

**PREDIKSI HARGA INDEX SAHAM GABUNGAN  
IHSG MENGGUNAKAN METODE SUPPORT  
VECTOR MACHINE**

**PROPOSAL TUGAS AKHIR**



**Oleh:**

**BOYKE AGUNG NUGRAHA**

**162019009**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNOLOGI INDUSTRI  
INSTITUT TEKNOLOGI NASIONAL  
BANDUNG  
2023**

**PROGRAM STUDI SISTEM INFORMASI**  
**LEMBAR PENGESAHAN**  
**TUGAS AKHIR**  
**PREDIKSI HARGA INDEX SAHAM GABUNGAN IHSG**  
**MENGGUNAKAN METODE SUPPORT VECTOR MACHINE**



Yang dipersiapkan dan disusun oleh :

**Boyke Agung Nugraha  
162019009**

Telah dipertahankan di depan Dewan Pengaji  
Pada Tanggal **3 Agustus 2023**

Susunan Dewan Pengaji  
Dosen Pembimbing                              Dosen Pengaji

**Nur Fitrianti Fahrudin, S.Kom., M.T.**

NIP. 120160503

**Dosen Pengaji**

NIP.

Tugas Akhir ini telah diterima sebagai salah satu persyaratan  
Untuk memperoleh gelar Sarjana Komputer (S.Kom)  
Pada Tanggal **3 September 2023**

Ketua Program Studi Sistem Informasi

**Mira Musrini Barmawi, S.Si., M.T**

NIP. 120070201

## KATA PENGANTAR

Puji dan syukur dipanjatkan kehadiran Tuhan Yang Maha Esa atas segala limpahan rahmat dan karunia – Nya penulis dapat menyelesaikan tugas akhir yang berjudul “PREDIKSI HARGA INDEX SAHAM GABUNGAN IHSG MENGGUNAKAN METODE SUPPORT VECTOR MACHINE”.

Penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu, mendukung, dan membina selama penggeraan tugas akhir berlangsung. Secara khusus penulis ingin mengucapkan terima kasih kepada:

1. Ibu Nur Fitrianti Fahrudin, S.Kom., M.T. selaku dosen pembimbing, yang telah membantu memberikan pengetahuan, serta arahannya dalam melakukan penelitian ini.
2. Bapak Kurnia Ramadhan Putra, S.Kom., M.T. selaku dosen wali yang telah memberikan bimbingan dan arahan selama masa perkuliahan.
3. Seluruh dosen program studi Sistem Informasi ITENAS yang telah memberikan ilmu dan nasihat selama masa perkuliahan.
4. Ibu Mira Musrini Barmawi, S. Si., M.T. selaku ketua program studi Sistem Informasi Itenas.
5. Kedua Orang Tua dan keluarga yang telah memberi dukungan dalam menyelesaikan laporan ini.
6. Semua pihak yang tidak dapat disebutkan satu – persatu.

Penulis juga menyadari sepenuhnya bahwa dalam laporan tugas akhir ini masih banyak kekurangan dan jauh dari kata sempurna, oleh sebab itu penulis mengharapkan kritik dan saran serta usulan yang bersifat membangun untuk perbaikan penulis di masa yang akan datang.

Bandung, Februari 2023

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN .....	ii
KATA PENGANTAR .....	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR .....	vi
DAFTAR TABEL.....	vii
DAFTAR RUMUS .....	viii
DAFTAR SINGKATAN DAN ISTILAH .....	ix
BAB I PENDAHULUAN .....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah.....	3
1.3    Tujuan .....	4
1.4    Ruang Lingkup .....	4
1.5    Sistematika penulisan .....	4
BAB II KAJIAN PUSTAKA .....	6
2.1    Literatur Map .....	6
2.2 <i>Knowledge Discovery in Database</i> .....	7
2.3    Data Mining .....	8
2.4 <i>Cross-Industry Standard Process for Data Mining (CRISP-DM)</i> .....	10
2.5    Prediksi .....	12
2.6    Saham.....	13
2.7    Analisis Harga Saham.....	14
2.8    Time Series (Data Deret Waktu) .....	15
2.9    Stasioneritas .....	15
2.10    Uji stasioneritas.....	15
2.10.1    Metode visualisasi grafik .....	15
2.10.2    Uji Correlogram .....	15
2.10.3    Uji akar unit.....	16

2.11	Differencing .....	16
2.12	Support Vector Machine .....	16
2.13	Evaluasi.....	18
2.13.1	Root Mean Square Error .....	18
2.13.2	Mean Absolute Percentage Error .....	19
	BAB III METODOLOGI PENELITIAN .....	20
3.1	Alur Penelitian .....	20
3.2	Teknik Analisis Data Harga Saham.....	21
3.3	Data Collection .....	21
3.4	Data Understanding .....	23
3.5	Data Preparation .....	24
3.6	Modeling.....	25
3.7	Prediction .....	25
3.8	Evaluation .....	25
	DAFTAR PUSTAKA .....	26

## DAFTAR GAMBAR

Gambar 2.1 Literatur Map.....	6
Gambar 2. 2 CRISP-DM Sumber : (Purnama, 2020) .....	10
Gambar 2. 7 Root Mean Square Error .....	19
Gambar 2. 8 Mean Absolute Percentage Error .....	19
Gambar 3. 1 Alur Penelitian.....	20
Gambar 3. 2 Sumber Data.....	22
Gambar 3. 3 ADF Test .....	24

## **DAFTAR TABEL**

Tabel 3. 1 Data Understanding.....	23
------------------------------------	----

## **DAFTAR RUMUS**

## DAFTAR SINGKATAN DAN ISTILAH

No	Istilah	Keterangan	Pemakaian Pada Halaman
1	Prediksi	Prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu.	
2	Saham	Saham memiliki arti "hak yang dimiliki orang (pemegang saham) terhadap perusahaan berkat penyerahan bagian modal sehingga dianggap berbagi dalam pemilikan dan pengawasan".	
3	IHSG	IHSG adalah singkatan dari Indeks Harga Saham Gabungan. Secara internasional disebut juga dengan Indonesia Composite Index (ICI) dan ada juga yang menyebutnya dengan IDX Composite. Dengan kata lain, IHSG adalah indeks pasar saham yang digunakan Bursa Efek Indonesia. Indeks saham berisi daftar semua saham yang diperjualbelikan di Bursa Efek Indonesia (BEI).	
4	Histori	Catatan umat manusia atau peradaban dunia, tentang perubahan-perubahan yang terjadi serta mencatat dari beberapa peristiwa yang mampu dibuktikan dengan kenyataan (fakta).	
5	Stasioner	Stasioner adalah jika tidak ada tren didalam data deret waktu.	

