro", written over a light blue grid background.

Raafi Haykal Bimantoro  
(1914000017)

## HALAMAN DOKUMENTASI

(A) <b>NAMA</b> Raafi Haykal Bimantoro	<b>NIM</b> 1914000017
(B) <b>JUDUL SKRIPSI</b> “ <b>ANALISA PREDIKSI NILAI SAHAM BANK DIGITAL INDONESIA DENGAN MENGGUNAKAN ALGORITMA <i>SUPPORT VECTOR MACHINE</i></b> ”	
(C) <b>HALAMAN, TAHUN, GAMBAR, TABEL, LAMPIRAN</b> + <b>70</b> Halaman, <b>2023</b> , <b>8</b> Gambar, <b>15</b> Tabel, <b>0</b> Lampiran	
(D) <b>KATA KUNCI</b> <i>Support Vector Machine (SVM), Supervised Learning, Prediksi Saham, Regresi, Data Mining, Machine Learning</i>	
(E) <b>INTISARI</b> Skripsi dengan judul <b>ANALISA PREDIKSI NILAI SAHAM BANK DIGITAL INDONESIA DENGAN MENGGUNAKAN ALGORITMA <i>SUPPORT VECTOR MACHINE</i></b> merupakan penelitian yang bertujuan untuk melakukan prediksi dari harga saham Allo Bank dan Bank Jago menggunakan algoritma Support Vector Machine (SVM) sehingga dapat diketahui performa dari algoritma tersebut.	
(F) <b>DAFTAR ACUAN</b> 7 Jurnal (1959-2021), 11 <i>website</i> , 6 <i>E-book</i>	
(G) <b>DOSEN PEMBIMBING</b> Prof. Dr. Harya Damar Widiputra, S.T., M.Kom.	

## ABSTRAKSI

Saham merupakan salah satu instrumen dalam berinvestasi yang tersedia untuk masyarakat, terutama bagi masyarakat Indonesia. Sebagaimana komoditas yang diperjual belikan, saham juga mengalami kenaikan dan penurunan harga (fluktuasi) yang menciptakan untung rugi kepada investor. Untuk menghindari kerugian, investor harus mengetahui saham mana yang dapat menjanjikan keuntungan. Namun, hal tersebut tidak terjadi kepada masyarakat Indonesia yang pada tahun 2020 mengalami kerugian besar karena tidak memiliki pengetahuan yang cukup untuk berinvestasi saham. Untuk membantu dan mencegah terulangnya kerugian tersebut, diadakannya alat untuk memprediksi harga saham dengan menggunakan *artificial intelligence*. Dalam penelitian ini, akan diprediksi harga saham Allo Bank dan Bank Jago dengan menggunakan algoritma *Support Vector Machine* dengan *kernel linear*, *polynomial*, RBF, dan *gaussian*, serta dengan menggunakan algoritma KNN sebagai parameter untuk membandingkan performa yang dihasilkan oleh algoritma SVM. Pengujian kedua algoritma untuk memprediksi harga saham dilakukan sebanyak sepuluh kali percobaan, dan hasil yang diperoleh dari percobaan tersebut diakumulasi kemudian dirata-rata. Hasil yang diperoleh dalam memprediksi harga saham adalah berupa peringkat dari performa kedua algoritma yaitu SVM *kernel* RBF, KNN, dan SVM *kernel linear*, SVM *kernel polynomial*, dan SVM *kernel sigmoid*.

Kata kunci: *Support Vector Machine*, *K-nearest Neighbor*, Prediksi, Harga Saham

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Subhanahu Wata'ala yang telah memberikat petunju, Rahmat, dan nikmat-Nya kepada penulis sehingga penulis bisa menyelesaikan skripsi yang berjudul **“ANALISA PREDIKSI NILAI SAHAM BANK DIGITAL INDONESIA DENGAN MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*”** sebagai salah satu syarat kelulusan dan memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika di IKPIA Perbanas.

Penulis mendedikasikan penelitian ini kepada kedua orang tua dan kepada eyang kakung serta eyang putri yang telah membantu seluruh proses pendidikan yang penulis lalui, serta sebagai sosok yang selalu memberikan doa dan dukungan fisik maupun moral kepada penulis sampai pada penulis berhasil menyelesaikan penelitian ini.

Penulis mengucapkan terima kasih kepada Bapak Prof., Dr. Harya Damar Widiputra, S.T., M.Kom., selaku Wakil Rektor sekaligus Dosen Pembimbing Skripsi yang telah memberikan arahan yang mudah dipahami serta solusi yang efektif dalam mengatasi rintangan dalam penyusunan skripsi ini, Ibu Lucia Sri Istiyowati, M.Kom., selaku Dekan Fakultas Teknologi Informasi, Ibu Winny Purbaratri S.Kom, M.Kom., selaku Ketua Program Studi Teknik Informatika Institut Perbanas, serta para Dosen dan Staf IKPIA PERBANAS yang telah memberikan bantuan kepada penulis dalam penyelesaian skripsi.

Penulis juga mengucapkan terima kasih kepada teman teman Teknik Informatika Angkatan 2019 yang saling menyemangati dalam proses penulisan dan penelitian skripsi ini. Terima kasih yang terakhir ditujukan kepada Deva Alviana Larasati S.Li. untuk seluruh diskusi dan perjuangannya serta kepada seluruh pihak yang telah membantu namun tidak bisa disebutkan satu persatu.

Akhir kata, penulis berharap penelitian ini dapat membantu dan memberikan manfaat bagi pembacanya.

Jakarta, 27 Juli 2023

Penulis



Raafi Haykal Bimantoro

## DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PERNYATAAN .....	ii
LEMBAR PERNYATAAN BEBAS PLAGIAT.....	iii
HALAMAN DOKUMENTASI .....	iv
ABSTRAKSI .....	v
KATA PENGANTAR.....	vi
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	x
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN .....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah .....	4
1.3    Identifikasi Masalah .....	4
1.4    Batasan Masalah.....	4
1.5    Tujuan Penelitian.....	5
1.6    Manfaat Penelitian.....	5
BAB II LANDASAN TEORI .....	7
2.1.    Tinjauan Studi .....	7
2.1.1.    Saham.....	7
2.1.1.1.    Transaksi Jual Beli Saham .....	8
2.1.2. <i>Machine Learning</i> .....	9
2.1.3. <i>Supervised Learning</i> .....	10
2.1.4. <i>Support Vector Machine (SVM)</i> .....	11
2.1.4.1.    Kernel SVM.....	12



2.1.5.	<i>K-Nearest Neighbors (KNN)</i> .....	12
2.1.6.	Regresi.....	14
2.1.7.	<i>Machine Learning Evaluation Matrics</i> .....	14
2.1.7.1.	<i>Mean Square Error (MSE)</i> .....	15
2.1.7.2.	<i>Mean Absolute Error (MAE)</i> .....	15
2.1.8.	Google Colaboratory .....	16
2.1.9.	Python .....	16
2.1.10.	SKLearn.....	17
2.1.11.	Pandas.....	17
2.1.12.	Matplotlib.....	18
2.2.	Tinjauan Literatur .....	18
2.3.	Kerangka Penelitian .....	27
BAB III METODE PENELITIAN.....		30
3.1.	Tahapan Penelitian .....	32
3.2.	Sumber Data .....	37
3.3.	Objek Penelitian .....	38
3.3.1.	Data Saham Allo Bank.....	38
3.3.2.	Data Saham Bank Jago.....	39
3.4.	Instrumen Penelitian .....	41
3.5.	Metode Analisa Data .....	41
BAB IV HASIL DAN PEMBAHASAN.....		43
4.1.	Identifikasi Model <i>Machine Learning</i> .....	43
4.2.	Pengujian Model <i>Machine Learning</i> .....	44
4.2.1.	Pengumpulan Data .....	44
4.2.2.	<i>Pre-Processing Data</i> .....	46

4.2.2.1.	<i>Data Cleaning</i> .....	47
4.2.2.2.	<i>Data Manipulating</i> .....	49
4.2.2.3.	<i>Data Splitting</i> .....	53
4.2.2.4.	Normalisasi Data.....	55
4.2.3.	Implementasi SVM & KNN Serta Memilih <i>Kernel Function</i> .....	56
4.2.4.	Menyiapkan <i>Multioutput Machine Learning Model</i> .....	57
4.2.5.	Melatih Model <i>Machine Learning</i> .....	57
4.2.6.	Melakukan Prediksi.....	58
4.3.	Analisa Hasil .....	59
4.3.1.	Perbandingan Hasil Prediksi SVM Linear Dengan <i>Data Test</i> .....	59
4.3.2.	Perbandingan Masing-Masing <i>Kernel Function</i> SVM .....	60
4.3.3.	Perbandingan Algoritma SVM Dengan KNN.....	62
4.4.	Evaluasi Dan Penarikan Kesimpulan .....	63
BAB V PENUTUP.....		66
5.1.	Kesimpulan.....	66
5.2.	Saran .....	66
DAFTAR PUSTAKA .....		68

## DAFTAR GAMBAR

Gambar 2.1 Grafik Fluktuasi Harga Saham. ....	9
Gambar 2.2 Proses dari Algoritma <i>Supervised Learning</i> .....	10
Gambar 2.3 <i>Support Vector Machine Hyperplane</i> . ....	12
Gambar 2.4 Proses Pengelompokan Algoritma KNN.....	13
Gambar 3.1 <i>Flowchart</i> Metode DSR. ....	31
Gambar 3.2 <i>Flowchart</i> Tahapan Penelitian.....	32
Gambar 3.3 Data Harga Saham Allo Bank. ....	38
Gambar 4.1 Tampilan <i>Website</i> Kaggle.com .....	45

## DAFTAR TABEL

Tabel 2.1 Tinjauan Literatur .....	18
Tabel 2.2 Kerangka Penelitian .....	28
Tabel 3.1 Data Deskriptif Harga Saham Allo Bank .....	39
Tabel 4.1 Tabel <i>data field</i> dalam <i>dataset</i> harga saham.....	45
Tabel 4.2 <i>Dataset</i> Harga Saham ARTO Sebelum <i>Cleaning</i> .....	48
Tabel 4.3 <i>Dataset</i> Harga Saham ARTO Setelah <i>Cleaning</i> .....	48
Tabel 4.4 Data Sebelum Normalisasi Dari Variabel X_train .....	55
Tabel 4.5 Data Setelah Normalisasi Dari Variabel X_train.....	56
Tabel 4.6 Perolehan MSE dan MAE dari SVM <i>Kernel</i> Linear. ....	59
Tabel 4.7 Perolehan MSE dan MAE dari SVM <i>Kernel Polynomial</i> . ....	60
Tabel 4.8 Perolehan MSE dan MAE dari SVM <i>Kernel</i> Gaussian/RBF. ....	61
Tabel 4.9 Perolehan MSE dan MAE dari SVM <i>Kernel</i> Sigmoid.....	62
Tabel 4.10 Perolehan Nilai MSE dan MAE dari Algoritma KNN. ....	63
Tabel 4.11 Perbandingan MSE KNN dengan <i>Kernel</i> SVM. ....	64
Tabel 4.12 Perbandingan MAE KNN dengan <i>Kernel</i> SVM. ....	64

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Menurut Reilly dan Norton dalam bukunya yang berjudul *Investments* yang terbit pada tahun 2007, definisi investasi adalah suatu komitmen untuk menanamkan dana pada periode tertentu untuk mendapatkan pembayaran di masa depan sebagai kompensasi bagi investor. Investasi banyak dilakukan oleh masyarakat dunia sebagai cara untuk mengelola keuangan pribadi maupun badan perusahaan. Berdasarkan definisi investasi yang tertera, khalayak umum dapat mengetahui tujuan individu melakukan investasi yaitu mendapatkan keuntungan. Tetapi, tidak menutup kemungkinan bahwa akan terjadi kerugian. Individu dapat mengatakan investasinya mengalami keuntungan apabila harga beli dari instrumen investasi lebih murah dibandingkan dari harga jualnya, dan sebaliknya. Untung dan rugi tersebut merupakan hal yang wajar dalam berinvestasi, yang dimana untung dan rugi tersebut dipengaruhi oleh banyak macam faktor. Hal tersebut melahirkan kesimpulan bahwa, untuk melakukan investasi, perlu adanya strategi yang matang. Untuk dapat melakukan investasi, individu perlu mengenal dan paham tentang instrumen apa yang akan dia jadikan sebagai tempat berinvestasi. Instrumen investasi tersebut beragam, diantaranya yang populer di masyarakat adalah emas, properti, reksa dana, dan saham.

Saham merupakan salah satu instrumen investasi yang sudah dikenal oleh masyarakat. Saham sendiri memiliki definisi sebagai suatu tanda bukti kepemilikan seseorang atau badan atas suatu korporasi merupakan definisi dari Darmadji dan Fakhruddin pada bukunya yang berjudul *Pasar Modal di Indonesia* yang terbit pada tahun 2006. Proses jual beli saham terjadi di pasar modal yang berjalan setiap harinya. Dimana proses tersebut dimulai dengan sebuah harga awal yang ditentukan oleh badan korporasi yang akan menjual sahamnya, harga awal tersebut tidak akan terlalu jauh dari harga saham pada

satu hari sebelumnya. Lalu investor akan melihat harga awal tersebut kemudian memutuskan apakah akan membeli saham tersebut di harga yang telah ditentukan sebelumnya atau ingin melakukan penawaran dengan harga baru. Selanjutnya transaksi baru akan tercatat apabila harga jual yang diajukan oleh korporasi tersebut menemukan calon investor yang mengajukan harga beli yang sama. Dengan terjadinya transaksi, harga saham dapat mengalami fluktuasi tergantung dari transaksi yang terjadi. Fluktuasi tersebut juga memiliki faktor diluar dari transaksi jual beli. Faktor yang dimaksud adalah; *Earning Per Share*, *Price Earning Ratio*, *Debt to Equity Ratio*, dan *Market Value Added* (Umayah, R., et al, 2019). Beberapa faktor sebelumnya juga ikut mengambil peran dalam terjadinya proses transaksi jual beli saham tersebut, sesuai dengan tujuan investasi yaitu untuk mendapatkan keuntungan dalam jangka waktu tertentu.

Pada tahun 2020 sampai dengan tahun 2021 silam, investasi khususnya pada instrumen saham mengalami kenaikan minat, sehingga menjadi populer di kalangan masyarakat Indonesia. Faktor yang membuat hal tersebut terjadi dan faktor paling jelas yaitu dengan adanya Pandemi COVID-19 yang melanda dunia mengakibatkan limitasi dari mobilitas masyarakat. Dengan terbatasnya mobilitas tersebut, membuat masyarakat memiliki banyak waktu luang sehingga dimanfaatkan oleh sebagian besar kalangan untuk mencari kegiatan baru, yaitu dengan melakukan investasi. Hal tersebut juga didorong dengan adanya iklan pada media sosial tentang aplikasi *trading* saham yang memudahkan masyarakat umum untuk melakukan investasi melalui *trading*. Aplikasi yang mengiklankan dirinya tersebut juga menghadirkan *Brand Ambassador* dan *Affiliator* sebagai bagian dari bentuk promosinya tersebut. Efek samping yang muncul dari investasi yang telah dilakukan oleh masyarakat dapat dibilang mengenaskan. Karena awamnya wawasan masyarakat terkait prosedur investasi, mengakibatkan kerugian yang cukup besar. Ditambah aplikasi *trading* tersebut dinyatakan sebagai aplikasi ilegal oleh Otoritas Jasa Keuangan (OJK, 2020). Lalu, *Brand Ambassador* dan *Affiliator* dari aplikasi

tersebut dijadikan sebagai tersangka karena telah menipu sebanyak 144 korban dengan total kerugian mencapai Rp.83 Miliar (Nufus, W.H., 2022).