- 6 ADF test Augmented Dickey Fuller Test adalah pengujian yang dilakukan terhadap data deret waktu (*time series*) untuk mengetahui apakah data deret waktu tersebut stasioner atau tidak. Jika nilai absolut statistik Augmented Dickey-Fuller (ADF) lebih besar dari nilai kritisnya, maka data yang diamati menunjukkan stasioner dan jika sebaliknya nilai absolut statistik Augmented Dickey-Fuller (ADF) lebih kecil dari nilai kritisnya maka data tidak stasioner.
- 7 EDA Exploratory Data Analysis adalah sebuah proses kritis dalam melakukan investigasi awal pada data dengan tujuan menemukan pola, anomali, menguji hipotesis dan dapat memeriksa asumsi dengan bantuan statistik ringkasan kemudian representasi grafis (visualisasi).
- 8 Differencing *Differencing* adalah menghitung perubahan atau selisih nilai observasi. *Differencing* digunakan untuk menstasionerkan data yang tidak stasioner.
- 9 Regresi *Regresi* adalah suatu teknik analisis untuk mengidentifikasi relasi atau hubungan diantara dua variabel atau lebih.
- 10 ROI Return On Investment
- 11 *Time series*
- 12 KDD
- 13 CRISP-DM
- 14 ANN
- 15 SVM

- 16 RMSE Root Mean Square Error, Model prediksi dikatakan paling baik apabila nilai RMSE adalah mendekati 0 (nol).
- 17 MAE Mean Absolute Error, Model prediksi dikatakan paling baik apabila nilai MAE adalah mendekati 0 (nol).
- 18 MAPE Mean Absolute Percentage Error, Nilai MAPE dapat diinterpretasikan atau ditafsirkan ke dalam 4 kategori yaitu; <10% = sangat akurat, 10-20% = baik, 20-50% = wajar, >50% = tidak akurat.

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Bursa saham merupakan instrumen keuangan jangka panjang yang dijual belikan di dalam bursa pasar saham di Indonesia(Purwanti, 2017). Saham didefinisikan sebagai kepunyaan atau kepemilikan atau tanda penyertaan atau kepemilikan seseorang atau badan usaha dalam sebuah perusahaan atau perseroan terbatas(Suyanto & Nursanti, 2017). Perdagangan saham pada lantai bursa mempunyai daya tarik tersendiri karena bisa menghasilkan *Return On Investment* yang besar, Menurut (Maulita & Arifin, 2018) *Return On Investment* atau *Return On Assets* merupakan kemampuan untuk menghasilkan laba dari aktivitas yang sedang dilakukan. Pertimbangan atau keputusan untuk pembelian biasanya dilakukan dengan melihat prospek perusahaan berdasarkan histori perusahaan sebelumnya(Wibowo, 2011). Sebelum memutuskan untuk membeli sebuah saham pelaku bisnis atau investor harus melaksanakan analisa dengan melihat grafik harga saham sebagai representatif histori perusahaan sebelumnya.

**Pergerakan harga saham acak.**  
Bursa saham merupakan suatu yang kompleks karena banyak sekali faktor yang akan mempengaruhi pergerakan harga saham. Pergerakan harga saham pada studi sebelumnya seringkali dianggap sebagai pergerakan yang acak dan tidak beraturan (Lusiana et al., 2018). Tetapi menurut teori *Efficient Market Hypothesis* yang dikemukakan oleh(Shleifer, 2000) pergerakan harga saham tidak sepenuhnya acak dan sepenuhnya tidak beraturan namun kompleksitasnya saja yang sangat tinggi dan sering kali berubah rubah. Belakangan ini didapatkan beberapa literatur yang mengemukakan bahwa pergerakan harga saham memungkinkan untuk di prediksi(Sugandhi et al., 2019)(Sugandhi et al., 2019). Dengan adanya prediksi ini memungkinkan pelaku bisnis yang ada di bidang ini untuk mengetahui pergerakan harga saham yang bisa membantu analisa pelaku bisnis dalam

*solusi bisa ANN bisa machine learning.*

pengambilan keputusan. Namun dalam prakteknya prediksi harga saham tidak mudah dan melibatkan banyak cabang keilmuan(Rosyida et al., 2020). Untuk melakukan prediksi ada salah satu metode Bernama *Artificial Neural Network* (ANN) (Kurniawan, 2018). Metode ANN berbasis dari penggunaan Neural Network merupakan metode yang banyak digunakan dalam melakukan Forecasting atau peramalan pada harga saham. Pada awalnya metode ini menjadi salah satu alternatif dalam pemecahan masalah ekonomi yang kompleks dimana model tradisional tidak dapat menanganinya. ANN bisa mempelajari dan mengidentifikasi pola yang berkorelasi antara set data yang di inputkan dengan nilai target yang di inginkan. Namun sebelum bisa melakukan prediksi diperlukan training terlebih dahulu sebagai learning prosess untuk ANN. Setelah lama digunakan dalam Forecasting harga saham(Imelda A.Muis & Muhammad Affandes, 2015). Mencoba membandingkan ANN dengan metode lainnya yaitu hybrid dengan berbasiskan Support Vector Machine (SVM). Ternyata akurasi prediksi metode berbasis SVM jauh lebih baik dari metode berbasis ANN. Support Vector Machine (SVM) adalah sistem pembelajaran dengan memakai ruang hipotesis yang berisikan fungsi-fungsi linear didalam sebuah fitur yang mempunyai dimensi tinggi dan dapat dilatih dengan menggunakan algoritma(Monika Parapat & Tanzil Furqon, 2018). SVM adalah metode yang memiliki performa lebih baik terutama dalam simulasi potensi prediksi harga dalam pasar modal jika dibandingkan dengan metode lainnya. Ada beberapa kelebihan SVM daripada ANN, diantaranya adalah hasil solusi yang didapatkan SVM bersifat global dan unik tetapi memiliki interpretasi geometrik yang mudah dibaca (Pramansah et al., 2022).

Dalam melakukan prediksi saham dibutuhkan data harga saham atau histori harga saham sebelumnya. Data yang dibutuhkan untuk melakukan prediksi adalah data *time series*. Data *time series* adalah data dari serangkaian pengamatan yang berurutan berdasarkan waktu dengan jarak waktu yang sama(Mahfud Al et al., 2020). Jenis data *time series* seringkali ditemui dalam keseharian dikarenakan data tersebut di dapatkan melalui interval waktu harian, mingguan bahkan bulanan.

*↳ sifasi'*

Data *time series* terbagi menjadi dua domain yang pertama adalah domain waktu dan domain frekuensi. Domain waktu mendalam tentang kestasioneran data, signifikansi auto korelasi, peramalan dan penaksiran regresi model deret waktu. Sedangkan domain frekuensi mendalam tentang frekuensi data atau pola yang tersembunyi pada data yang tercatat berdasarkan waktu. Tujuan dari prediksi menggunakan data *time series* adalah mengetahui hal istimewa atau pola yang ada didalam data tersebut. Data *time series* yang penulis gunakan adalah data pergerakan saham yang ada didalam bursa Indonesia (Indeks Harga Saham Gabungan IHSG) yang tercatat sejak tahun 1990 sampai tahun 2022. Pada *Exploratory Data Analysis* (EDA) yang penulis lakukan pada data *time series* IHSG ditemukan masalah pada data *time series* IHSG yaitu kestasioneran data IHSG yang kurang baik yang menyebabkan data *time series* IHSG harus di transformasikan terlebih dahulu sebelum dilakukan prediksi.