Tidak ada orang yang menginginkan investasinya mengalami kerugian. Untuk itu, para calon investor harus dibekali dengan ilmu yang mumpuni, mampu mempunyai strategi sehingga dapat mencegah kegagalan seminimal mungkin, dan mempunyai wawasan mengenai alat-alat yang dapat membantu mereka demi mendapatkan keuntungan dalam investasi. Salah satu alat yang dapat digunakan oleh para calon investor adalah algoritma *machine learning*. *Machine learning* dapat diaplikasikan sebagai alat untuk melakukan prediksi dari harga saham yang akan dilakukannya investasi oleh calon investor. *Machine learning* bekerja dengan cara mempelajari data-data saham terdahulu dan melihatnya sebagai sebuah pola, kemudian pola tersebut akan dijadikan sebuah model prediksi. Model prediksi harga saham yang terbilang akurat dapat membantu para calon investor dalam mempertimbangkan pengambilan keputusan transaksi saham karena pergerakan harga saham yang tidak dapat ditebak ini akan menyulitkan investor dalam melakukan prediksi. Algoritma yang akan digunakan dalam penelitian ini adalah *Support Vector Machine*. Dimana algoritma ini dipilih adalah karena performa dan waktu proses yang cukup baik dalam kategorinya.

Dalam penelitian ini, akan dilakukannya analisa dengan menjadikan saham dari dua bank digital yang ada di Indonesia, antara lain Bank Jago dan Allo Bank. Dipilih nya bank digital tersebut dikarenakan kedua perusahaan tersebut berhasil mendapatkan nasabah secara masal dengan program promosinya pada tahun 2021 silam (cari referensinya). Parameter analisa yang digunakan adalah RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*). RMSE sendiri merupakan selisih antara nilai sebenarnya dengan nilai prediksi. Sedangkan MAE merupakan nilai kesalahan rata-rata yang eror dari nilai sebenarnya dengan nilai prediksi. Penulis berharap, dengan adanya penelitian ini khalayak dapat menjadi calon investor yang cerdas dengan memanfaatkan salah satu algoritma *machine learning* yang dipercaya dalam melakukan

prediksi pergerakan nilai saham secara mandiri sebelum melakukan investasi.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang ada, berikut adalah rumusan masalah yang terdapat dalam penelitian:

1. Bagaimana cara menghindari terjadinya kerugian yang dialami oleh masyarakat yang antusias dan memiliki ketertarikan tinggi dalam melakukan investasi saham, tetapi tidak diimbangi dengan pengetahuan berinvestasi yang mumpuni baik secara teori maupun praktik?
2. Apakah model *machine learning* merupakan sebuah alternatif yang dapat dipercaya oleh masyarakat yang ingin berinvestasi dalam memprediksi harga saham ?

## **1.3 Identifikasi Masalah**

Berikut merupakan identifikasi masalah yang dapat disimpulkan, yaitu;

1. Perlu diketahuinya algoritma machine learning yang sesuai untuk memprediksi harga saham.
2. Bagaimana kinerja algoritma SVM dan KNN dalam memprediksi harga saham.
3. Menemukan konfigurasi algoritma yang paling sesuai dalam memprediksi harga saham.
4. Membandingkan dan menganalisa kedua algoritma tersebut untuk mengetahui manakah algoritma yang lebih sesuai digunakan dalam memprediksi harga saham.

## **1.4 Batasan Masalah**

Di dalam penulisan ini, penulis melakukan pembatasan dari bahasan masalah yang akan dilakukan penelitian. Bahasan masalah tersebut yaitu:



1. Penulis hanya akan menggunakan algoritma *Support Vector Machine* dan algoritma *K-Nearest Neighbour* sebagai metode untuk memprediksi nilai saham data harga saham yang digunakan yaitu data saham Allo Bank dan Bank Jago.
2. Satuan harga yang digunakan dalam penelitian ini sebagai variabel terikat yaitu *data field close*.
3. Penelitian ini menggunakan model algoritma prediksi yang didapatkan melalui *data training* masing-masing objek dikarenakan perbedaan data yang akan diprediksi.
4. Hasil prediksi yang diperoleh yaitu 3 hari yang akan datang, dan hasil prediksi yang diperoleh berupa nilai normalisasi.
5. Hasil analisa yaitu berupa hasil perhitungan *Mean Square Error* dan *Mean Absolute Error*.

### **1.5 Tujuan Penelitian**

Berikut adalah tujuan dari dilakukannya penelitian ini:

1. Ditemukannya algoritma yang dapat diimplementasikan dalam memprediksi harga saham
2. Mendapatkan data performa dari hasil prediksi harga saham menggunakan algoritma SVM dan KNN
3. Mendapatkan konfigurasi algoritma yang paling sesuai dalam memprediksi harga saham.

### **1.6 Manfaat Penelitian**

Manfaat penelitian itu terbagi menjadi dua, yaitu manfaat bagi penelitian dan manfaat segi keilmuan. Berikut merupakan paparan dari keduanya:

1. Manfaat bagi penelitian  
Sisi teknis dalam penelitian ini diharapkan bisa menjadi landasan dari pengembangan *tools* yang lebih mumpuni dalam mempermudah khalayak umum dalam melakukan investasi.

2. Manfaat segi keilmuan:

Penelitian ini diharapkan dapat memperoleh informasi atau pengetahuan dari kinerja kedua algoritma *machine learning* dalam penerapannya sebagai *tools* untuk memprediksi harga saham dan menjadi literatur yang menghadirkan alternatif bagi khalayak umum dalam berinvestasi pada saham.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Tinjauan Studi**

Penelitian ini terdiri dari beberapa unsur yang memerlukan pembahasan lebih lanjut sehingga dapat dimengerti oleh khalayak umum. Dalam subbab ini akan dipaparkan oleh penulis secara mendalam terkait unsur-unsur tersebut sehingga proses penelitian dapat diteliti secara sistematis serta isi penelitian dapat dimengerti secara terarah.

##### **2.1.1. Saham**

Definisi saham menurut Darmadji dan Fakhrudin dalam bukunya yang berjudul Pasar Modal di Indonesia pada tahun 2012 yaitu adalah tanda penyertaan atau kepemilikan seseorang atau badan dalam suatu perusahaan atau perseroan terbatas (Darmadhi & Fakhrudin, 2012). Proses transaksi jual beli saham dilakukan melalui pasar modal. Langkah awal bagi individu agar dapat melakukan proses transaksi saham yaitu, individu tersebut harus sudah dapat menentukan saham kepercayaannya untuk berinvestasi. Saham-saham yang dimuat dalam pasar modal merupakan saham milik perusahaan yang mana perusahaan tersebut telah memutuskan untuk membuka kepemilikan sahamnya kepada khalayak umum. Dengan dibukanya saham milik perusahaan kepada khalayak umum, hal ini dapat berdampak positif bagi kedua belah pihak, diantaranya masyarakat dan perusahaan itu sendiri. Hal tersebut dapat terjadi karena masyarakat mempunyai opsi lain untuk menabung yaitu melalui investasi saham di pasar modal. Perusahaan pemilik saham juga mendapatkan kemudahan untuk memenuhi kebutuhan investor secara langsung. Dalam melakukan transaksi saham tentu terdapat prosedur, syarat, dan ketentuan yang berlaku. Beberapa unsur ini akan dijelaskan dalam subbab selanjutnya.

#### 2.1.1.1. Transaksi Jual Beli Saham

Dalam proses transaksi jual beli saham, terdapat beberapa kosa kata dalam satuan harga saham yang harus dipahami terlebih dahulu. Dalam subbab ini dipaparkan kosa kata tersebut diantaranya;

- *Open* (harga pembukaan)  
*Open* adalah harga yang terjadi pada transaksi pertama suatu saham.
- *Close* (harga penutupan/final)  
*Close* adalah harga yang terjadi pada transaksi terakhir suatu saham.
- *High* (harga tertinggi)  
*High* adalah harga tertinggi transaksi yang tercapai pada suatu saham.
- *Low* (harga terendah)  
*Low* adalah harga terendah transaksi yang tercapai suatu saham.
- *Bid* (minat beli)  
*Bid* adalah harga jual yang diminati pembeli untuk melakukan transaksi.
- *Ask* (minat jual)  
*Ask* adalah harga yang diminati penjual untuk melakukan transaksi

Kegiatan jual beli saham terjadi setiap hari, dimulai dengan sebuah tanda, yaitu munculnya harga *open* sebagai harga awal saham yang berlaku pada hari tersebut. Harga *open* diambil secara acak dengan catatan bahwa harga acak tersebut tidak terlalu jauh dari harga terakhir pada hari sebelumnya. Calon

investor lalu akan melihat urutan dari harga *ask* dan menyesuaikan dengan *bid* yang diajukan. Apabila telah menemukan harga yang sesuai, maka transaksi dapat dilakukan dan akan tercatat di pasar modal. Harga saham akan berubah apabila telah terjadi transaksi pertama dengan harga *bid* yang berbeda dengan harga *open*. Transaksi pertama tersebut akan merubah harga *close* sesuai dengan harga yang tercatat dalam transaksi. Hal tersebut akan terus mengulang sampai transaksi terakhir terlaksana. Kemudian akan terlihat keseluruhan data harga saham yang terjadi pada hari itu, mulai dari harga *open* sampai dengan harga *close*.



**Gambar 2.1 Grafik Fluktuasi Harga Saham.**

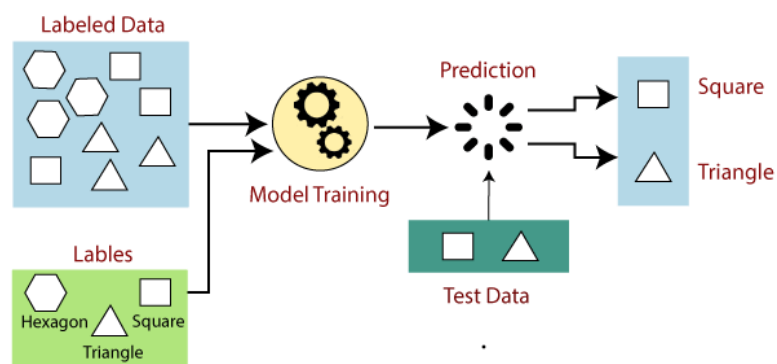
### **2.1.2. *Machine Learning***

Definisi *Machine Learning* menurut Arthur Samuel pada bukunya dengan judul *Some Studies in Machine Learning Using the Game of Checkers* yang terbit pada tahun 1988 berupa sebuah pertanyaan yang berbunyi “Bagaimana komputer dapat melakukan pembelajaran untuk menyelesaikan sebuah masalah tanpa diprogram secara eksplisit?”. Untuk mencapai automasi tersebut, komputer perlu melakukan proses “belajar” yang mempunyai konsep hampir sama seperti manusia, yaitu dengan mengembangkan algoritma dari melihat dan mempelajari pola berdasarkan data terdahulu. Dalam *Machine Learning* terdapat empat jenis pembelajaran, diantaranya;

*supervised learning, unsupervised learning, semi unsupervised learning, dan reinforcement learning.*

### 2.1.3. *Supervised Learning*

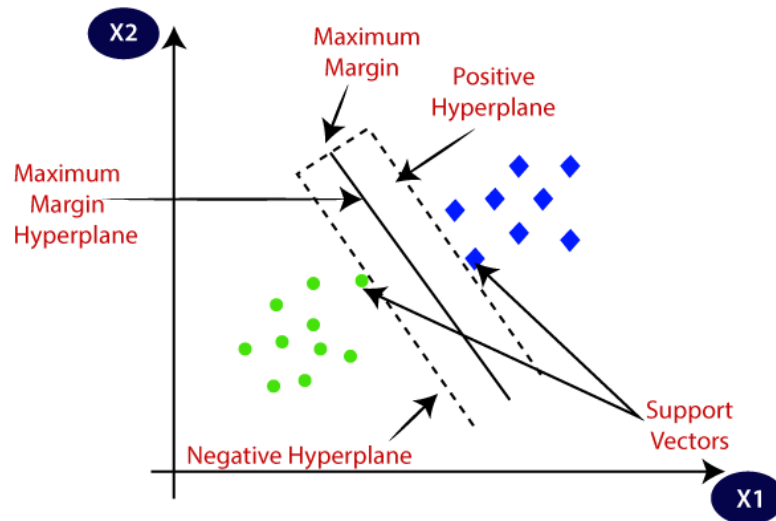
*Supervised learning* merupakan salah satu cabang dari *machine learning* yang berupa algoritma. Cara kerja dari *supervised learning* yaitu algoritma akan menerima contoh *input* dan *output* yang diinginkan, sehingga algoritma tersebut akan melihat sebuah pola berdasarkan contoh yang telah diberikan (Abijono, H., et al, 2021). Proses dari algoritma ini memiliki 2 tahap, yaitu proses *training* dan proses *testing*. Dalam proses *training*, algoritma ini akan membagi data yang sudah memiliki label menjadi 2, yaitu data latih (*data train*) dan *data test*. Data latih akan digunakan sebagai bahan pembelajaran dan *data test* akan digunakan sebagai data untuk dilakukan pengujian performa algoritma. Dengan dibaginya data, secara teori algoritma akan mempelajari data latih untuk melihat pola dan ketergantungan data satu sama lain serta menerapkan pola dalam data latih ke dalam *data test*. Hal tersebut akan memiliki *output* berupa hasil prediksi ataupun klasifikasi dari data test (Russel, S.J., 2010).



**Gambar 2.2** Proses dari Algoritma *Supervised Learning*.

#### 2.1.4. *Support Vector Machine (SVM)*

*Support Vector Machine* atau disingkat dengan (SVM) merupakan salah satu algoritma yang tergolong dalam metode *Supervised Learning* yang biasa digunakan untuk melakukan klasifikasi dan/atau regresi data. SVM dapat melakukan klasifikasi atau regresi pada tipe data yang linear maupun *non-linear* (Han, J., et al, 2022). SVM diperkenalkan pertama kali oleh Vapnik pada tahun 1992 sebagai konsep unggulan dalam bidang *pattern recognition*. Algoritma ini dapat memilih model otomatis dan tidak memiliki masalah *overfitting*. Cara kerja dari *support vector machine* yaitu memasukan tiap nilai di dalam data kedalam plot/diagram yang mana tiap nilai dalam plot/diagram disebut sebagai *support vector*. *Support vector* nanti akan membentuk kelompok dalam plot/diagram dan nantinya dipisahkan oleh garis linear yang disebut *hyperplane*. Yang mana *hyperplane* ini akan diperoleh dari perolehan kalkulasi algoritma *support vector machine* tersebut dengan cara memperhitungkan jarak terjauh dari dua kelompok *support vector*. Jarak terjauh antara 2 kelompok *support vector* tersebut disebut sebagai *margin*.



**Gambar 2.3 Support Vector Machine Hyperplane.**

Terdapat dua kategori SVM yaitu *Support Vector Machine Classification* dan *Support Vector Machine Regression* (Fadilah, W. R. U., et al, 2020)

#### 2.1.4.1. Kernel SVM

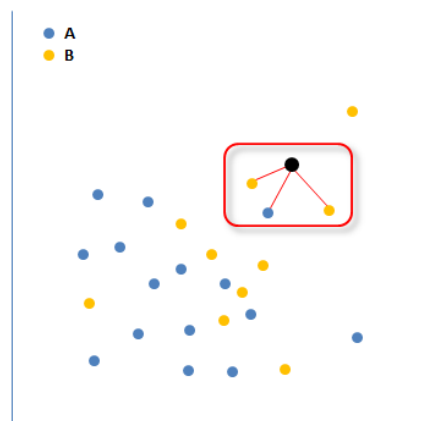
Kernel atau *Kernel function* merupakan sebuah fitur yang terdapat dalam model *machine learning* SVM yang berfungsi untuk menyelesaikan masalah yang tidak dapat diklasifikasi atau diprediksi secara linear (masalah *non-linear*). *Kernel function* dapat menyelesaikan masalah *non-linear* dengan menaikkan dimensi dari *support vector* dalam sebuah diagram yang berdimensi  $n$  menggunakan salah satu rumus kernel yang dipilih sehingga menghasilkan diagram baru dengan dimensi  $n+1$  atau lebih. Rumus kernel yang tersedia dan umum digunakan dalam SVM antara lain yaitu: *polynomial*, *gaussian/RBF (Radial Basis Function)*, dan *sigmoid (hyperbolic tangen)*.

#### 2.1.5. K-Nearest Neighbors (KNN)

K-Nearest Neighbors atau disingkat dengan (KNN) merupakan salah satu algoritma yang tergolong dalam metode



*Supervised Learning* yang biasa digunakan untuk melakukan klasifikasi dan/atau regresi data. Konsep dari algoritma ini adalah kelompok data dengan fitur yang serupa memiliki kemungkinan berada dalam satu kelompok data dalam diagram atau plot. Cara kerja dari algoritma ini yaitu menentukan sejumlah atau beberapa data yang bisa disebut sebagai  $n/K$  terdekat, dari data yang akan dilakukan prediksi atau klasifikasi dengan tujuan untuk mengetahui kedalam kelompok mana data baru ini akan digolongkan. Dari data  $n/K$  terdekat ini nantinya data baru tersebut akan digolongkan berdasarkan mayoritas dari kelompok data mana yang terdapat di dalam daftar data  $K$  terdekat tersebut.



**Gambar 2.4 Proses Pengelompokan Algoritma KNN.**

Berdasarkan dari gambar 2.4 di atas, dapat dilihat bahwa terdapat 2 kelompok data yaitu data A dengan warna biru dan data B dengan warna kuning. Selanjutnya dapat dilihat juga terdapat data berwarna hitam yang belum termasuk kedalam 2 kelompok data sebelumnya. Data hitam ini nantinya akan dicari sejumlah  $K$  data terdekat (dalam studi kasus gambar di atas, jumlah  $K$  terdekatnya = 3) lalu berdasarkan mayoritas dari ketiga  $K$  terdekat yang diperoleh, data hitam akan digolongkan ke dalam data  $K$  mayoritas tersebut. Mayoritas data dalam ke 3  $K$  data terdekat adalah data B yang

berwarna kuning. Berdasarkan kesimpulan tersebut, data hitam akan digolongkan ke dalam data B.

#### **2.1.6. Regresi**

Regresi merupakan salah satu metode yang sering digunakan dalam dunia statistika untuk memprediksi suatu nilai baru berdasarkan data yang sudah diperoleh. Menurut Nawari (2010), regresi merupakan investigasi terkait relasi fungsional antara variabel-variabel berbeda. Relasi antara variabel tersebut dituliskan dalam sebuah model matematika (Wahyono, T., 2010). Regresi juga diterapkan pada algoritma *machine learning*. Sebagai contoh algoritma yang menggunakan konsep regresi adalah SVM dan KNN. Secara sederhana, cara kerja dari regresi adalah melihat ketergantungan atau relasi dari dua variabel terikat dan variabel bebas.

#### **2.1.7. *Machine Learning Evaluation Metrics***

Setelah diterapkannya model *machine learning* dari suatu masalah atau studi kasus yang dipelajari, dibutuhkan upaya dalam melihat performa dari apa yang telah dihasilkan oleh algoritma *machine learning* tersebut. Upaya yang dimaksud dinamakan *evaluation metrics*. *Evaluation metrics* umum digunakan sebagai salah satu bagian dari *framework* dalam melakukan evaluasi pada berbagai studi. Dalam penerapannya di bidang regresi dan *machine learning*, *evaluation metrics* digunakan untuk membandingkan hasil prediksi yang diperoleh dari model *machine learning* dengan data sesungguhnya yang berasal dari data testing (Botchkarev, A., 2018).

Terdapat berbagai opsi pilihan *evaluation metrics* yang disesuaikan dengan tujuan dari penerapan model *machine learning* itu sendiri. Sebagai contoh opsi yang tersedia pada *evaluation metrics* yang digunakan dalam melakukan NLP dengan melakukan

regresi data akan berbeda. Untuk melakukan regresi, opsi yang tersedia dari *evaluation matrices* adalah; *Mean Square Error* (MSE), *Mean Absolut Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan opsi-opsi yang tersedia lainnya.

#### 2.1.7.1. *Mean Square Error* (MSE)

MSE adalah rata-rata kesalahan kuadrat antara nilai sebenarnya (*data testing*) dan nilai hasil prediksi. MAE juga merupakan salah satu *evaluation matrices* yang umum digunakan dalam analisa *time-series* data. Cara melakukan pengukuran performa menggunakan MSE yaitu, semakin kecil atau mendekati nol nilai MSE yang diperoleh, maka prediksi yang dilakukan memiliki performa yang lebih bagus. Berikut merupakan rumus dari MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (A_i - F_i)^2$$

Dimana:

$n$  = ukuran sampel

$A_i$  = data test ke  $i$

$F_i$  = data prediksi ke  $i$

#### 2.1.7.2. *Mean Absolute Error* (MAE)

MAE adalah rata-rata selisih mutlak nilai sebenarnya (*data testing*) dengan nilai hasil prediksi. MAE juga merupakan salah satu *evaluation matrices* yang umum digunakan dalam analisa *time-series* data. Cara melakukan pengukuran performa menggunakan MAE sama seperti dengan menggunakan MSE. Berikut adalah rumus dari MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - F_i|$$

Dimana:

$n$  = ukuran sampel

$A_i$  = data test ke  $i$

$F_i$  = data prediksi ke  $i$

#### 2.1.8. Google Colaboratory

Google Colaboratory atau biasa disebut dengan Google Colab merupakan sebuah *website* dari Google yang menyediakan layanan *cloud computing* khusus dalam melakukan pemrograman menggunakan bahasa Python untuk berbagai macam kebutuhan, sebagai contohnya yaitu *data science* dan *machine learning*. Google Colab sendiri memiliki antarmuka dan cara kerja yang hampir sama dengan Spyder atau Jupyter Notebook, yaitu pengguna akan mengetikkan kode Python ke dalam *code cell* dan dapat mengeksekusinya satu demi satu. Salah satu keunggulan dari Google Colab yaitu terdapat pada pelayanannya yang berupa *cloud computing* yang berarti semua proses komputasi yang terjadi pada *website* dilakukan didalam server yang dimiliki oleh Google. Dengan kata lain, peneliti tidak perlu melakukan proses instalasi *library* atau pustaka Python yang dibutuhkan ke dalam perangkat penelitian sehingga peneliti bisa menghemat penggunaan *disk* penyimpanan (Google).

#### 2.1.9. Python

Python merupakan bahasa pemrograman tingkat tinggi (*high level programming language*) yang diciptakan oleh Guido van Rossum pada akhir tahun 1980 (Kuhlman, D., 2009). Maksud dari bahasa pemrograman tingkat tinggi yaitu, Python dibuat dengan

bahasa yang sangat mirip dengan bahasa manusia yang kompleks dan bukan berupa deretan angka sebagaimana cara komputer berkomunikasi. Python tidak menggunakan titik koma (semicolon/ “ ; ”) sebagai penanda 1 baris kode dan kurung kurawal (curly brackets/ “ { } “) untuk menyatakan scope kode dalam sintaksnya. Melainkan menggunakan *enter (new line)* sebagai akhir dari baris kode dan indentasi sebagai *scope* kode pada sintaksnya.

Python merupakan salah satu bahasa pemrograman yang populer. Hal itu disebabkan oleh Python memiliki *code environment* berukuran besar dan *library* yang lebih variatif, serta dalam Python, dimuat *library* yang merupakan unsur dari *machine learning*.

#### **2.1.10. SKLearn**

Scikit learn atau yang biasa disebut sklearn merupakan sebuah pustaka (*library*) dari bahasa pemrograman Python yang dikembangkan berdasarkan pustaka python terkenal lainnya yaitu Numpy, SciPy, dan Matplotlib, untuk mempermudah proses yang dilakukan oleh *data scientist* dan *machine learning engineer*. Sklearn dikembangkan pertama kali oleh David Cournapeau sebagai bagian dari acara “Google Summer of Code Project” dan Matthieu Brucher melanjutkan proses pengembangan pustaka ini sebagai bagian dari tesis yang dikerjakannya (Scikit Learn).

#### **2.1.11. Pandas**

Pandas adalah pustaka dari bahasa pemrograman Python yang digunakan untuk melakukan pembersihan, pemrosesan, manipulasi, dan analisis data penelitian. Pustaka ini dibangun berdasarkan 2 pustaka Python yang terkenal lainnya, yaitu Numpy dan Matplotlib. Pandas sendiri mulai dikembangkan pada tahun 2008 oleh AQR Capital Management dan mulai menjadi *open source* pada tahun 2009 dan terus dikembangkan hingga sekarang.

### 2.1.12. Matplotlib

Matplotlib merupakan sebuah pustaka dari bahasa pemrograman Python yang berfungsi sebagai “alat” untuk melakukan visualisasi data untuk mempermudah penyajian dan pemahaman pengguna terhadap data yang sedang diolah. Matplotlib memiliki fitur visualisasi data yang beragam dan juga interaktif. Keberagaman fitur tersebut terletak pada banyaknya pilihan diagram yang dimiliki oleh pustaka ini dalam penyajian data. Lalu letak fitur interaktif yang ditawarkan oleh pustaka ini berada pada kemampuan untuk melakukan pembesaran gambar diagram secara otomatis sesuai dengan kehendak dari pengguna, dan juga visualisasi tiga dimensi yang disediakan oleh pustaka ini. Terlebih lagi terdapat banyak kemungkinan oleh pengguna untuk melakukan kostumisasi dalam penyajian data sehingga posibilitas untuk membuat penyajian data lebih menarik menjadi sangat luas.

## 2.2. Tinjauan Literatur

**Tabel 2.1 Tinjauan Literatur**

No	Nama dan Tahun Penelitian	Judul dan Lokasi Penelitian	Metode	Hasil Penelitian
1	Widya Rizka Ulul Fadilah, Dewi Agfiannisa, Yufis Azhar (2020)	Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode <i>Support Vector Machine</i>	<i>Support Vector Machine</i> K-Nearest Neighbour	Hasil dari penelitian ini berupa tingkat akurasi dari algoritma <i>Support Vector Machine</i> yaitu sebesar 0.964 dan RMSE sebesar 0.09 yang dibandingkan dengan algoritma K-Nearest

				<p>Neighbour dengan akurasi sebesar 0.945 dan RMSE sebesar 0.116 ketika kedua algoritma tersebut memprediksi harga saham dari PT. Telekomunikasi Indonesia. Sehingga dapat ditarik kesimpulan bahwa algoritma <i>Support Vector Machine</i> memiliki tingkat akurasi lebih tinggi dan RMSE yang lebih rendah.</p>
2	Eka Patriya (2020)	Implementasi <i>Support Vector Machine</i> Pada Prediksi Harga Saham Gabungan (IHSG)	<i>Support Vector Machine</i>	<p>Hasil dari penelitian ini berupa RMSE training dan testing yang masing-masing memiliki nilai 14.334 dan 20.281 serta MAPE training dan testing sebesar 0.211% dan 0.251%. dalam melakukan prediksi dari IHSG.</p>

3	Ratih Febrilia Tri Wulandari, Dian Anubhakti (2021)	Implementasi Algoritma <i>Support Vector Machine</i> (SVM) dalam Memprediksi Harga Saham PT. Garuda Indonesia TBK	<i>Support Vector Machine</i>	Hasil penelitian ini berupa nilai akurasi dari prediksi yang dilakukan oleh algoritma <i>Support Vector Machine</i> pada data saham PT. Garuda Indonesia dalam kurun waktu 18 maret 2019 hingga 23 april 2021 yaitu sebesar 0.545.
4	Falentino Sembiring, Mayang Gunawan, Rosalinda Hakim, Vemi Januarita Putri (2023)	Komparasi Pergerakan Saham Apple Dan Samsung Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	<i>Support Vector Machine</i>	Hasil penelitian ini yaitu didapatkannya <i>Rate Of Change</i> (ROC) dari Saham Samsung sebesar 0,435% dan dari Saham Apple sebesar 0,359%. Data yang digunakan memiliki rentang waktu dari awal 2017 sampai dengan 2022.
5	Sudriyanto, Rian Hidayad, Rafsanjai Akbar Ronaldo,	Optimasi Parameter <i>Support Vector Machine</i> Menggunakan Algoritma	<i>Support Vector Machine</i>	Hasil penelitian ini yaitu didapatkannya nilai RMSE sebesar 140.000 +/- 5.698 dan SE sebesar 19629.215 +/- 1609.864.



	Riangga Aji Prasetiyo, Setyo Agung Edho Wicaksono (2022)	Genetika untuk Meningkatkan Prediksi Pergerakan Harga Saham .		Sedangkan dengan metode <i>Support Vector Machine</i> Berbasis Algoritma Genetika berada pada nilai RMSE sebesar 101.208 +/- 9.475 dan SE sebesar 10323.858 +/- 1956.237.
6	Rusma Insan Nurachim (2019)	Pemilihan Model Prediksi Indeks Harga Saham yang Dikembangkan Berdasarkan Algoritma <i>Support Vector Machine</i> (SVM) atau <i>Multilayer Perceptron</i> (MLP) Studi Kasus Saham PT. Telekomunikasi Indonesia TBK.	<i>Support Vector Machine, Multilayer Perceptron</i>	Hasil penelitian ini berupa akurasi dari keda algoritma yaitu <i>Support Vector Machine</i> dan <i>Multilayer Perceptron</i> . Keduanya masing-masing memiliki akurasi prediksi tertinggi sebesar 90.9% pada eksperimen ke 15 dan 92.5% pada eksperimen ke 6. Kesimpulan yang yang dapat ditarik dari penelitian tersebut adalah, algoritma MLP mengungguli SVM

				dalam melakukan prediksi sebesar 1.4%.
7	Yudi Ramdhani, Ade Mubarak (2019)	Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi	<i>Support Vector Machine</i>	Hasil penelitian ini berupa nilai RMSE sebesar 22.662 nilai terbilang belum optimal untuk hasil prediksi, sehingga dilakukan optimasi parameter pada algoritma tersebut untuk meningkatkan hasil prediksi. Setelah dilakukan optimasi parameter menggunakan algoritma genetika hasil yang diperoleh yaitu nilai RMSE sebesar 10.495.
8	Anisa Aulia, Bella Aprianti, Yusuf Supriyanto,	Prediksi Harga Emas dengan Menggunakan Algoritma <i>Support Vector Regression</i> (Svr) dan <i>Linear Regression</i> (LR)	<i>Support Vector Regression</i> , <i>Linear Regression</i>	Hasil penelitian ini berupa perbandingan dari MSE algoritma SVM dengan <i>Linear Regression</i> yang masing-masing bernilai 7.524 dan 4.044 dengan rasio

	Chaerur Rozikin (2022)			<p>perbandingan data <i>training</i> dan testing sebesar 1:4 pada data harga emas dari 1 januari 2019 hingga 29 november 2021.</p> <p>Berdasarkan MSE tersebut, dapat ditarik kesimpulan bahwa algoritma <i>Linear Regression</i> memiliki akurasi yang lebih baik dibandingkan algoritma <i>Support Vector Regression</i>.</p>
9	Elfa Aufa Nida, Agung Budi Prasetijo, Risma Septiana (2020)	<p>Analisis Kinerja Algoritma <i>Support Vector Machine</i> (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk.</p>	<i>Support Vector Machine</i>	<p>Hasil dari penelitian ini berupa sebuah perbandingan akurasi dari penerapan algoritma SVM pada data saham ELNUSA dan saham sub sektor yang masing-masing memiliki nilai yaitu; 92,34% dan 91,25%.</p> <p>Kedua akurasi tersebut didapatkan setelah melakukan uji testing algoritma</p>

				menggunakan nilai eksponen 2 dan menggunakan data non-agresif
--	--	--	--	---

#### Perbedaan dengan Penelitian terdahulu

Penelitian yang dilakukan oleh Widya Rizka Ulul Fadilah dkk, Dewi Agfiannisa, dan Yufis Azhar (2020) dengan judul Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode *Support Vector Machine* menggunakan algoritma *Support Vector Machine* untuk melakukan prediksi terhadap harga saham PT. Telekomunikasi Indonesia dan membandingkan hasilnya dengan algoritma KNN. Penelitian tersebut menghasilkan perbandingan dengan masing-masing algoritma yaitu 0.964 dan RMSE sebesar 0.09 dan 0.945 dan RMSE sebesar 0.116 . Perbedaan yang ada dengan penelitian ini adalah perbedaan metode yang digunakan dan parameter analisis. Metode yang akan digunakan pada penelitian ini adalah algoritma SVM sebagai *tools* untuk melakukan prediksi harga saham bank digital dan tidak melakukan prediksi dengan algoritma yang lain sebagai parameter pembanding dalam melakukan analisis prediksi harga saham bank digital.