Data *time series* atau deret waktu yang tidak stationer akan menimbulkan hasil regresi palsu dan menyebabkan hasil prediksi kurang baik. Stasioneritas adalah tidak adanya pertumbuhan dan penurunan data dalam kurun waktu tertentu. Data bisa dikatakan stationer jika pola data tersebut berada di keseimbangan nilai rata rata konstan. Data *time series* dikatakan stationer jika tidak ada tren didalam data deret waktu tersebut (Sherly Aktivani, 2021).

Berdasarkan latar belakang ini dengan mengadopsi data sampel yaitu data *time series* IHSG yang memiliki kestasioneran data yang kurang baik maka dapat dilakukan penelitian untuk mencari dan menangani ketidakstasioneran data dan prediksi menggunakan metode SVM, yang pada akhirnya bisa menghasilkan prediksi yang lebih akurat dan diharapkan bisa membantu investor dan pelaku bisnis dalam bursa saham dalam pengambilan keputusan bisnis.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang ada, permasalahan yang diangkat dalam penelitian ini adalah sebagai berikut:

L D solusi ???

1. Bagaimana mengatasi data non S.<sup>4</sup>
2. ~ II — akurasi supt dlm  
Mem prediksi saham .
3. ~ a — solusi dapat meningkatkan SVMT  
Bagaimana cara melakukan prediksi dengan data yang memiliki kestasioneran data yang kurang baik dan metode *support vector machine* pada harga saham IHSG di indonesia?

#### 1.3 Tujuan

*Sampaikan dg rumus masalah*

Tujuan dari penelitian Tugas Akhir ini adalah untuk :

Untuk menangani data yang memiliki kestasioneran data yang kurang baik dan melakukan prediksi harga saham IHSG di indonesia dengan menggunakan metode *support vector machine*.

#### 1.4 Ruang Lingkup

Ruang lingkup dalam penelitian kali ini adalah, data harga saham yang dipakai adalah harga saham Index Saham Gabungan (IHSG) pertimbangannya adalah IHSG merupakan index yang perhitungannya mewakili semua perusahaan yang tercatat di bursa efek indonesia, jadi diharapkan IHSG bisa mewakili pergerakan harga saham di Indonesia.

#### 1.5 Sistematika penulisan

Sistem penulisan yang digunakan dalam pembuatan laporan ini adalah sebagai berikut:

### BAB I : PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, tujuan, tujuan penulisan laporan, batasan masalah dan sistematika penulisan laporan yang dilakukan penulis di program studi Sistem Informasi Institut Teknologi Nasional Bandung.

### BAB II : LANDASAN TEORI

Bab ini menjelaskan mengenai teori – teori yang dipakai dan digunakan dalam melakukan penelitian ini.

### BAB III : METODOLOGI PENELITIAN

Pada bagian kali ini adalah bagian dimana penjelasan mengenai metode yang digunakan dalam penelitian, mulai dari studi literatur dan perumusan masalah, analisis, dan pengumpulan data.

#### **BAB IV : ANALISIS DAN PERANCANGAN**

Dalam bab ini menjelaskan mengenai hasil dari analisis prediksi menggunakan metode yang sudah dijelaskan pada bab sebelumnya.

#### **BAB V : PENUTUP**

Pada bab kali ini menyajikan kesimpulan dan saran penelitian yang disampaikan oleh penulis.

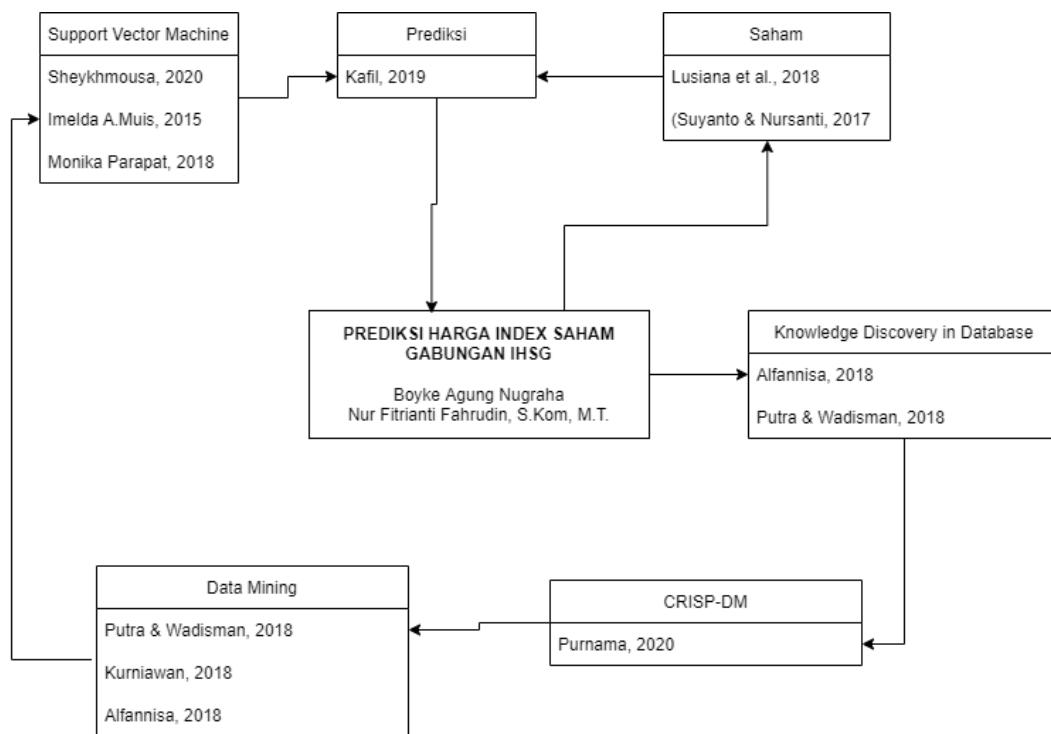
## BAB II

### KAJIAN PUSTAKA

Pada bab kali ini berisi tentang kajian pustaka yang diambil dari beberapa sumber seperti jurnal, literatur dan penelitian terdahulu yang relevan dengan topik penelitian ini.

#### 2.1 Literatur Map

Pada penelitian ini digunakan literatur map untuk mengarahkan penelitian supaya analisis yang didapatkan bisa tepat. Pada setiap proses penelitian harus menggunakan dasar analisis sehingga bisa menghasilkan keputusan yang sesuai. Hal ini harus mempertimbangkan teori yang telah dibahas dan sumber teori lainnya, literatur Map pada penelitian ini divisualisasikan pada gambar dibawah.