Penelitian yang dilakukan oleh Eka Patriya (2020) dengan judul Implementasi *Support Vector Machine* Pada Prediksi Harga Saham Gabungan (IHSG) menggunakan algoritma SVM untuk melakukan prediksi harga saham gabungan (IHSG) menghasilkan RMSE *training* dan testing sebesar 14.334 dan 20.281serta MAPE *training* dan testing sebesar 0.211% dan 0.251%. Perbedaan penelitian tersebut dengan penelitian ini terletak pada objek yang akan diteliti. Objek yang diteliti oleh Eka Patriya yaitu data dari Indeks Harga Saham Gabungan atau (IHSG), sedangkan objek yang akan diteliti adalah 2 data historis harian dari Allo Bank dan Bank Jago.

Penelitian yang dilakukan oleh Ratih Febrilia Tri Wulandari dan Dian Anubhakti (2021) dengan judul Implementasi Algoritma *Support Vector Machine* (SVM) dalam Memprediksi Harga Saham PT. Garuda Indonesia TBK menggunakan algoritma SVM untuk melakukan prediksi pada harga saham PT. Garuda Indonesia menghasilkan akurasi sebesar 0.545. Perbedaan dari penelitian di atas dengan penelitian ini adalah perbedaan objek penelitian. Dimana penelitian yang dilakukan oleh Ratih Febrilia Tri Wulandari dan Dian Anubhakti menjadikan saham dari PT. Garuda Indonesia sebagai objek penelitian, dan peneliti menjadikan saham dari Allo Bank dan Bank Jago sebagai objek penelitian.

Penelitian yang dilakukan oleh Falentino Sembiring, Mayang Gunawan, Rosalinda Hakim, dan Vemi Januarita Putri (2023) dengan judul Komparasi Pergerakan Saham Apple Dan Samsung Menggunakan Algoritma *Support Vector Machine* (SVM) melakukan komparasi dari pergerakan saham Apple dan Samsung dengan menggunakan algoritma SVM untuk melakukan prediksi pergerakan harga saham dari kedua perusahaan tersebut yang menghasilkan nilai *Rate Of Change* (ROC) dari Saham Samsung sebesar 0,435% dan dari Saham Apple sebesar 0,359%. Perbedaan dengan penelitian ini adalah objek yang akan dilakukan penelitian dan tujuan akhir dari penelitian. Perbedaan objek penelitian terletak pada Falentino Sembiring et, al, yang menjadikan saham perusahaan Apple dan Samsung, dan penulis yang menjadikan saham dari Allo Bank dan Bank Jago. Kemudian pada tujuan akhir penelitian, Falentino Sembiring et, al, bertujuan akhir untuk melihat akurasi dari prediksi pergerakan harga saham Apple dan Samsung, sementara peneliti bertujuan untuk menganalisis akurasi dari SVM untuk melakukan prediksi pada saham Allo Bank dan Bank Jago secara *general*.

Penelitian yang dilakukan oleh Sudriyanto, Rian Hidayad, Rafsanjai Akbar Ronaldo, Riangga Aji Prasetyo, dan Setyo Agung Edho Wicaksono (2022) yang berjudul Optimasi Parameter *Support Vector Machine* Menggunakan Algoritma Genetika untuk Meningkatkan Prediksi Pergerakan Harga Saham melakukan optimasi parameter algoritma SVM dengan menggunakan algoritma genetika untuk melihat peningkatan dari akurasi prediksi harga saham. Perbedaan dengan

penelitian ini terletak pada hasil akhirnya. Sudriyanto et, al. berfokus untuk meneliti perbedaan akurasi prediksi algoritma SVM setelah adanya peningkatan dengan menerapkan optimasi parameter algoritma SVM dengan menggunakan algoritma genetika. Sedangkan peneliti bertujuan untuk melihat tingkat akurasi prediksi harga saham dengan menggunakan algoritma SVM.

Penelitian yang dilakukan oleh Rusman Insan Nurachim (2019) dengan judul Pemilihan Model Prediksi Indeks Harga Saham yang Dikembangkan Berdasarkan Algoritma *Support Vector Machine* (SVM) atau *Multilayer Perceptron* (MLP) Studi Kasus Saham PT. Telekomunikasi Indonesia memiliki hasil berupa pernyataan yaitu algoritma MLP mengungguli SVM dalam melakukan prediksi sebesar 1.4%. Perbedaan dengan penelitian ini berada pada metode yang digunakan dan objek penelitian. Rusman Insan Nurachim dalam penelitiannya menggunakan dua jenis algoritma yaitu SVM dan MLP dalam melakukan penelitian terhadap harga saham PT. Telekomunikasi Indonesia sebagai objek yang akan diteliti. Sementara peneliti berfokus untuk menggunakan satu algoritma yaitu SVM untuk meneliti objek penelitian yaitu data harga saham dari Allo Bank dan Bank Jago.

Penelitian yang dilakukan oleh Yudi Ramdhani dan Ade Mubarok (2019) dengan judul Analisis *Time Series* Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi menggunakan metode SVM untuk melakukan prediksi dari penutupan harga saham Antam yang menghasilkan nilai RMSE 22.662. Dan nilai RMSE 10.495 setelah dilakukannya optimasi. Perbedaan dengan penelitian ini yaitu objek penelitian. Objek yang dijadikan sebagai bahan penelitian oleh Yudi Ramdhani dan Ade Mubarok adalah data harga saham dari Antam. Sementara peneliti menjadikan harga saham dari Allo Bank dan Bank Jago sebagai objek yang akan dilakukan penelitian.

Penelitian yang dilakukan oleh Anisa Aulia, Bella Aprianti, Yusuf Supriyanto, dan Chaerur Rozikin (2022) dengan judul Prediksi Harga Emas dengan Menggunakan Algoritma *Support Vector Regression* (SVR) dan *Linear Regression* (LR) melakukan Prediksi Harga Emas dengan Menggunakan Algoritma *Support Vector Regression* dan *Linear Regression* mendapatkan hasil yaitu MSE dari kedua

algoritma yang masing-masing sebesar 7.524 untuk SVM dan 4.044 untuk LR. Perbedaan dengan penelitian ini yaitu berada pada metode yang digunakan. Anisa Aulia, et, al. menggunakan algoritma SVR dan LR sebagai metode dalam melakukan penelitian. Dan peneliti berfokus untuk menggunakan 1 algoritma saja yaitu SVM sebagai metode dalam melaksanakan penelitian.

Penelitian yang dilakukan oleh Elfa Aufa Nida, Agung Budi Prasetyo, dan Risma Septiana (2020) dengan judul Analisis Kinerja Algoritma *Support Vector Machine* (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk. menganalisa kinerja algoritma SVM untuk mendukung pengambilan keputusan dalam melakukan transaksi jual/beli saham PT Elnusa mendapatkan hasil yaitu akurasi prediksi sebesar 92,34%. Perbedaan dari penelitian ini berada pada tujuan akhir penelitian. Penelitian terdahulu ini memiliki tujuan akhir untuk yaitu mendukung pengambilan keputusan berupa melakukan transaksi jual beli dari saham PT. Elnusa. Sedangkan peneliti bertujuan untuk melihat akurasi yang dihasilkan oleh algoritma SVM.

### 2.3. Kerangka Penelitian

SVM merupakan sebuah metode *machine learning* yang dapat digunakan untuk melakukan prediksi, baik dalam kasus regresi maupun klasifikasi. Teknik SVM digunakan dengan tujuan mencari fungsi pemisah yang optimal, yaitu *hyperplane*, untuk memisahkan *data input* yang memiliki nilai variabel target yang berbeda (Fachrurrazi, 2011). Pada dasarnya, SVM termasuk kedalam *machine learning* yang dikembangkan dengan asumsi kelinieran, namun terdapat metode untuk merubah kelinieran SVM untuk menyelesaikan berbagai masalah tidak linear yang dikembangkan oleh Scholkopf and Smola pada tahun 2002 (Santosa B, 2010). Metode untuk merubah orientasi penyelesaian masalah menggunakan SVM dari linier menjadi *non-linier* disebut sebagai *kernel trick*. *Kernel trick* bekerja dengan cara menaikkan tingkatan dimensi dari suatu input lalu mencari *hyperplane* yang paling optimal dari jumlah  $n$  kelompok yang diinput. Salah satu masalah yang bisa dipecahkan dengan menggunakan SVM yaitu memprediksi harga saham.

Prediksi harga saham dilakukan guna menjadi penentu dalam pengambilan keputusan berinvestasi bagi para investor. Selain itu, alasan dari diprediksinya harga saham diakibatkan oleh saham yang menjadi salah satu tolak ukur atau parameter dari suatu perusahaan. Hal tersebut dikarenakan bahwa saham cenderung menggambarkan nilai atau *value* dari perusahaan yang dimaksud. Semakin tinggi dan stabil harga saham yang dimiliki, maka *value* perusahaan tersebut akan dipandang positif terutama bagi investor. Prediksi harga saham sekiranya diperlukan bagi khalayak umum sebagai acuan apabila suatu saat hendak melakukan investasi saham dan diperlukan secara khusus untuk investor guna memperkuat keputusan dalam berinvestasi yang akan diambil.

Berdasarkan dari kebutuhan yang telah dipaparkan pada paragraf sebelumnya, penelitian ini diharapkan menjadi acuan dan solusi dalam memprediksi harga saham menggunakan metode SVM.

**Tabel 2.2 Kerangka Penelitian**

<p style="text-align: center;"><b>MASALAH</b></p> <p style="text-align: center;">PREDIKSI HARGA SAHAM BANK DIGITAL ALLO BANK DAN BANK JAGO MENGGUNAKAN METODE <i>SUPPORT VECTOR MACHINE</i></p>		
<p style="text-align: center;"><b>EKSPERIMEN / ANALISA</b></p>		
<i>Tools</i>	Data	Metode



Google Colaboratory  Python  SKLearn  Pandas  Matplotlib	Data historis saham Allo Bank dan Bank Jago dimulai dari tahun 2020 sampai dengan 2022 yang bersumber dari laman kaggle.com.	<i>SUPPORT VECTOR MACHINE  K-NEAREST NEIGHBOUR</i>
<b>PENGUJIAN DAN VALIDASI</b>		
Membandingkan hasil prediksi model <i>machine learning</i> SVM dengan <i>data test</i> dan hasil prediksi model <i>machine learning</i> KNN yang bersumber dari data historis saham Allo Bank dan Bank Jago, serta menghitung <i>Mean Square Error</i> dan <i>Mean Absolute Error</i> dari kedua model <i>machine learning</i> .		
<b>HASIL</b>		
Nilai <i>error</i> yang dihasilkan oleh algoritma SVM dan KNN terhadap prediksi data saham Allo Bank dan Bank Jago selama periode tahun 2020 sampai dengan tahun 2022, serta analisis yang berdasar pada nilai <i>error</i> kedua algoritma.		

### **BAB III**

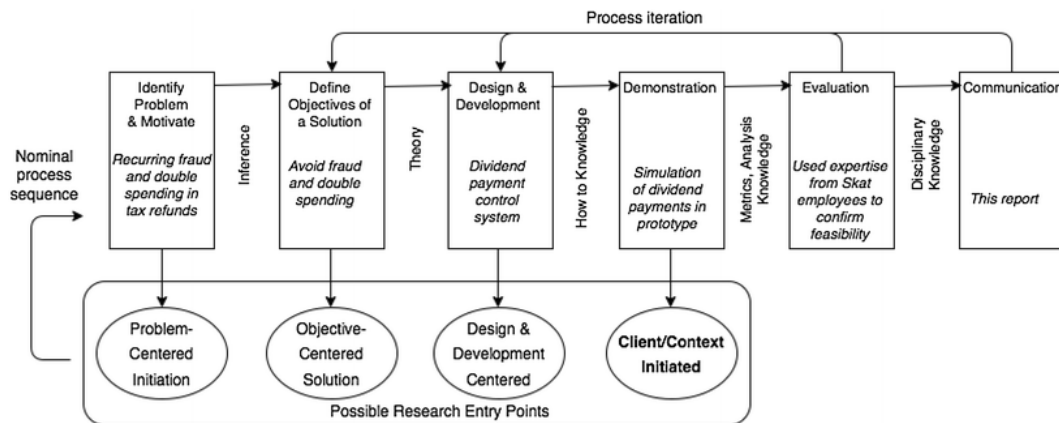
#### **METODE PENELITIAN**

Definisi dari metode penelitian yaitu langkah-langkah atau tindakan yang diambil oleh seorang peneliti untuk menghimpun informasi atau data serta melakukan penyelidikan lebih lanjut terhadap informasi atau data yang telah dikumpulkan dalam tahap sebelumnya (Sari, M. et al, 2022). Sebelum seorang peneliti melakukan penelitiannya, mereka diwajibkan untuk memahami inti dan maksud dari metode penelitian terlebih dahulu. Dengan begitu, seorang peneliti tersebut akan lebih mudah dalam menentukan dan melaksanakan hal-hal terkait penelitian. Definisi lain dari metode penelitian adalah suatu prosedur teratur dan terstruktur yang dilakukan oleh peneliti untuk mengumpulkan data dan informasi yang relevan dengan subjek atau objek penelitian yang sedang diteliti (Botchkarev, A., 2018). Terdapat banyak jenis dari metode penelitian. Dalam penelitian ini, metode penelitian yang digunakan adalah metode DSR (*Design Science Research*).

DSR atau *design science research* merupakan sebuah metode penelitian yang pada dasarnya merupakan sebuah kerangka berpikir dalam melakukan *problem solving*. Metode DSR bertujuan untuk menciptakan pengetahuan atau informasi tentang bagaimana seharusnya manusia menciptakan sesuatu (artifak) untuk meraih tujuan yang diinginkan (*Design Knowledge/DK*). Sebagai contoh, DK yang ada dalam penelitian ini adalah bagaimana langkah-langkah atau prosedur dalam membuat model *machine learning* yang dapat digunakan untuk memprediksi harga saham. Atau contoh lain DK dalam disiplin IT diantaranya, bagaimana cara membuat dan menstruktur *database* serta bagaimana cara dalam menerjemahkan hasil analisa data untuk membuat keputusan yang efektif (Vom Brocke, J. et al, 2020).

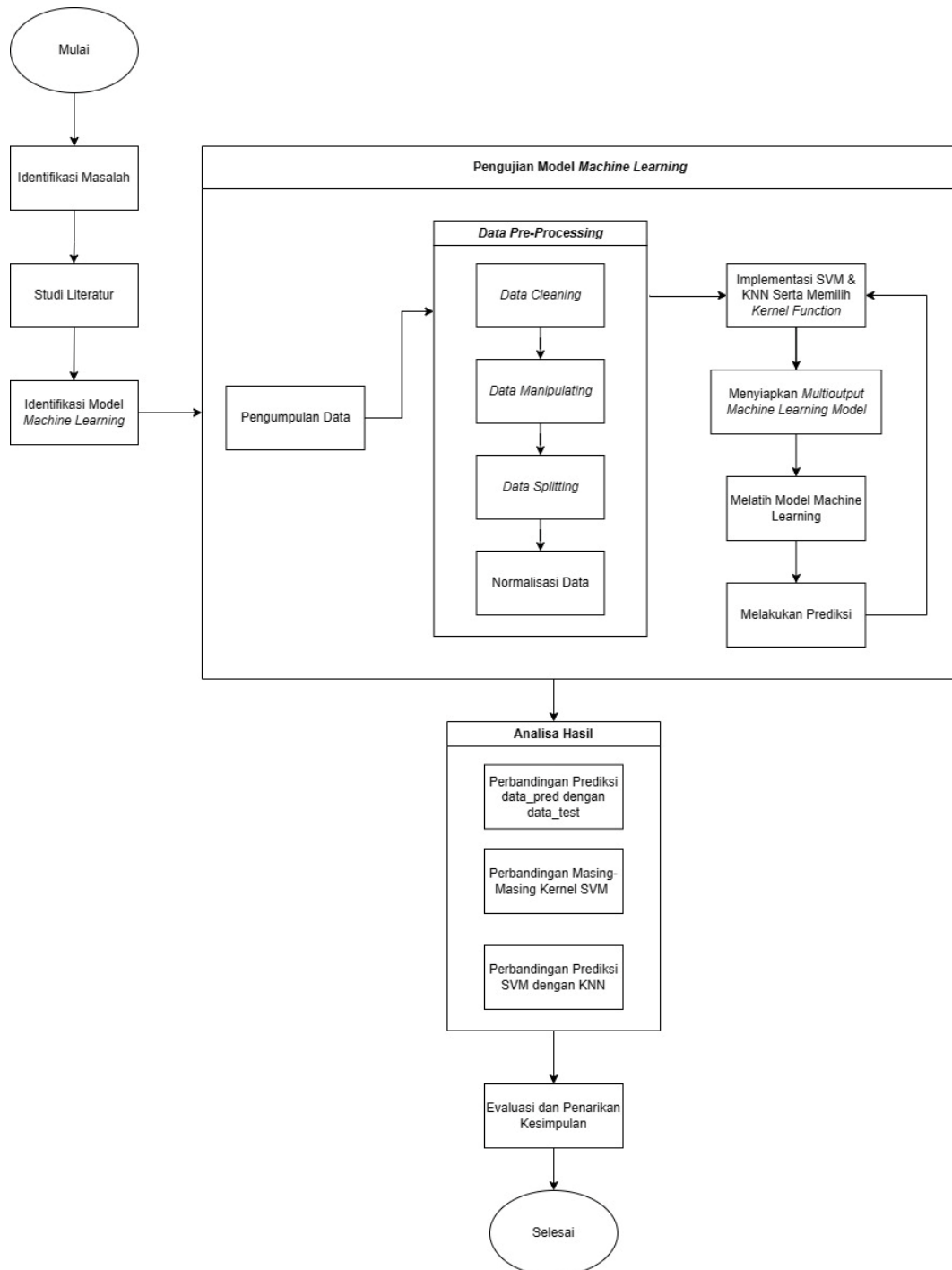
Prosedur yang dilakukan dalam melakukan metode DSR terdiri dari 6 langkah berdasarkan jurnal '*Introduction to Design Science Research*' karya Brocke, Jan Vom et al. Dalam jurnalnya Brocke menjelaskan bahwa prosedur dalam metodologi DSR adalah; 1. Mengidentifikasi masalah (*Identify Problem &*

*Motivate*), 2. Mendefinisikan tujuan dari solusi (*Define Objective of a Solution*), 3. Proses *design* dan *development* (*Design & Development*), 4. Demonstrasi artifak (*Demonstartion*), 5. Evaluasi (*Evaluation*), 6. Mengkomunikasikan Artifak (*Communication*). Untuk *flowchart* dari metodologi DSR dapat dilihat pada gambar dibawah.



**Gambar 3.1 *Flowchart* Metode DSR.**

### 3.1. Tahapan Penelitian



**Gambar 3.2 Flowchart Tahapan Penelitian**

## **1. Identifikasi Masalah**

Tahapan yang pertama dilakukan yaitu mengidentifikasi masalah yang akan dijadikan sebagai bahan penelitian. Masalah yang dapat dijadikan sebagai penelitian bisa didapatkan dari berbagai macam studi kasus pada kehidupan nyata. Sebagai contoh, masalah yang dibahas dalam penelitian ini.

## **2. Studi Literatur**

Tahap ini meliputi pengumpulan jurnal dan literatur relevan dengan penelitian yang dilakukan, yaitu melakukan prediksi terhadap harga saham dengan menggunakan algoritma SVM dan topik yang berkuat pada bidang yang serupa.

## **3. Identifikasi Model *Machine Learning***

Setelah dilakukannya studi literatur terkait penelitian yang serupa, tahap yang selanjutnya dilakukan adalah mengidentifikasi model *machine learning* yang akan digunakan dalam penelitian.

## **4. Pengujian Model *Machine Learning***

Tahap ini meliputi semua proses dari pembuatan model *machine learning* hingga mendapatkan hasil prediksi dari harga saham sesuai dengan algoritma yang dipilih. Berikut merupakan penjabarannya:

- **Pengumpulan Data**

Tahap ini dilakukan pengumpulan data yang akan digunakan untuk melakukan prediksi. Data yang dikumpulkan tersebut berupa data historis harian dari saham Allo Bank dan Bank Jago selama periode tahun 2020 sampai dengan awal tahun 2023.

- ***Data Pre-Processing***

Pada tahap ini akan dilakukan berbagai macam persiapan dalam pengolahan data sebelum pengimplementasian model *machine learning*. Berikut merupakan langkah-langkah dalam melakukan data *pre-processing* dalam melakukan penelitian ini:

- ***Data Cleaning***

Berdasarkan namanya, proses data *cleaning* merupakan proses yang dilakukan untuk membersihkan data dari data yang tidak dipakai atau data yang dapat mengakibatkan terjadinya *output* dari prosedur yang buruk. Karena di dalam pelaksanaan *data mining* terdapat istilah yaitu “*Garbage in, garbage out*”. Demi dipastikannya keberhasilan dalam penelitian ini, maka tahap ini menjadi salah satu tahapan yang wajib dilaksanakan. Dalam penelitian ini, proses data *cleaning* dilaksanakan dalam 2 bagian; bagian pertama yang dilakukan secara manual, dan yang kedua didalam google colaboratory.

- ***Data Manipulating***

*Data manipulating* merupakan proses mengubah atau memodifikasi data mentah dengan tujuan agar data tersebut dapat dijadikan input kedalam model *machine learning*. Proses data *manipulating* yang akan dilakukan dalam penelitian ini yaitu, melakukan modifikasi dari data harga saham Indonesia untuk menghasilkan data harga saham tiga hari yang akan datang, dimana data baru tersebut akan dijadikan sebagai data *testing* atau data target dalam proses untuk memprediksi harga saham.

- ***Data Splitting***

*Data splitting* merupakan proses pembagian data berdasarkan rasio yang diinginkan oleh peneliti yang bertujuan untuk dibaginya peran data menjadi *data train* dan *data test* dalam *input* kepada model *machine learning*. Rasio pembagian data yang digunakan dalam penelitian ini adalah 80% untuk *data train* dan 20% untuk *data test*.

- **Normalisasi Data**

Proses normalisasi data merupakan proses transformasi data sebelum diproses menggunakan model *machine learning* yang dilakukan dengan cara memperkecil *range* data numerik yang akan digunakan menjadi diantara angka 0 sampai 1 tanpa mengubah jarak asli antara data awal tersebut. Normalisasi data bertujuan untuk menghindari *error* yang terjadi dalam proses implementasi *machine learning* yang diakibatkan oleh perbedaan selisih data yang besar dan *range* data keseluruhan yang banyak.

- **Implementasi SVM & KNN Serta Memilih *Kernel Function***

Pada tahap ini nanti akan dipanggilnya algoritma SVM dan KNN serta ditentukannya *kernel* mana yang akan digunakan untuk melakukan prediksi harga saham. Kernel yang pertama kali digunakan adalah kernel linear terlebih dahulu. Hal tersebut bertujuan untuk melihat apakah data harga saham bisa diprediksi secara linear atau tidak, dan apakah hasil prediksi secara linear lebih memiliki performa yang baik atau buruk dibandingkan dengan hasil prediksi secara *non-linear*.

- **Menyiapkan *Multioutput Machine Learning Model***

Dalam tahap ini, akan terjadi proses persiapan dalam implementasi dari *multioutput machine learning model* yang nantinya akan digunakan dalam proses prediksi data saham 3 hari yang akan datang. *Multioutput Machine Learning Model* ini nantinya akan terimplementasi kedalam 2 metode *machine learning* yang digunakan yaitu SVM dan KNN.

- **Melatih Model *Machine Learning***

*Multioutput machine learning model* yang telah disiapkan pada tahapan sebelumnya akan melakukan proses training. *Data training* yang digunakan merupakan hasil *splitting* yang telah dilakukan pada tahapan sebelumnya.

- **Melakukan Prediksi**

Model *machine learning* yang telah selesai melewati proses *training* selanjutnya akan melakukan prediksi menggunakan *data testing*. Prediksi ini nantinya akan dilakukan oleh kedua model *machine learning* yang masing masing menggunakan metode SVM dan KNN.

## **5. Analisa Hasil**

Proses ini nantinya akan menganalisa hasil performa model *machine learning* yang diperoleh setelah melakukan prediksi harga saham. Analisa tersebut dilakukan dengan membandingkan nilai MSE dan MAE dari masing masing hasil yang diperoleh. Dipilihnya MSE dan MAE sebagai *evaluation matrices* disebabkan oleh umumnya penggunaan kedua parameter tersebut didalam penelitian yang berfokus pada *machine learning*. Hal



tersebut dapat dilihat pada bab 2 subbab tinjauan literatur di tabel 2.1. Proses analisa hasil akan meliputi beberapa analisa berikut:

- **Perbandingan Prediksi data\_pred Dengan data\_test**

Analisa dilakukan dengan membandingkan MSE dan MAE dari hasil prediksi dari model *machine learning* menggunakan algoritma SVM dengan *data test* nya.

- **Perbandingan Masing-masing Kernel SVM**

Analisa dilakukan dengan membandingkan MSE dan MAE dari hasil prediksi dari berbagai macam kernel SVM.

- **Perbandingan Prediksi SVM Dengan KNN**

Analisa dilakukan dengan membandingkan membandingkan MSE dan MAE dari hasil prediksi model *machine learning* SVM dan hasil prediksi model *machine learning* KNN.

## **6. Evaluasi dan Penarikan Kesimpulan**

Tahapan terakhir ini akan melakukan evaluasi dari penelitian yang telah dilakukan serta evaluasi dari model *machine learning* yang digunakan, selanjutnya akan ditarik sebuah kesimpulan berdasarkan hasil dari proses analisa sebelumnya.

### **3.2. Sumber Data**

Pengumpulan data diperlukan untuk melakukan penelitian. Karena data-data ini akan digunakan sebagai landasan utama dalam penelitian. Data diperoleh melalui *website* kaggle.com. Data ini berisi catatan historis harga saham mulai dari tahun 2001 sampai dengan awal 2023.

Penelitian ini menggunakan *dataset* yang memiliki lima *field*, yaitu; *date*, *open*, *high*, *low*, dan *close*. Berdasarkan *field* data tersebut, *field* akan

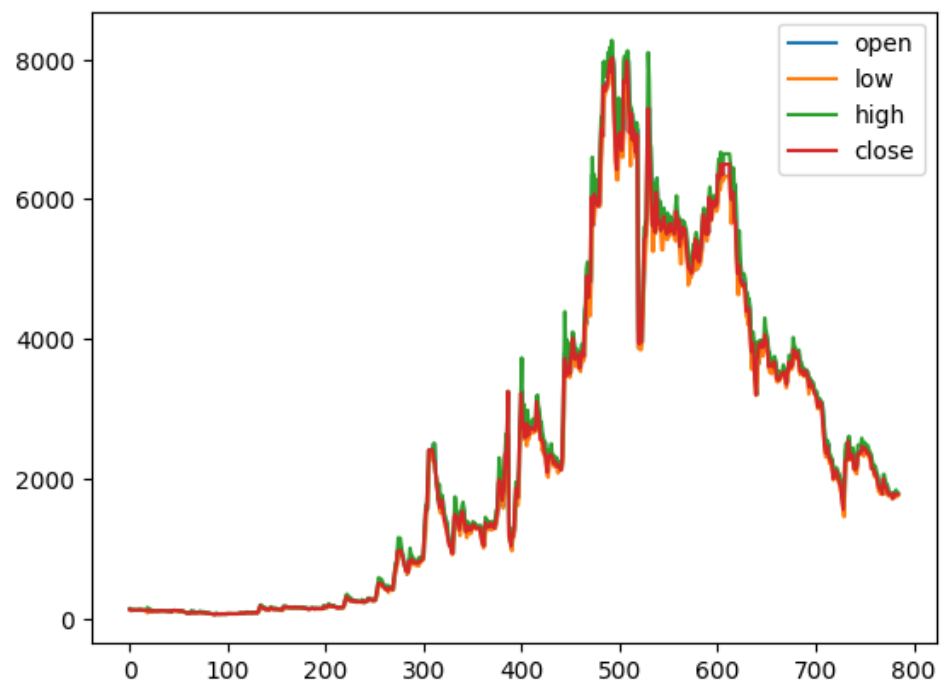
dikelompokkan menjadi dua variabel yaitu *independent* variabel dan *dependent* variabel oleh SVM dan KNN secara otomatis. Keseluruhan *dataset* kemudian akan kembali dibagi menjadi dua untuk kebutuhan *training* dan *testing*.

### 3.3. Objek Penelitian

Objek penelitian merupakan sebuah objek yang akan diteliti dalam penelitian. Objek penelitian dalam penelitian ini yaitu berupa data historis harga saham dari Allo Bank dan Bank Jago dari periode tahun 2020 sampai dengan awal tahun 2023. Berikut merupakan jabaran dari kedua objek penelitian.

#### 3.3.1. Data Saham Allo Bank

Allo bank Pada tahun melakukan Penawaran Saham Perdana (IPO) kepada masyarakat pada tahun 2015 dengan kode BBHI. Data yang digunakan dalam penelitian ini merupakan data historis harga saham Allo Bank. Berikut merupakan *line plot* dari harga saham Allo Bank periode tahun 2020 sampai dengan awal tahun 2023.



**Gambar 3.3 Data Harga Saham Allo Bank.**

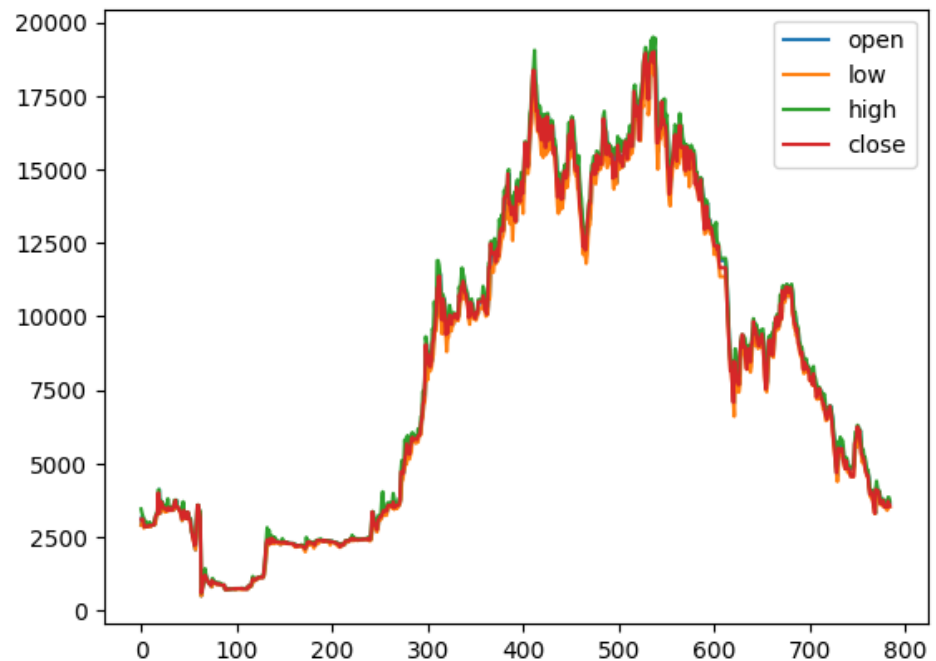
Berdasarkan *line plot* diatas, dapat dilihat bahwa saham BBHI memiliki pergerakan harga yang fluktuatif selama periode dari tahun 2020 sampai dengan awal tahun 2023. Penjelasan mengenai data deskriptif dari harga saham BBHI akan disajikan dalam tabel dibawah.