Gambar 2.1 Literatur Map

## **2.2 *Knowledge Discovery in Database***

*Knowledge discovery in Database* atau KDD diartikan sebagai pengambilan informasi yang potensial dan implisit. Proses KDD melibatkan proses Data Mining (Proses untuk mencari infomasi dengan menggunakan pencarian dan relasi yang tersembunyi dalam kumpulan data yang sangat banyak) kemudian mendapatkan informasi yang mudah dipahami. KDD bisa diartikan untuk keseluruhan proses untuk mendapatkan dan bisa mengidentifikasi pola dalam sebuah kumpulan data, dimana pola dari sekumpulan data tersebut bersifat baru, bisa dimengerti dan bermanfaat (Alfannisa Annurullah Fajrin, 2018).

*Knowledge discovery in Database* adalah keseluruhan proses non-trivial untuk mencari, mendapatkan dan mengidentifikasi sebuah pola atau pattern didalam sebuah data, dimana pola data bersifat baru, sah dan bisa dimengerti. Proses KDD secara garis besar dapat dijelaskan dari beberapa uraian dibawah. Dalam KDD ada beberapa proses pemilihan data (Data Selection), pembersihan (Preprocessing/Cleaning), transformasi (Transformation), penggalian data (Data Mining), evaluasi (evaluation) (Putra & Wadisman, 2018).

### **1. *Data Selection***

Pemilihan atau penyeleksian data dilakukan dari sekumpulan atau sebuah database suatu instansi. Penyeleksian data perlu dilakukan sebelum tahap penggalian data dilakukan. Data yang sudah diseleksi akan bisa digunakan untuk proses penggalian data.

### **2. *Preprocessing / Cleaning***

Sebelum terjadinya proses penggalian data atau Data Mining dilakukan harus adanya proses yang dilakukan yaitu proses pembersihan data yang menjadi fokus dan titik berat dalam Knowledge discovery in Database. Proses pembersihan kumpulan data mencakup antara lain menghapus atau menyeleksi data yang terdapat duplikasi data, menyeleksi dan memeriksa data yang inkonsisten dan yang terakhir adalah melakukan perbaikan data yang terjadi kesalahan contohnya ada kesalahan dalam penulisan atau tipografi.

### **3. *Transformation***

Transformasi data atau coding yaitu adalah sebuah proses data yang telah dipilih dari proses sebelumnya yaitu proses pembersihan data. Proses coding transformation dalam Knowledge discovery in Database adalah proses yang kreatif dan tergantung dengan jenis atau pola yang akan dicari dan dipecahkan dalam basis data.

#### 4. *Data Mining*

Penggalian data atau Data Mining merupakan proses pencarian informasi atau pola yang menarik dalam sebuah kumpulan data yang menggunakan teknik atau metode tertentu. Metode, teknik, atau algoritma dalam penggalian data mining sangat bervariasi. Memilih algoritma atau metode yang tepat sangatlah bergantung pada tujuan pada proses keseluruhan Knowledge discovery in Database.

#### 5. *Interpretation / Evaluation*

Interpretasi atau evaluasi pada pola informasi yang didapatkan dari proses penggalian data harus disampaikan dan ditampilkan dalam bentuk yang mudah dimengerti dan bisa baca oleh pihak – pihak yang berkepentingan. Tahap interpretasi atau evaluasi ini mencakup menyeleksi apakah informasi yang didapatkan dan ditemukan dengan fakta yang ada sebelumnya.

### 2.3 Data Mining

Penggalian data atau Data Mining adalah proses untuk mencari dan mendapatkan informasi dengan mencari pola dan relasi yang tidak terlihat dalam timbunan data yang banyak. Data Mining atau sering disebut dengan *Knowledge discovery in Database* (KDD) merupakan kegiatan mencari yang terdiri dari pemakaian dan pengumpulan data historis untuk menemukan sebuah keteraturan baru dalam sekumpulan tumpukan data, hubungan atau pola data yang berukuran besar. Hasil dari Data Mining bisa digunakan dimasa yang akan datang sebagai pengambilan keputusan yang baru(Kurniawan, 2018).

Data Mining adalah sebuah inti dari proses *Knowledge discovery in Database* (KDD), meliputi algoritma yang mengekstrak data, membuat model dan mencari pola dalam sebuah tumpukan data yang sebelumnya belum ditemukan. KDD bersifat otomatis, bisa didefinisikan sebagai sebuah proses pengorganisasian untuk

mengidentifikasi yang berguna untuk penemuan pola dari kumpulan daya yang kompleks. Data Mining adalah salah satu jawaban untuk menyelesaikan dengan cara menganalisa daya yang sudah ada sebelumnya, dimana data-data tersebut sudah tersimpan pada komputer(Putra & Wadisman, 2018).

Data Mining bukan suatu bidang yang baru sama sekali. Data Mining bertujuan untuk membuat perbaikan teknik yang sudah ada (tradisional) sehingga bisa melakukan pengolahan data; Jumlah data yang jumlahnya sangat banyak, Data yang berbeda beda sifat, Dimensi data yang banyak dan kompleks. Berikut pengelompokan penggalian data atau data mining menjadi beberapa bagian yaitu (Alfannisa Annurullah Fajrin, 2018):

#### 1. Klasifikasi

Klasifikasi adalah mendapatkan target variabel kategori. Contohnya ; penggolongan masyarakat berdasarkan umur dan bisa di kategorikan menjadi beberapa kategori contohnya ; lansia, remaja, dan balita

#### 2. Estimasi

Estimasi adalah jenis didalam data mining yang hampir mirip dengan klasifikasi, yang membedakannya adalah kecuali variabel target estimasi lebih mengarah ke numerik daripada kearah kategori.

#### 3. Deskripsi

Deskripsi adalah cara untuk mendapatkan visual pola dan kecenderungan yang ada didalam suatu data yang dimiliki.

#### 4. Pengklasteran

Pengklasteran adalah mengelompokan record, memperhatikan atau melakukan pengamatan dan membentuk kelas objek yang memiliki kemiripan.

#### 5. Prediksi

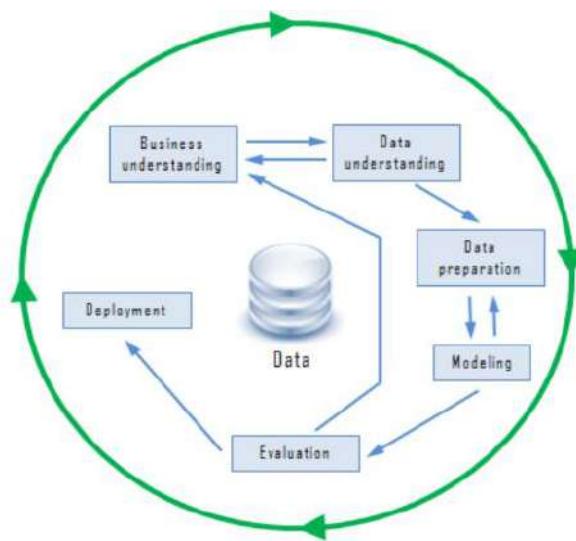
Prediksi adalah menerka sebuah hasil atau nilai yang belum diketahui sebelumnya dan berfungsi juga untuk memperkirakan sebuah nilai di masa depan.

#### 6. Asosiasi

Asosiasi memiliki fungsi untuk menemukan atribut yang sering muncul dalam satu waktu. Didalam bisnis sebenarnya biasanya disebut dengan analisis keranjang belanja.

#### **2.4 Cross-Industry Standard Process for Data Mining (CRISP-DM)**

*CRISP-DM* adalah salah satu metodologi data mining yang dibuat oleh himpunan perusahaan yang ada di eropa pada tahun 1996 dan ditetapkan sebagai standar baru dalam suatu proses data mining(Purnama, 2020).



Gambar 2. 2 CRISP-DM Sumber : (Purnama, 2020)

Menurut (Purnama, 2020) proses CRISP-DM terdiri dari enam fase yaitu:

1. Fase Pemahaman Bisnis (Business Understanding Phase)

Fase pertama adalah memahami tujuan dan kebutuhan dari sudut pandang business, kemudian mengartikan pengetahuan ini ke dalam pendefinisan masalah dalam data mining. Selanjutnya akan ditentukan rencana dan strategy untuk mencapai tujuan tersebut.

2. Fase Pemahaman Data (Data Understanding Phase)

Fase ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi sebuah masalah kualitas data atau untuk menemukan adanya

bagian yang menarik dari data yang dapat digunakan untuk hipotesa didalam informasi yang tersembunyi.

### 3. Fase Pengolahan Data (Data Preparation Phase)

Fase ini meliputi semua langkah dalam membangun dataset akhir (data yang akan diproses pada tahap pemodelan/modeling) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, record, dan atribut-atribut data, termasuk proses pembersihan dan transformation data untuk kemudian dijadikan masukan dalam tahap pemodelan (modeling).

### 4. Fase Pemodelan (Modeling Phase)

Fase ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang baik. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga pada tahap ini masih mungkin kembali ke tahap sebelumnya.

### 5. Fase Evaluasi (Evaluation Phase)

Fase ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisis data. Pada tahap ini akan dilakukan evaluasi terhadap kinerja dan kualitas model sebelum yang digunakan menentukan apakah model tersebut dapat mencapai tujuan yang ditetapkan pada fase awal (Business Understanding).

### 6. Fase Penyebaran (Deployment Phase)

Pada tahap ini,knowledge atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh user. Tahap deployment dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang.

## 2.5 Prediksi

Prediksi merupakan proses memperkirakan dengan cara sistematis tentang suatu yang paling mungkin terjadi di masa yang akan datang berdasarkan data atau informasi di masa lalu dan sekarang, supaya kesalahan dalam melakukan perkiraan dapat diperkecil. Prediksi tidak harus mendapatkan jawaban secara pasti dalam suatu kejadian yang di masa yang akan datang akan terjadi, melainkan mengusahakan untuk mencari jawaban sedekat mungkin yang akan terjadi. Pengertian prediksi sama dengan perkiraan atau ramalan namun menurut kamus besar bahasa Indonesia, prediksi merupakan hasil dari suatu kegiatan memprediksi atau memperkirakan atau meramal nilai pada masa depan dengan menggunakan data masa lalu. Prediksi adalah tata cara untuk membuat informasi berdasarkan fakta tentang sebuah situasi masa depan dengan berdasarkan informasi yang telah ada dimasa lalu. Ramalan memiliki tiga bentuk yang utama adalah; prediksi, proyeksi, dan perkiraan.

### 1. Prediksi

Prediksi adalah ramalan yang memiliki dasar dari asumsi teoritik yang tegas. Asumsi ini bisa berbentuk teoritis (contohnya hukum berkurangnya nilai mata uang), proporsi teoritis (contohnya bahwa perpecahan yang ada di masyarakat diakibatkan oleh kesenjangan sosial)

### 2. Proyeksi

Proyeksi adalah ramalan yang berdasarkan pada kecenderungan masa lalu dan masa kini ke masa yang akan datang. Proyeksi memiliki pertanyaan yang tegas atas argument yang didapat dari metode-metode tertentu.

### 3. Perkiraan

Perkiraan adalah ramalan yang berdasarkan penilaian informatif dari penilaian pakar tentang apa yang akan ada di masa depan.

Berikut adalah pengelompokan dari peramalan yang akhirnya adalah pemerolehan informasi tentang perubahan yang akan datang(Kafil, 2019).

## 2.6 Saham

Saham adalah surat berharga sebagai bukti atau pernyataan atau kepemilikan seseorang dalam suatu perusahaan, khususnya perusahaan yang menjual sahamnya di bursa saham. Harga saham biasanya mengalami fluktuasi yang susah untuk diprediksi sehingga mengakibatkan para investor memiliki ketidakpastian dalam mendapatkan return dari saham yang dibelinya. Faktor yang memiliki pengaruh dalam pergerakan harga saham contohnya faktor yang dikarenakan dari internal (hasil untung dari perusahaan yang dibagikan kepada deviden) dan eksternal (suku bunga yang ada didalam negara dan kebijakan moneter dan fisikal serta keadaan perekonomian pada suatu negara) (Lusiana et al., 2018).

Saham adalah salah satu sarana bagi investor untuk berinvestasi, investasi saham mempunyai daya tarik untuk investor dikarenakan investor mempunyai harapan untuk mempunyai keuntungan yang berupa capital gain ataupun deviden yang tinggi. Harga saham mempunyai nilai nominal yang terdapat didalam surat bukti kepemilikan bagian dari suatu perusahaan atau perseroan terbatas. Harga saham sangat penting untuk melakukan pengukuran kinerja perusahaan dan sebagai penentu resiko dan hasil yang akan didapat dimasa yang akan datang. Saham juga bisa mencerminkan nilai atau prestasi dari suatu perusahaan, saham perusahaan akan diminati investor jika memiliki prestasi yang baik jika dilihat dari laporan keuangan yang dipublikasikan dan dilaporkan oleh perusahaan. Perusahaan berkewajiban untuk melaporkan laporan keuangan pada periode sebelumnya, laporan keuangan ini berguna untuk investor untuk membantu pengambilan keputusan yang dilakukan oleh investor. Harga saham perusahaan terjadi fluktuasi setiap waktunya ketidakstabilan ini menyulitkan investor dalam berinvestasi terlebih investor juga harus mempertimbangkan berbagai macam informasi yang terkait dengan perusahaan (Suyanto & Nursanti, 2017).