**Tabel 3.1 Data Deskriptif Harga Saham Allo Bank**

	<i>Min</i>	<i>max</i>	<i>mean</i>	<i>std</i>	<i>var</i>
<i>open</i>	57	8025	2341,256	2241,558	5,024
<i>low</i>	54	7850	2273,448	2182,499	4,763
<i>high</i>	61	8275	2431,980	2319,318	5,379
<i>close</i>	58	8025	2343,515	2243,036	5,031

### 3.3.2. Data Saham Bank Jago

Bank Jago melakukan IPO dan terdaftar pada Bursa Efek Indonesia dalam kode ARTO. Data yang digunakan dalam penelitian ini merupakan data historis harga saham Bank Jago. Berikut merupakan *line plot* dari harga saham Bank Jago periode tahun 2020 sampai dengan awal tahun 2023.



**Gambar 3.4. Data Harga Saham Bank Jago.**

Berdasarkan *line plot* diatas, dapat dilihat bahwa saham BBHI memiliki pergerakan harga yang fluktuatif selama periode dari tahun 2020 sampai dengan awal tahun 2023. Penjelasan mengenai data deskriptif dari harga saham BBHI akan disajikan dalam tabel dibawah.

	<i>min</i>	<i>max</i>	<i>mean</i>	<i>std</i>	<i>var</i>
<i>open</i>	492	19100	8211,143	5495,179	3.019
<i>low</i>	492	18700	7990,963	5344,127	2.855
<i>high</i>	590	19500	8426,035	5616,246	3.154
<i>close</i>	590	19000	8204,165	5484,524	3.008

**Tabel 3.2. Data Deskriptif Harga Saham Bank Jago**

### 3.4. Instrumen Penelitian

Berikut adalah instrumen yang digunakan dalam penelitian guna mengetahui hasil akhir dari penelitian ini:

- Perangkat Keras

Berikut ini merupakan spesifikasi dari perangkat keras yang penulis gunakan untuk mencari dan mengolah data dan melakukan uji algoritma:

1. *Processor* intel i3 1115G4
2. RAM 8 Gigabyte
3. Solid State Drive (SSD) 500 Gigabyte

- Perangkat Lunak

Adapun perangkat lunak yang dibutuhkan dalam melakukan penelitian ini yaitu:

1. Windows 11 64 bit
2. Google Colaboratory
3. Python3.8.1
4. SKLearn
5. Pandas
6. Matplotlib

### 3.5. Metode Analisa Data

Metode penelitian yang digunakan dalam subjek penelitian ini adalah metode penelitian kuantitatif. Hal tersebut dipilih berdasarkan data yang diperlukan

dalam penelitian ini yaitu data historis harga saham dari kedua bank digital pilihan yang ada di Indonesia. Serta didapatkannya hipotesa diawal yaitu bagaimana hasil yang akan diperoleh dari prediksi menggunakan kedua algoritma *machine learning* yang digunakan dalam penelitian. Dalam menganalisa data, diperlukannya objek penelitian, metode pengolahan data, dan parameter pembanding. Objek penelitian yang dibutuhkan dalam penelitian ini adalah data historis dari harga saham Allo Bank dan Bank Jago, metode pengolahan data yang digunakan adalah algoritma *Support Vector Machine*, dan parameter pembanding nya adalah algoritma *K-Nearest Neighbor*.

Analisa data akan dilakukan setelah didapatkannya hasil prediksi data saham menggunakan algoritma *Support Vector Machine*. Dalam implementasinya, algoritma *Support Vector Machine* tidak akan menggunakan modifikasi kernel apapun. Hal tersebut ditetapkan karena penelitian ini dilakukan untuk mempermudah khalayak umum dalam melakukan prediksi harga saham tanpa pengetahuan terkait teori dan implementasi *machine learning* yang mendalam.

Metode yang digunakan dalam menganalisa data prediksi saham yaitu metode komparatif antara hasil prediksi dengan data testing dari algoritma *Support Vector Machine*, lalu dilakukan komparasi selanjutnya antara hasil prediksi algoritma *Support Vector Machine* dengan hasil prediksi dengan algoritma *K-Nearest Neighbor*. Variabel yang akan dibandingkan dari kedua algoritma tersebut yaitu masing masing MSE dan MAE yang diperoleh dari hasil prediksi. Dengan membandingkan MSE dan MAE, akan dihasilkannya kesimpulan dari analisa yang akan menyatakan salah satu algoritma paling optimal dalam implementasinya untuk memprediksi harga saham berdasarkan dua saham bank digital yang dipilih dalam penelitian ini.

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

Pada bab ini akan dipaparkan langkah-langkah atau prosedur yang dilakukan dan hasil yang diperoleh dari penelitian. Hal-hal yang akan dipaparkan di bawah mengikuti prosedur dari gambar *flowchart* yang telah dilampirkan pada bab sebelumnya (gambar 3.1) dimulai dari identifikasi model *machine learning* dikarenakan bagian identifikasi masalah dan studi literatur sudah dibahas pada bab satu dan bab dua. Pemaparan dibawah juga dilengkapi dengan *source code* dari implementasi algoritma *machine learning* SVM dan KNN. *Source code* yang dipaparkan pada penjelasan di bawah hanya merupakan satu dari dua *source code* dalam pengembangan model *machine learning*, yaitu dalam pengembangan model *machine learning* untuk memprediksi harga saham ARTO (Bank Jago). Tidak dipaparkannya proses pengembangan model *machine learning* dalam memprediksi harga saham BBHI dikarenakan kemiripan proses yang terjadi, perbedaan mencolok yang terdapat dari kedua proses tersebut terdapat dalam penamaan variabel dan hasil yang diperoleh.

#### **4.1. Identifikasi Model *Machine Learning***

Proses identifikasi model *machine learning* dimulai pada waktu yang sama dengan proses studi literatur. Alasan dipilihnya algoritma SVM dan KNN untuk memprediksi harga saham BBHI dan ARTO adalah sebagai berikut:

1. Keduanya merupakan algoritma *machine learning* yang tergolong ke dalam cabang *supervised learning* dan bisa dilakukan untuk melakukan klasifikasi/regresi.
2. Kedua algoritma tersebut merupakan algoritma yang umum digunakan untuk melakukan klasifikasi/regresi data yang dapat dilihat pada referensi karya tulis dalam bagian studi literatur.
3. Konsep dasar yang cukup sederhana untuk dipahami dari kedua algoritma.

4. Adanya *kernel trick* dalam SVM yang dapat menyelesaikan masalah *non-linear* dan bisa sangat membantu dalam meningkatkan performa algoritma tersebut.

## 4.2. Pengujian Model *Machine Learning*

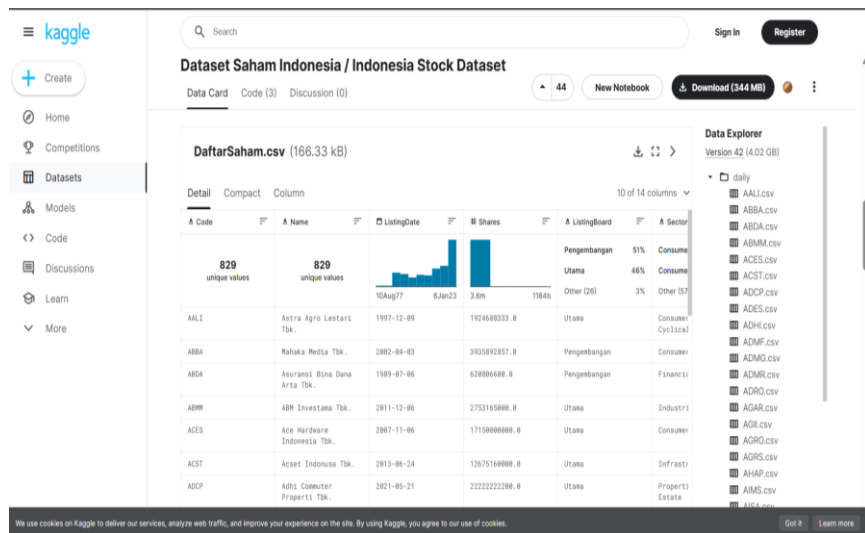
Tahapan ini merupakan proses pembuatan model *machine learning* dengan menggunakan algoritma SVM dan KNN untuk memprediksi nilai saham BBHI dan ARTO. Tahapan pembuatan yang akan dijelaskan di bawah akan dibagi menjadi enam bagian, yaitu; pengumpulan data; data *pre-processing*; menyiapkan *multioutput machine learning model*; memilih *kernel function*; melatih model *machine learning*; dan melakukan prediksi. Selanjutnya pada bagian data *pre-process* akan dibagi kembali menjadi empat bagian, yaitu; *data cleaning*; *data manipulating*; *data splitting*; dan normalisasi data.

### 4.2.1. Pengumpulan Data

Setelah berhasil mengidentifikasi algoritma yang akan digunakan dalam memprediksi harga saham, tahapan selanjutnya adalah melakukan pengumpulan data. Langkah pertama yang dilakukan dalam proses pengumpulan data yaitu dengan melakukan pencarian pada *website* google dengan *keyword* “Dataset Harga Saham Indonesia”. Hasil dari pencarian menampilkan beberapa *website* yang menyediakan *dataset* yang diinginkan, namun *website* kaggle.com yang dipilih dikarenakan *website* kaggle merupakan *website* yang memiliki reputasi sebagai *website* penyedia *dataset* untuk melakukan proses *data mining*, *data science*, dan *machine learning*. Walaupun dalam penelitian ini menjadikan *website* kaggle sebagai sumber utama dalam memperoleh objek penelitian tidak menutup kemungkinan bahwa objek penelitian bisa didapatkan pada *website-website* terpercaya dalam menyediakan dataset lainnya. *Link* lengkap dari *website* kaggle yang dijadikan sebagai sumber utama dalam memperoleh objek penelitian adalah



“<https://www.kaggle.com/datasets/7415fcfb5010d225fc1951e426d81c89cf473603d49f46bc6553bd3fe2688b68>”. Tampilan singkat dari halaman *website* berdasarkan *link* yang sudah dilampirkan diatas adalah sebagai berikut.



**Gambar 4.1 Tampilan Website Kaggle.com**

Data diperoleh dari *website* kaggle dengan cara diunduh ke dalam instrumen penelitian yaitu laptop dalam format file csv dengan total data asli sebanyak 3757 *entry*. Setelah mengalami tahap *pre-processing* diperoleh kembali data sebanyak 1570 *entry* dengan rentang periode yang dimulai pada tahun 2020 sampai dengan awal tahun 2023. Rentang periode tersebut dipilih untuk diteliti dikarenakan tingkat popularitas kedua bank tersebut sedang melejit yang disebabkan oleh promosi yang dilakukan oleh kedua belah pihak. Terdapat dua data yang diunduh, yaitu data historis harga saham Allo Bank dan Bank Jago. Berikut merupakan struktur dataset dari kedua data yang telah diunduh:

**Tabel 4.1 Tabel data field dalam dataset harga saham**

Nama Data Field	Jumlah Data
-----------------	-------------

<i>Timestamp</i>	788
<i>Open</i>	788
<i>Low</i>	788
<i>High</i>	788
<i>Close</i>	788
<i>Volume</i>	788

Keterangan:

- *timestamp* = tanggal dari dilakukannya trading saham.
- *open* = harga saham pembukaan di awal hari.
- *low* = harga terendah saham dalam hari yang tertera.
- *high* = harga tertinggi saham dalam hari yang tertera.
- *close* = harga saham yang tercatat dalam akhir hari.
- *volume* = jumlah saham yang terjual pada hari yang tertera

Dalam penelitian ini, *field* yang digunakan yaitu *field open*, *low*, *high*, dan *close* sebagai variabel independen. Variabel dependen atau variabel target yang akan digunakan yaitu *data field close* dalam tiga hari yang akan datang.

#### 4.2.2. *Pre-Processing Data*

Yang dimaksud dengan *pre-processing* data adalah langkah-langkah, tahapan, atau prosedur yang dilakukan sebelum data mentah dapat dijalankan pada proses utama dalam suatu prosedur yang lebih besar. Tujuan dilakukannya *pre-processing* secara sederhana dapat dijelaskan sebagai proses yang mengubah data mentah menjadi data siap pakai. Langkah-langkah atau prosedur yang dikategorikan sebagai bagian dari *pre-processing* data dalam penelitian ini terbagi menjadi 4 proses, yaitu *data cleaning*, *data manipulating*, *data splitting*, dan normalisasi data.

#### 4.2.2.1. *Data Cleaning*

Dalam tahap pertama, *cleaning* yang dilakukan pertama kali adalah mengerutkan data awal dari dataset harga saham BBHI sebanyak 1933 *entry* dan dataset harga saham ARTO sebanyak 1824 *entry* dengan rentang waktu dari akhir tahun 2015 sampai dengan awal tahun 2023, menjadi rentang waktu dari tahun 2020 sampai dengan awal tahun 2023 yang dilakukan secara manual dengan menggunakan aplikasi Microsoft Excel 2019. Kemudian, data yang diperoleh dengan rentang waktu yang sesuai akan kembali dihapus dari 3 data terakhir. Hal tersebut berkaitan dengan proses yang akan dilaksanakan pada bagian *data manipulating* nantinya. Setelah dilakukan *data cleaning* secara manual ini, diperoleh kembali *dataset* kedua bank tersebut dengan jumlah yang sama yaitu sebanyak 785 *entry data*.

Tahapan *data cleaning* kedua yang dilakukan di dalam Google Colaboratory bertujuan untuk menghapus *data field* yang tidak diperlukan, yaitu *data field timestamp* dan *volume*. Tidak diperlukannya kedua *data field* tersebut disebabkan oleh tidak adanya relevansi antara kedua *data field* yang telah disebutkan dengan perhitungan prediksi yang akan dilakukan.