## 2.7 Analisis Harga Saham

Dalam beberapa literatur yang penulis lihat ada 2 pendekatan umum dalam melakukan analisis harga di bursa yaitu analisis fundamental dan analisis teknikal.

Didalam analisis fundamental para analis fundamental saham meyakini nilai dari intrinsik suatu harga saham, analisis fundamental ini didasarkan dengan pendekatan literasi pada laporan keuangan perusahaan yang dilakukan perusahaan setiap kuartalnya, jika suatu saham memiliki harga yang rendah dari nilai intrinsiknya maka harga saham tersebut sedang di kondisi undervalued dimana kondisi ini investor diharapkan bisa melakukan pembelian saham tersebut, selain itu analisis fundamental juga melibatkan beberapa aspek dari perusahaan yaitu return on investment, return on assets dan liabilitas ratio (debt).

Analisis teknikal adalah analisis yang dilakukan terhadap harga saham melalui data historical itu sendiri, pergerakan harga saham bisa dilihat pada rentang waktu sebelum-sebelumnya bahkan Analisa harga saham menggunakan analisis teknikal ini bisa menggunakan data histori dari beberapa tahun yang lalu. Pergerakan harga saham dilihat berdasarkan grafik yang divisualisasikan pada beberapa penyedia grafik harga saham yang ada di Indonesia contoh saja yang penulis gunakan adalah website yahoo finance. Chart atau grafik yang disediakan oleh penyedia grafik saham digunakan untuk melakukan analisis teknikal, pada chart atau grafik ini analis bisa melihat data historical dari harga saham terkait pada rentang waktu tertentu. Pada analisis harga saham menggunakan analisis teknikal juga terdapat beberapa indicator yang bisa dipakai dalam membantu analisis teknikal diantaranya adalah exponential moving average (EMA), moving average (MA), stochastic oscillator, dan moving average convergence and divergence (MACD).

## 2.8 Time Series (Data Deret Waktu)

Data *time series* adalah data dari serangkaian pengamatan yang berurutan berdasarkan waktu dengan jarak waktu yang sama. Jenis data *time series* seringkali ditemui dalam keseharian dikarenakan data tersebut di dapatkan melalui interval waktu harian, mingguan bahkan bulanan(Mahfud Al et al., 2020).

Data *time series* terbagi menjadi dua domain yang pertama adalah domain waktu dan domain frekuensi. Domain waktu mendalam tentang kestationeran data, signifikansi auto korelasi, peramalan dan penaksiran regresi model deret waktu. Sedangkan domain frekuensi mendalam tentang frekuensi data atau pola yang tersembunyi pada data yang tercatat berdasarkan waktu(Mahfud Al et al., 2020).

## 2.9 Stasioneritas

Stasioneritas adalah tidak adanya pertumbuhan dan penurunan data. data bisa dikatakan stationer jika pola data tersebut berada di keseimbangan nilai rata rata konstan. Data time series dikatakan stationer jika tidak ada tren didalam data deret waktu tersebut.

## 2.10 Uji stasioneritas

### 2.10.1 Metode visualisasi grafik

Untuk mencari apakah sebuah data deret waktu stasioner atau tidak secara visual dapat dilihat dari plot atau grafik data observasi terhadap kurun waktu tertentu. Jika data stationer maka grafik akan cenderung konstan atau dengan amplitude yang relatif tetap (tidak ada tren naik atau turun)(Sherly Aktivani, 2021).

### 2.10.2 Uji Correlogram

*Correlogram* bisa digunakan untuk mencari signifikansi autokorelasi dan kestasioneran data. Fungsi autokorelasi adalah hubungan antara autokorelasi data deret waktu dengan lagnya Tetapi harus diperhatikan karena uji *correlogram* ini dilakukan secara visual sehingga harus dilakukan pengujian statistik tertentu untuk mendapatkan keputusan yang lebih akurat dengan memakai uji akar unit atau *unit root test*(Sherly Aktivani, 2021).

### **2.10.3 Uji akar unit**

Uji akar unit dikembangkan oleh *Dickey* dan *Fuller* atau yang sering dikenal sebagai uji akar unit *Dickey-Fuller*. Pengujian stasionaritas data yang sering digunakan adalah pengujian akar unit (unit root test) dengan jenis pengujian *Augmented Dickey Fuller Test* (ADF test).

### **2.11 Differencing**

*Differencing* atau perbedaan dilakukan untuk menangani data yang tidak stationer atau menstasionerkan data nonstationer, jika data deret waktu nonstationer maka data akan dibuat mendekati stationer dengan melakukan *differencing* (Yuliyanti et al., 2022).

### **2.12 Support Vector Machine**

*Support Vector Machine* atau SVM adalah teknik machine learning yang termasuk golongan linear yang di generalisasi, metode ini memakai regresi atau klasifikasi yang digunakan untuk menghasilkan keputusan dengan kombinasi linear dari variabel yang ada pada dataset. Dengan memanfaatkan data histori SVM membuat bentuk fungsi matematika untuk membuat peta terhadap variabel, hasil akhir SVM berbentuk klasifikasi atau regresi untuk melakukan forecasting atau prediksi (Sheykhmousa et al., 2020).

Konsep SVM adalah kombinasi harmonis dari komputasi yang sudah ada dari puluhan tahun lalu. Prinsip dari SVM adalah bekerja secara linier dan dikembangkan untuk bisa diterapkan pada kasus yang non linier menggunakan metode kernel trick yang bisa mencari hyperplane dengan mengubah dan mentransformasikan dataset ke bidang vector yang berdimensi tinggi (feature space), selanjutnya dilakukan proses klasifikasi pada feature space. Penentuan kernel fungsi yang dipakai sangat berpengaruh pada hasil prediksi (Imelda A.Muis & Muhammad Affandes, 2015).

SVM adalah salah satu metode learning machine yang bekerja dengan prinsip *Structural Risk Minimization* (SRM) yang memiliki tujuan untuk mencari hyperplane yang paling baik yang bisa memisahkan dua buah class pada space

input. Metode ini memakai hipotesis yang berupa fungsi linear dalam sebuah ruang fitur yang memiliki dimensi tinggi dengan menggunakan bias learning yang berasal dari teori statistik. Data pada suatu dataset diberikan variabel  $x_i$ , sedangkan untuk dataset diberikan  $y_i$ . Metode SVM memiliki 2 dataset yaitu kelas pertama dipisahkan dengan hyperplane yang memiliki nilai 1 dan kelas selain hyperplane memiliki nilai -1 (Monika Parapat & Tanzil Furqon, 2018).

$$X_i \cdot W + b \geq 1 \text{ untuk } Y_i = 1$$

$$X_i \cdot W + b \leq -1 \text{ untuk } Y_i = -1$$

Keterangan :

$X_i$ = data ke -i

$W$ = Bobot support vector (tegak lurus dengan hyperplane)

$b$ = nilai bias

$Y_i$ = kelas data ke-i

Bobot vector atau  $W$  merupakan garis vector yang tegak lurus antara titik pusat koordinat dengan garis hyperplanr, bias ( $B$ ) adalah koordinat garis relative dari titik koordinat.

$$b = -12(w.x+ + w.x-)$$

$$w = \sum \alpha_i y_i x_i n_i = 1$$

Keterangan :

$b$ =nilai bias

$w.x+$ = nilai bobot untuk kelas data positif

$w.x^-$ = nilai bobot untuk kelas data negative

$w$ = bobot vector

$\alpha_i$ = nilai bobot data ke-i

$y_i$ = kelas data ke-i

$x_i$ = data ke-i

H1 adalah hyper plane pendukung dari kelas +1 yang memiliki fungsi  $wx_1+b=+1$ .