Berikut adalah *source code* yang digunakan untuk melakukan *data cleaning* untuk menghapus *data field timestamp* dan *volume*:

```
import pandas as pd
data = pd.read_csv("/content/ARTO-3.csv")
del data['volume']
del data['timestamp']
```

Berdasarkan *source code* di atas, dapat dilihat bahwa file dengan format .csv yang bernama ‘ARTO-3.csv’ di *import* kedalam sistem dan dibuat menjadi variabel dengan nama data. Selanjutnya *data field volume* dan *timestamp* akan dihapus dari variabel data tersebut. Setelah mengeksekusi *source code* diatas, berikut merupakan perubahan dalam *dataset* harga saham ARTO yang diperoleh

**Tabel 4.2 *Dataset* Harga Saham ARTO Sebelum *Cleaning***

	<i>timestamp</i>	<i>open</i>	<i>low</i>	<i>high</i>	<i>close</i>	<i>volume</i>
0	2020-01-01	137	121	137	125	0
1	2020-01-02	125	110	139	117	272100
2	2020-01-03	117	111	125	120	16100
3	2020-01-06	113	113	124	124	201200
4	2020-01-07	120	120	122	121	10400
...	...	...	...	...	...	...
780	2022-12-28	1730	1730	1800	1775	1148200
781	2022-12-29	1780	1760	1790	1785	382400
782	2022-12-30	1785	1765	1840	1765	1549400
783	2023-01-02	1775	1740	1795	1775	413600
784	2023-01-03	1775	1760	1795	1775	290600

**Tabel 4.3 *Dataset* Harga Saham ARTO Setelah *Cleaning***

<i>open</i>	<i>low</i>	<i>high</i>	<i>close</i>
137	121	137	125
125	110	139	117
117	111	125	120
113	113	124	124
120	120	122	121

...	...	...	...
1730	1730	1800	1775
1780	1760	1790	1785
1785	1765	1840	1765
1775	1740	1795	1775
1775	1760	1795	1775

#### 4.2.2.2. Data Manipulating

Dalam penjelasan di subbab *data cleaning* telah disebutkan bahwa total keseluruhan data dikurangi tiga sehingga menghasilkan total *entry* dalam *dataset* menjadi 785. Hal tersebut bertujuan untuk mendapatkan total *entry* data yang sama banyaknya dan mendapatkan *output* yang sesuai ketika mengolah data saham menggunakan model *machine learning*. Perlu diingat bahwa *output* yang diinginkan adalah prediksi harga saham 3 hari mendatang.

Untuk memperoleh data saham 3 hari kedepan, pertama yang harus dilakukan adalah membuat *dataset* baru yang akan digunakan sebagai *data target* dalam model *machine learning*. *Dataset target* dapat dibuat dengan *source code* berikut

```
close = []
close_fix = []
for x in data.index:
    close.append(data['close'][x])
```

*Source code* diatas mendeklarasikan dua Python *list*, yaitu *close* dan *close\_fix*. Kemudian dilakukan *looping* dengan tujuan untuk memasukkan semua data *entry* dari *data field* *close* kedalam *list* *close* yang sudah dideklarasikan sebelumnya. Hasil dari *source code* tersebut adalah python

*list* yang berisikan dengan nilai dari harga saham dari *data field close*.

```
close_fix = close  
close_1 = []  
close_2 = []  
close_3 = []
```

*Source code* di atas membuat isi dari *list close\_fix* sama dengan *list close*. Hal ini bertujuan untuk menyimpan nilai dari *list close* ke variabel yang lain guna mengembalikan nilai *close* jika dibutuhkan. Kemudian terdapat deklarasi dari tiga Python *list* baru, yaitu *close\_1*, *close\_2*, dan *close\_3*. Ketiga *list* tersebut dibuat untuk mendapatkan tiga nilai *close* dalam tiga hari kedepan sesuai dengan *output* yang diinginkan, yaitu prediksi saham tiga hari yang akan datang.

```
for x in range(3):  
    if x == 0:  
        close = close_fix  
    if x == 1:  
        close = close_fix  
    if x == 2:  
        close = close_fix
```

Kemudian dilakukan *looping* sebanyak tiga kali dan *if statement* dengan jumlah sama yang masing masing dilakukan untuk mengisi *list close\_1*, *close\_2*, dan *close\_3* secara bergantian, apabila nilai *x* sama dengan kondisi dalam *if statement* yang telah dibuat. Misalkan nilai *x* sama dengan 0, maka akan dilakukan pengisian *list close\_1* dari *list close* yang sudah berisikan harga dari *data field close*. Hal tersebut berlaku untuk nilai *x* sama dengan 1 dan 2 yang masing masing akan mengisi *list close\_2* dan *close\_3*.

potongan *code* berikut:

```
close = close_fix
```

berfungsi untuk mengembalikan nilai dari *list* *close* setelah memenuhi salah satu kondisi yang ada pada *if statement*. Hal tersebut perlu dilakukan untuk alasan yang akan dijelaskan selanjutnya.

```
del close[0]
```

Potongan *code* diatas bertujuan untuk menghapus *entry* data yang pertama didalam *list* *close*. Perlu dihapusnya *entry* data pertama dalam *list* *close* berdasarkan oleh asumsi berikut. Diasumsikan bahwa nilai dari *list* *close* sama dengan *n*, maka dibutuhkan untuk *list* *close\_1* bernilai *n+1* untuk alasan *list* *close\_1* merupakan *list* yang berisi harga saham 1 hari kedepan, untuk alasan tersebut perlu dihapusnya *entry* data pertama dalam *list* *close*.

Hal tersebut juga berlaku dalam proses pengisian *list* *close\_2* dan *close\_3* yang mana nantinya nilai *list* yang telah disebutkan akan bernilai *n+2* dan *n+3*. Hanya saja, untuk menghapus dua dan tiga *entry* data untuk *list* *close\_2* dan *close\_3* perlu diberlakukan *looping* sebanyak dua dan tiga kali juga. Berikut merupakan terjemahan asumsi tersebut kedalam bahasa pemrograman python:

```
#looping untuk menghapus dua entry list close
untuk mengisi close_2
for x in range(1):
    del close[0]
#looping untuk menghapus tiga entry list close
untuk mengisi close_3
for x in range(2):
    del close[0]
    break
```

Selanjutnya *list* *close* yang sudah dihapus tiga data pertama nya akan dimasukan kedalam *list* *close\_1*, *close\_2*, dan *close\_3* berdasarkan jumlah data yang dihapus. Apabila dihapus satu data pertama maka *list* *close* tersebut akan dimasukan kedalam *list* *close\_1* dan seterusnya. Berikut merupakan terjemahan logika diatas kedalam bahasa pemrograman python.

```
close_1.extend(close)
close_2.extend(close)
close_3.extend(close)
```

Proses diatas apabila digabungkan dan diterjemahkan kedalam bahasa pemrograman python akan terlihat seperti dibawah.

```
close_fix = close
close_1 = []
close_2 = []
close_3 = []
for x in range(3):
    if x == 0:
        close = close_fix
        del close[0]
        close_1.extend(close)
    if x == 1:
        close = close_fix
        for x in range(1):
            del close[0]
        close_2.extend(close)
    if x == 2:
        close = close_fix
        for x in range(2):
            del close[0]
        break
    close_3.extend(close)
```

Keseluruhan *source code* tersebut menghasilkan tiga *list* yang masing masing berjumlah 784 *entry* data untuk *list* *close\_1*, 783 *entry* data untuk *list* *close\_2*, dan 782 *entry* data untuk *list* *close\_3*. Tidak samanya jumlah dari ketiga *list* yang telah dibuat akan mengakibatkan *error* ketika data



tersebut dimasukan kedalam model *machine learning*. Untuk menghindari hal tersebut, perlu ditambahkannya tiga data baru. Tiga data baru tersebut diperoleh dari proses *cleaning* manual ketika mengurangi total data sebanyak 3 *entry*. Ketiga data tersebut akan kembali dimasukan kedalam masing masing *list* sesuai dengan kekurangan data yang ada. Berikut merupakan *source code* untuk melakukan *input* secara manual(tanpa *looping*) kedalam masing masing *list*.

```
close_1.extend([3510])
close_2.extend([3510, 3270])
close_3.extend([3510, 3270, 3060])
```

Proses *data manipulating* selanjutnya yaitu mengubah ketiga *list* di atas menjadi python *dictionary*, dan selanjutnya menjadikan *dictionary* tersebut kedalam pandas *dataframe*. Berikut merupakan terjemahan langkah-langkah diatas kedalam bahasa pemrograman python.

```
close_1.extend([3510])
close_2.extend([3510, 3270])
close_3.extend([3510, 3270, 3060])
data_target = {
    'close+1': close_1,
    'close+2': close_2,
    'close+3': close_3,
}
target = pd.DataFrame(data_target)
```

#### 4.2.2.3. Data Splitting

Proses *data splitting* dilakukan dengan rasio 80% untuk *data training* dan 20% untuk *data testing*. Perlu diketahui bahwa terdapat banyak rasio *splitting* yang dapat digunakan, dan akan terjadi hasil yang kemungkinan

berbeda apabila menggunakan rasio *splitting* yang juga berbeda.

*Data splitting* diperlukan dalam proses pembuatan dan pengembangan model *machine learning* dikarenakan proses ini dibutuhkan sebagai bentuk *learning* yang dilakukan terlebih dahulu agar model *machine learning* bisa melakukan *pattern recognition*. Serta diperlukannya *data splitting* untuk melihat seberapa akurat dan berapa besar nilai kesalahann (*error*) dari *pattern recognition* yang dilakukan.

Berikut merupakan *source code* dalam melakukan *data splitting*

```
from sklearn.model_selection import  
train_test_split  
  
x_train, x_test, y_train, y_test =  
train_test_split(data, target, test_size=0.2)
```

Berdasarkan *source code* di atas, proses *splitting* dilakukan dengan bantuan *library* (pustaka) python yang bernama skleran. kemudian dideklarasikan empat buah variabel yang dapat dikelompokkan menjadi dua kelompok, yaitu variabel *train* dan variabel *test*. Ke empat variabel tersebut akan diisi berdasarkan variabel data dna target yang dilakukan dengan syntax *train\_test\_split* dengan rasio *test\_size* sama dengan 0,2 atau 20%.

#### 4.2.2.4. Normalisasi Data

Normalisasi dilakukan dengan melakukan *import library* sklearn dan mendeklarasikan ke empat variabel baru dengan tujuan membedakan mana variabel yang belum dan sudah ternormalisasi. Berikut merupakan *source code* dari proses normalisasi data.

```
from sklearn.preprocessing import
StandardScaler
std = StandardScaler()

std.fit(x_train)
X_train = std.transform(x_train)
std.fit(x_test)
X_test = std.transform(x_test)
std.fit(y_train)
Y_train = std.transform(y_train)
std.fit(y_test)
Y_test = std.transform(y_test)
```

Berikut perbedaan dari data yang belum ternormalisasi dan yang sudah ternormalisasi.

**Tabel 4.4 Data Sebelum Normalisasi Dari Variabel X\_train**

	<i>open</i>	<i>low</i>	<i>high</i>	<i>Close</i>
1	8200	7800	8250	7825
2	2322	2272	2322	2272
3	7625	7425	8025	7800
4	12100	12000	12175	12175
5	12100	12000	12650	12175

**Tabel 4.5 Data Setelah Normalisasi Dari Variabel X\_train**

	<i>open</i>	<i>Low</i>	<i>high</i>	<i>Close</i>
1	- 0.01785391	- 0.05073684	- 0.04682103	- 0.08404185
2	- 1.08755979	- 1.08573598	- 1.10162067	- 1.09726488
3	-0.1224951	- 0.12094752	- 0.08685644	- 0.08860345
4	1.18779451	1.16624841	1.1364478	1.10653626
5	0.69188629	0.73562286	0.73609368	0.70967689

#### 4.2.3. Implementasi SVM & KNN Serta Memilih *Kernel Function*

Dalam proses ini akan melakukan *import* algoritma SVM dan KNN dari *library* python yaitu sklearn, sekaligus memilih salah satu jenis *kernel function* yang akan digunakan. Sebagai pengingat, kernel yang digunakan yaitu linear, *polynomial*, gaussian/RBF, dan *sigmoid* secara berurutan. Berikut merupakan *source code* untuk mengimplementasikan kedua algoritma dari *library* sklearn

```
from sklearn.svm import SVR
svr = SVR()
from sklearn.neighbors import KNeighborsRegressor
knn = KNeighborsRegressor()
```

Penggunaan *kernel function* jenis linear sudah diaplikasikan sebagai *default* dari konfigurasi penggunaan algoritma SVM. Sedangkan untuk melakukan pemilihan *kernel function* jenis

*polynomial*, *gaussian/RBF*, dan *sigmoid* secara berurutan adalah sebagai berikut.

```
svr = SVR(kernel='poly')
svr = SVR(kernel='rbf')
svr = SVR(kernel='sigmoid')
```

#### 4.2.4. Menyiapkan *Multioutput Machine Learning Model*

Untuk bisa menghasilkan *output* lebih dari satu dalam melakukan prediksi, perlu digunakan salah satu modul yang tersedia pada *library* *sklearn* yang bernama *multioutput regressor*. Kemudian *multioutput regressor* akan digabungkan dengan algoritma SVM dan KNN sehingga hasil prediksi yang diperoleh bisa menghasilkan *output* yang diinginkan. Berikut merupakan *source code* untuk menggunakan *multioutput regressor*.

```
from sklearn.multioutput import MultiOutputRegressor
mor = MultiOutputRegressor(svr)
moknn = MultiOutputRegressor(knn)
```

#### 4.2.5. Melatih Model *Machine Learning*

Setelah berhasil mengeksekusi *source code* sebelumnya, langkah selanjutnya adalah melatih model *machine learning*. Dalam pelatihannya, model *machine learning* membutuhkan dua parameter, yaitu *data train*, dan *data test*. Untuk menghasilkan *output* hasil prediksi tiga hari kedepan, dibutuhkan *data test* dengan bentuk struktur dan data dari harga saham tiga hari kedepan juga, hal tersebut sudah berhasil dirancang dalam tahap *data manipulating*.

Dalam proses pelatihannya, masing masing harga saham yang akan diprediksi dilatih dengan harga saham itu sendiri. Yang dimaksud dari pernyataan sebelumnya adalah untuk memprediksi harga saham BBHI, maka model *machine learning* yang akan memprediksi harga saham BBHI akan dilatih dengan data saham BBHI saja. Hal yang sama terjadi dengan model *machine learning* yang akan memprediksi saham ARTO. Hal tersebut dilakukan untuk memisahkan kedua data agar tidak terjadinya kebingungan (*human error*) dalam memperoleh hasil prediksi.

Berikut merupakan *source code* untuk melakukan pelatihan model *machine learning*.

```
mor = mor.fit(X_train, Y_train)
moknn = moknn.fit(X_train, Y_train)
```

Perlu diingat bahwa model *machine learning* dengan algoritma SVM dan KNN melakukan pelatihan dengan menggunakan data yang telah dinormalisasi, dan akan menghasilkan *output* dalam nilai normalisasi juga.

#### 4.2.6. Melakukan Prediksi

Langkah terakhir dalam proses pengujian model *machine learning* yaitu melakukan prediksi. Kedua saham akan diprediksi dengan masing masing algoritma yaitu SVM dan KNN, serta akan mengalami pengulangan untuk algoritma SVM untuk dilakukan prediksi dengan menggunakan *kernel function* yang berbeda.

Berikut merupakan *source code* untuk melakukan prediksi menggunakan *multioutput regressor* SVM dan KNN.

```
y_pred = mor.predict(X_test)
```

```
knn_pred = moknn.predict(X_test)
```

Variabel `y_pred` merupakan variabel hasil prediksi yang dihasilkan oleh algoritma SVM dan `knn_pred` merupakan variabel hasil prediksi yang dihasilkan oleh algoritma KNN. Dalam melakukan prediksi, dibutuhkan satu parameter dari *data test* yaitu variabel `X_test`. Hal ini bertujuan untuk menghasilkan prediksi harga saham dari kedua algoritma yang nantinya akan disamakan dengan variabel `Y_test` untuk melakukan uji performa.

### 4.3. Analisa Hasil

Pada tahapan ini akan dilakukan analisa hasil dari perolehan prediksi harga saham BBHI dan ARTO dengan menggunakan algoritma SVM dengan *kernel function* linear, *polynomial*, gaussian/RBF, dan *sigmoid* serta algoritma KNN. Hasil yang dicantumkan pada pembahasan ini merupakan hasil rata rata dari sepuluh kali percobaan dalam memprediksi harga saham dari kedua saham bank digital, yaitu saham ARTO dan BBHI. Analisa hasil dilakukan dengan cara membandingkan MSE dan MAE dari masing masing perolehan hasil. Proses perbandingan akan dibagi menjadi tiga, yaitu perbandingan hasil prediksi algoritma SVM secara linear dengan *data test* nya sendiri, perbandingan perolehan hasil prediksi dari masing masing *kernel function* SVM, dan perbandingan perolehan hasil prediksi algoritma SVM dan KNN.

#### 4.3.1. Perbandingan Hasil Prediksi SVM Linear Dengan *Data Test*

Berikut merupakan nilai MSE dan MAE yang diperoleh dari model *machine learning* algoritma SVM dengan *kernel* linear:

**Tabel 4.6 Perolehan MSE dan MAE dari SVM *Kernel* Linear.**

linear	<i>field</i>	MSE	MAE
	close+1	0.0335	0.0589

ARTO	close+2	0.0110	0.0765
	close+3	0.0157	0.0890
BBHI	close+1	0.0110	0.0698
	close+2	0.0188	0.0890
	close+3	0.0269	0.1054

Berdasarkan nilai MSE dan MAE yang diperoleh, terdapat perbedaan nilai MSE dan MAE dari hasil prediksi data saham ARTO dengan BBHI. Saham ARTO memiliki nilai MSE yang lebih besar dari BBHI pada *field* close+1, namun mengungguli BBHI pada *field* close+2 dan close+3. Kemudian nilai MAE yang diperoleh dari saham ARTO lebih kecil pada ketiga *field* dibandingkan dengan nilai MAE yang diperoleh dari saham BBHI.

Namun, kedua nilai tersebut dapat dikatakan sebagai nilai yang kecil, sehingga model *machine learning* dengan algoritma SVM *kernel* linear memiliki kinerja yang cukup akurat.

#### 4.3.2. Perbandingan Masing-Masing *Kernel Function* SVM

Hasil perolehan nilai MSE dan MAE yang akan dipaparkan dibawah merupakan nilai MSE dan MAE yang diperoleh menggunakan model *machine learning* SVM *kernel polynomial*, gaussian/RBF, dan *sigmoid* secara berurutan. Tidak disertakannya nilai MSE dan MAE dari *kernel* linear dikarenakan nilai tersebut sudah dipaparkan pada sub bab sebelumnya, Berikut merupakan hasil perolehan nilai MSE dan MAE dari masing masing *kernel function* algoritma SVM.

**Tabel 4.7 Perolehan MSE dan MAE dari SVM *Kernel Polynomial*.**

<i>poly</i>	<i>field</i>	MSE	MAE
ARTO	close+1	0.1853	0.3308
	close+2	0.1957	0.3535
	close+3	0.2026	0.3575



BBHI	close+1	0.6216	0.4549
	close+2	0.6320	0.4583
	close+3	0.6414	0.4636

Dengan menggunakan *kernel polynomial* didapatkan nilai MSE dan MAE yang lebih besar dibandingkan dengan hasil yang didapatkan menggunakan *kernel linear*. Terdapat perbedaan yang signifikan pada nilai MSE yang diperoleh dari hasil memprediksi saham ARTO dan saham BBHI, dimana nilai MSE yang diperoleh dari hasil prediksi saham ARTO berada pada *range* 0.18 sampai dengan 0.20 sedangkan nilai MSE dari saham BBHI berada pada *range* 0.62 sampai dengan 0.64.

Nilai MAE yang diperoleh dengan menggunakan *kernel polynomial* juga lebih besar dari hasil yang diperoleh menggunakan *kernel linear*. Dan untuk perbedaan nilai MAE dari saham ARTO memiliki nilai yang lebih kecil dari nilai MAE dari saham BBHI. Dimana saham ARTO memiliki nilai MAE dengan *range* 0.33 sampai dengan 0.35 sedangkan saham BBHI memiliki nilai MAE dengan *range* 0.45 sampai dengan 0.46.

**Tabel 4.8 Perolehan MSE dan MAE dari SVM *Kernel Gaussian/RBF*.**

RBF	<i>field</i>	MSE	MAE
ARTO	close+1	0.0136	0.0572
	close+2	0.0103	0.0701
	close+3	0.0138	0.0859
BBHI	close+1	0.0136	0.0777
	close+2	0.0202	0.0791
	close+3	0.0359	0.0359

Berdasarkan table 4.5 di atas, dapat dilihat bahwa nilai MSE dan MAE yang diperoleh dengan menggunakan *kernel gaussian/RBF* memiliki nilai yang kecil, dan hampir sama dengan hasil yang diperoleh menggunakan *kernel linear*. Perbandingan nilai MSE dan MAE dari hasil prediksi saham ARTO memiliki nilai yang lebih kecil dari hasil prediksi saham BBHI pada hampir semua *field*, kecuali nilai MAE pada *field close+3*, dimana hasil prediksi BBHI memiliki nilai MAE yang lebih kecil dibandingkan dengan hasil prediksi ARTO. Lalu untuk sisa nilai MSE dan MAE selain pada nilai MAE dalam *field close+3*, saham ARTO memiliki nilai MSE dan MAE dalam *range* 0.010 sampai dengan 0.013 dan 0.05 sampai dengan 0.08 sedangkan hasil prediksi saham BBHI memiliki nilai MSE dan MAE dalam *range* 0.01 sampai dengan 0.03 dan 0.03 sampai dengan 0.7

**Tabel 4.9 Perolehan MSE dan MAE dari SVM *Kernel Sigmoid*.**

<i>sigmoid</i>	<i>field</i>	MSE	MAE
ARTO	close+1	281.4252	14.2342
	close+2	281.3625	14.2327
	close+3	281.3396	14.2320
BBHI	close+1	293.8726	13.0513
	close+2	327.5678	13.0601
	close+3	327.3549	47.1768

Berdasarkan nilai MSE dan MAE yang diperoleh dengan menggunakan *kernel sigmoid*, dapat disimpulkan bahwa *kernel sigmoid* bukanlah *kernel* yang cocok digunakan dalam memprediksi harga saham dikarenakan nilai MSE dan MAE yang didapatkan sangat besar.

#### 4.3.3. Perbandingan Algoritma SVM Dengan KNN

Berikut merupakan nilai MSE dan MAE yang diperoleh dengan menggunakan model *machine learning* algoritma KNN.

**Tabel 4.10 Perolehan Nilai MSE dan MAE dari Algoritma KNN.**

KNN	<i>field</i>	MSE	MAE
ARTO	close+1	0.0066	0.0589
	close+2	0.0116	0.0787
	close+3	0.0248	0.0890
BBHI	close+1	0.0099	0.0536
	close+2	0.0227	0.0800
	close+3	0.0303	0.1033

Nilai MSE untuk kedua saham yang diperoleh dengan menggunakan model *machine learning* algoritma KNN pada *field* close+1 jauh lebih kecil dibandingkan dengan perolehan nilai MSE dengan menggunakan seluruh *kernel function* algoritma SVM. Tetapi, nilai MSE untuk kedua saham dari algoritma KNN pada *field* close+2 dan close+3 lebih besar dibandingkan dengan perolehan nilai MSE dari algoritma SVM *kernel* gaussian/RBF. Kemudian untuk nilai MAE, algoritma KNN memiliki nilai yang lebih kecil pada *field* close+1 dalam saham BBHI dibandingkan dengan semua hasil perolehan *kernel* SVM.

#### **4.4. Evaluasi Dan Penarikan Kesimpulan**

Setelah melakukan perbandingan dari masing masing algoritma dan *kernel function*, langkah terakhir yang dilakukan adalah evaluasi dan penarikan kesimpulan untuk menentukan algoritma dan konfigurasi yang paling sesuai untuk melakukan prediksi harga saham Bank Jago (ARTO) dan Allo Bank (BBHI).

Dalam proses ini akan dilakukan eliminasi dari dua *kernel function* yaitu *polynomial* dan *sigmoid* dikarenakan hasil perolehan nilai MSE dan MAE dari masing masing *kernel function* memiliki selisih yang besar dari algoritma dan *kernel function* yang lainnya. Proses penarikan kesimpulan akan dilakukan

berdasarkan pernyataan yang telah diberikan pada sub-bab sebelumnya yang disajikan dengan menggunakan tabel agar penyajian lebih terstruktur. Berikut merupakan tabel yang akan dijadikan acuan dalam evaluasi dan penarikan kesimpulan.

**Tabel 4.11 Perbandingan MSE KNN dengan *Kernel* SVM.**

	<i>field</i>	KNN	LINEAR	GAUSSIAN/RBF
		MSE	MSE	MSE
ARTO	close+1	0.0066	0.0335	0.01366
	close+2	0.0116	0.0110	0.01037
	close+3	0.0248	0.0157	0.01383
BBHI	close+1	0.0099	0.0110	0.01367
	close+2	0.0227	0.0188	0.02029
	close+3	0.0303	0.0269	0.03595

**Tabel 4.12 Perbandingan MAE KNN dengan *Kernel* SVM.**

	<i>field</i>	KNN	LINEAR	GUASSIAN/RBF
		MAE	MAE	MAE
ARTO	close+1	0.0589	0.0589	0.0572
	close+2	0.0787	0.0765	0.0701
	close+3	0.0890	0.0890	0.0859
BBHI	close+1	0.0536	0.0698	0.0777
	close+2	0.0800	0.0890	0.0791
	close+3	0.1033	0.1054	0.0359

Berdasarkan kedua tabel di atas dapat disimpulkan bahwa untuk memprediksi harga saham ARTO pada waktu hari t+1 atau yang diwakilkan oleh *field* close+1, algoritma yang cocok digunakan adalah KNN atau SVM dengan *kernel* gaussian/RBF karena menghasilkan nilai MSE dan MAE yang lebih akurat. Tetapi untuk melakukan prediksi harga saham pada hari t+2 dan t+3 atau yang diwakilkan oleh *field* close+2 dan close+3, algoritma yang cocok digunakan adalah SVM dengan *kernel* gaussian/RBF, karena menghasilkan

nilai MSE dan MAE yang lebih kecil dibandingkan dengan algoritma dan *kernel* lainnya.

Sementara untuk memprediksi harga saham BBHI pada hari  $t+1$  atau yang diwakilkan oleh *field*  $close+1$ , algoritma yang paling sesuai untuk digunakan adalah algoritma KNN karena perolehan MSE dan MAE yang lebih kecil. Dan untuk hari  $t+2$  dan  $t+3$  atau yang diwakilkan dengan  $close+2$  dan  $close+3$ , algoritma yang paling sesuai yaitu SVM dengan *kernel* linear atau SVM dengan *kernel* gaussian/RBF. Hal tersebut dikarenakan nilai MSE paling kecil diperoleh dengan menggunakan algoritma SVM *kernel* linear, dan nilai MAE paling kecil diperoleh dengan algoritma SVM *kernel* gaussian/RBF

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Setelah melakukan penelitian untuk memprediksi harga saham ARTO dan BBHI dari 1570 data, diperoleh kesimpulan antara lain:

1. Algoritma SVM dan KNN adalah algoritma yang dapat digunakan dalam memprediksi harga saham, dengan perolehan *range* MSE dan MAE dari algoritma SVM sebesar 0.0110 sampai dengan 0.0335 dan 0.0589 sampai dengan 0.1054, dan perolehan *range* nilai MSE dan MAE sebesar 0.066 sampai dengan 0.0303.
2. Telah diperoleh performa yang dihasilkan oleh algoritma SVM yaitu berupa peringkat dari *kernel function* berdasarkan nilai MSE dan MAE dalam memprediksi harga saham ARTO dan BBHI. Berikut peringkat *kernel function* yang didapatkan: gaussian/RBF, linear, *polynomial*, dan *sigmoid*.
3. Algoritma KNN adalah algoritma yang paling akurat digunakan dalam memprediksi harga saham pada hari  $t+1$  atau *field close+1* dan algoritma SVM dengan *kernel* gaussian/RBF untuk memprediksi harga saham pada  $t+2$  dan  $t+3$  atau *field close+2* dan *close+3*.

#### 5.2. Saran

Setelah melakukan penelitian, terdapat beberapa saran yang dapat dilakukan untuk penelitian lebih lanjut. Saran tersebut diantaranya adalah:

1. Digunakan beberapa pilihan algoritma lainnya dalam memprediksi harga saham.
2. Melakukan konfigurasi yang berbeda dalam melakukan *data splitting*, *data training*, dan dilakukannya pengembalian nilai dari data yang telah dinormalisasi dalam bentuk penyajian hasil prediksi.

3. Membuat output yang dihasilkan oleh model machine learning lebih dari tiga.
4. Mengembangkan aplikasi untuk memprediksi harga saham yang mengimplementasikan model machine learning yang telah dikembangkan dalam penelitian ini.

## DAFTAR PUSTAKA

- Abijono, H., Santoso, P., & Anggreini, N. L. (2021). Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *G-Tech: Jurnal Teknologi Terapan*, 4(2), 315-318.
- Afifah, Luthfia. Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi. Diakses pada 10 April 2023. <https://ilmudatapy.com/algoritma-k-nearest-neighbor-knn-untuk-klasifikasi/>.
- Ajaib. (2020). Cara Trading Saham untuk Pemula Agar Tidak Salah Langkah. Diakses pada 10 April 2023. <https://ajaib.co.id/cara-trading-saham-untuk-pemula-agar-tidak-salah-langkah/>.
- Allo Bank. Profil Allo Bank. Diakses pada 10 April 2023. <https://www.allobank.com/about-us/profiles>.
- Bank Jago. Tentang Jago. Diakses pada 10 April 2023. <https://jago.com/id/whoweare>.
- Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*.
- Darmadji, T., & Fakhruddin, H. M. (2012). Pasar Modal di Indonesia, Edisi 3, Salemba Empat.
- Fadilah, W. R. U., Agfiannisa, D., & Azhar, Y. (2020). Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine. *Fountain of Informatics Journal*, 5(2), 45-51.
- Google. Welcome to Colab! Diakses pada 10 April 2023. <https://colab.research.google.com/>.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.



Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises* (pp. 1-227). Lutz: Dave Kuhlman.

Nufus, W. H. (2022). Jaksa: Korban Binomo Indra Kenz 144 Orang, Total Kerugian Rp 83,3 M. Diakses pada 10 April 2023. <https://news.detik.com/berita/d-6231006/jaksa-korban-binomo-indra-kenz-144-orang-total-kerugian-rp-833-m>.

OJK. (2020). Daftar Entitas Investasi Ilegal Yang Ditangani Satgas Waspada Investasi. Diakses pada 10 April 2023. <https://www.ojk.go.id/id/beritadankegiatan/siaranpers/Documents/Pages/Siaran-Pers--Satgas-Waspada-InvestasiKembali-Hentikan-140-Fintech-Peer-To-Peer-Lending-TanpaIzin/LAMPIRAN%2043%20Entitas%20Penawaran%20Investasi%20Tanpa%20Izin.pdf>.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.

Sari, M., Siswati, T., Suparto, A. A., Ambarsari, I. F., Azizah, N., Safitri, W., & Hasanah, N. (2022). *Metodologi penelitian*. Global Eksekutif Teknologi.

Scikit Learn. About Us. Diakses pada 10 April 2023. <https://scikit-learn.org/stable/about.html>.

Setia, Insan Cahaya. (2021). Support Vector Machine : SVM Implementation using Python. Diakses pada 10 April 2023. <https://insancs.medium.com/support-vector-machine-svm-implementation-using-python-4442e9a5babc>.

Singh, Taranvir. (2021). A practical approach to understand Machine learning Models. Diakses pada 10 April 2023. <https://medium.com/crossml/machine-learning-models-a8ce83b89fd6>

Tambunan, D. (2020). Investasi saham di masa pandemi COVID-19. *Jurnal Khatulistiwa Informatika*, 4(2), 117-123.

Umayah, R., Darna, N., & Basari, M. A. (2019). Faktor-faktor yang Mempengaruhi Harga Saham (Suatu Studi pada PT. Telkom Indonesia, Tbk yang Terdaftar di BEI Periode 2007-2017). *Business Management and Entrepreneurship Journal*, 1(4), 178-192.

Vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to design science research. *Design science research. Cases*, 1-13.

Wahyono, T. (2010). Analisis regresi dengan MS Excel 2007 dan SPSS 17. *Jakarta: Elex Media Komputindo. E-Book*.