Margin = | dH1-dH2 | = 2/w

Keterangan :

$dH1$  = jarak hyperplane pendukung kelas +1

$dH2$  = jarak hyprplane pendukung kelas -1

## 2.13 Evaluasi

Kesalahan prediksi adalah ukuran melihat kinerja suatu model prediksi yang digunakan untuk membandingkan nilai hasil prediksi dari model yang digunakan dengan data actual yang ada. Dalam penelitian ini pengukuran kesalahan prediksi menggunakan parameter *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE).

### 2.13.1 Root Mean Square Error

Nilai yang didapatkan RMSE adalah nilai rata – rata kuadrat dari total kesalahan pada algoritma atau model prediksi. Root Mean Square Error (RMSE) merupakan teknik yang bisa dibilang mudah diimplementasikan dan seringkali digunakan dalam penelitian yang berkaitan dengan prediksi (Putri & Sadikin, 1165).

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Gambar 2. 3 Root Mean Square Error

Keterangan :

$\hat{y}_i$  = Nilai Hasil Peramalan

$y_i$  = Nilai Aktual / Nilai Sebenarnya

$n$  = Jumlah Data

### 2.13.2 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah rata – rata diferensiasi antara nilai actual dan nilai peramalan, hasil dari MAPE adalah persentase nilai actual. MAPE dipakai untuk mencari persentase kesalahan antara nilai prediksi dan nilai actual(Putri & Sadikin, 1165).

$$100 \times \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Gambar 2. 4 Mean Absolute Percentage Error

Keterangan :

$\hat{y}_i$  = Nilai Hasil Peramalan

$y_i$  = Nilai Aktual / Nilai Sebenarnya

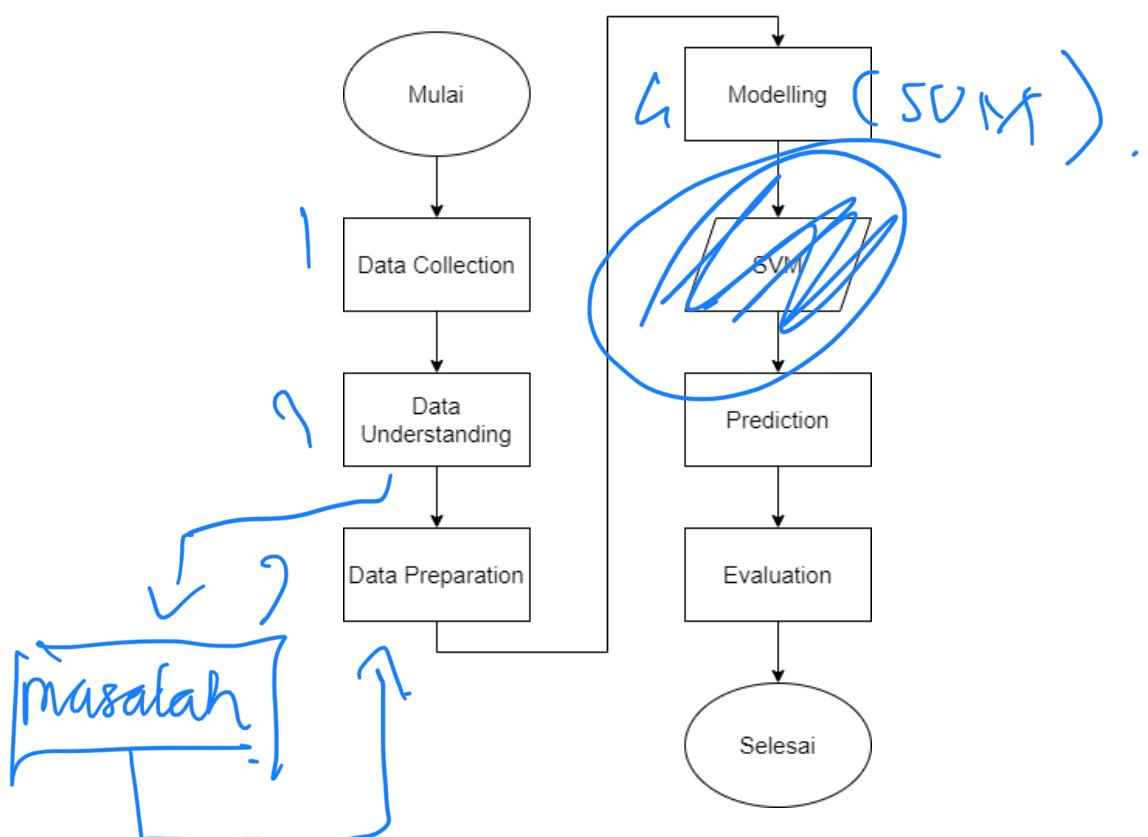
$n$  = Jumlah Data

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Alur Penelitian

Penelitian kali ini penulis akan melakukan prediksi terhadap harga saham IHSG namun sebelum melakukan prediksi ditemukan masalah pada data IHSG yang didapatkan penulis yaitu ketidakstationeran data IHSG yang kurang baik maka selanjutnya penulis akan mencari metode yang bisa mengubah data IHSG menjadi stationer lalu selanjutnya penulis juga menggunakan metode *Support Vector Machine* atau SVM adalah teknik machine learning yang termasuk golongan linear yang di generalisasi, metode ini memakai regresi atau klasifikasi yang digunakan untuk menghasilkan keputusan dengan kombinasi linear dari variabel yang ada pada dataset.



Gambar 3. 1 Alur Penelitian

Pendekatan penelitian yang akan dilakukan pada penelitian ini adalah pendekatan kuantitatif. Penelitian kuantitatif adalah proses mengambil pengetahuan yang memakai data berupa angka sebagai alat bantu menganalisis data tentang apa yang ingin diketahui. Menurut (Imron, 2019) Metode kuantitatif disebut kuantitatif dikarenakan data penelitian yang berupa angka dan analisis menggunakan statistic.

### 3.2 Teknik Analisis Data Harga Saham

Penelitian kali ini memakai analisis teknikal. Analisis teknikal dilakukan pertama kali pada tahun 1884 oleh *Charles Henry Dow*, dimana Dow mendapatkan perhitungan harga saham yang dipunyainya berdasarkan data time series. Teori Dow menyebutkan bahwa teori ini bertujuan untuk mengetahui harga pasar dalam jangka panjang berdasarkan data harga historis dimasa lalu. Objek dari analisis teknikal yaitu memprediksi dari time series data dengan metode peramalan.

### 3.3 Data Collection

Data collection adalah proses pengumpulan, pengukuran, dan analisis berbagai tipe informasi menggunakan teknik berstandar. Tujuan utama data collection adalah untuk mengumpulkan informasi dan data terpercaya sebanyak-banyaknya, yang kemudian dianalisis untuk membuat sebuah keputusan bisnis yang krusial. Data yang dikumpulkan adalah data harga saham IHSG dari situs *Yahoo Finance*. Data yang diambil adalah berupa harga buka, harga tutup, harga terendah, dan harga tertinggi. Adapun data lain yang tercantum dalam pengambilan data yaitu tanggal beserta jam harga tersebut terbentuk. Tahap akuisisi data adalah pengambilan data dari sumber data dari *repository online Yahoo Finance* tautan dari data yang diambil adalah sebagai berikut  
<https://finance.yahoo.com/quote/%5EKSE/history?period1=639360000&period2=1666828800&interval=1d&filter=history&frequency=1d&includeAdjustedClose>

=true&guccounter=1 dimana data ini adalah data yang akan dipakai dan digunakan sebagai data set pada penelitian ini.

Data yang didapatkan mempunyai format .csv sehingga untuk membuka dan melihat isi data yang ada peneliti melakukan perubahan format .csv menjadi .xlsx. berikut ini adalah data set yang peneliti dapatkan.

	A	B	C	D	E	F
1	Date	Open	High	Low	Close	
2	27/10/2021	6.664.373.047	6.670.809.082	6.590.504.883	6.602.208.984	
3	28/10/2021	6.578.583.008	6.600.247.070	6.509.877.930	6.524.076.172	
4	29/10/2021	6.561.712.891	6.596.769.043	6.551.539.063	6.591.346.191	
5	01/11/2021	6.618.122.070	6.627.851.074	6.552.889.160	6.552.889.160	
6	02/11/2021	6.557.944.824	6.573.693.848	6.480.009.766	6.493.274.902	
7	03/11/2021	6.497.865.234	6.561.810.059	6.485.250.977	6.552.129.883	
8	04/11/2021	6.588.188.965	6.617.861.816	6.583.169.922	6.586.442.871	
9	05/11/2021	6.600.722.168	6.608.424.805	6.550.117.188	6.581.785.156	
10	08/11/2021	6.599.404.785	6.637.011.230	6.592.055.176	6.632.296.875	
11	09/11/2021	6.640.221.191	6.669.924.805	6.633.208.008	6.669.924.805	
12	10/11/2021	6.665.548.828	6.683.145.020	6.651.797.852	6.683.145.020	
13	11/11/2021	6.694.583.008	6.704.463.867	6.671.227.051	6.691.341.797	
14	12/11/2021	6.710.853.027	6.714.158.203	6.646.315.918	6.651.054.199	
15	15/11/2021	6.661.849.121	6.675.595.215	6.610.823.242	6.616.028.809	
16	16/11/2021	6.598.408.203	6.654.515.137	6.592.228.027	6.651.207.031	
17	17/11/2021	6.667.861.816	6.679.304.199	6.650.494.141	6.675.804.199	
18	18/11/2021	6.653.700.195	6.669.786.133	6.621.692.871	6.636.469.238	
19	19/11/2021	6.651.778.809	6.720.987.793	6.651.778.809	6.720.263.184	
20	22/11/2021	6.731.515.137	6.754.463.867	6.690.817.871	6.723.386.230	
21	23/11/2021	6.716.157.227	6.732.897.949	6.662.538.086	6.677.875.977	
22	24/11/2021	6.697.423.828	6.698.414.063	6.668.054.199	6.683.276.855	
23	25/11/2021	6.694.957.031	6.751.001.953	6.694.796.875	6.699.346.191	
24	26/11/2021	6.688.568.848	6.691.711.914	6.544.895.996	6.561.553.223	
25	29/11/2021	6.552.803.223	6.617.287.109	6.487.748.047	6.608.290.039	
26	30/11/2021	6.605.796.875	6.647.477.051	6.533.932.129	6.533.932.129	
27	01/12/2021	6.544.517.090	6.593.074.219	6.494.497.070	6.507.676.758	
28	02/12/2021	6.517.338.867	6.586.895.996	6.484.578.125	6.583.819.824	
29	03/12/2021	6.588.014.160	6.600.165.039	6.536.907.227	6.538.505.859	

Gambar 3. 2 Sumber Data

### 3.4 Data Understanding

Proses *data understanding* yang digunakan adalah *Exploratory Data Analysis* (EDA). *Exploratory Data Analysis* (EDA) adalah bagian dari proses data *science*. EDA menjadi sangat penting sebelum melakukan *feature engineering* dan *modeling* karena dalam tahap ini kita harus memahami datanya terlebih dahulu. Pada penelitian ini dilakukan pendefinisian setiap atribut pada dataset yang digunakan dalam penelitian yang bertujuan untuk memahami data yang akan digunakan. Berikut adalah definisi dari setiap atribut dari data saham IHSG.

Atribut	Deskripsi	Kriteria
Date	Berisikan data waktu harga	09/04/1990 - 30/01/2023
<i>Open</i>	Berisikan data harga buka	Rp.223 – Rp.7318
<i>High</i>	Berisikan data harga yang tertinggi	Rp.223 - Rp.7377
<i>Low</i>	Berisikan data harga yang terrendah	Rp.223 - Rp.7278
<i>Close</i>	Berisikan data harga tutup	Rp.223 - Rp.7318

Tabel 3. 1 Data Understanding

Kemudian dilakukan *Augmented Dickey Fuller Test*. Ada beberapa uji akar unit yang sering digunakan yaitu *Augmented Dickey Fuller* (ADF) dan uji *Philip Peron* (PP) didalam peneltian ini penulis hanya memakai uji ADF saja. uji akar unit ini dikembangkan oleh *Dickey* dan *Fuller*. Langkah-langkah yang digunakan untuk melihat apakah data stasioner atau tidak yaitu dengan cara membandingkan antara nilai statistik ADF dengan nilai kritis ADF. Apabila nilai ADF lebih besar dari nilai kritisnya maka data tersebut stasioner dan jika nilai ADF lebih kecil dari nilai kritisnya maka data tersebut tidak stasioner selain itu Jika P-Value dalam uji ADF kurang dari tingkat nominal (0,05) maka data tersebut stationer dan jika uji ADF lebih dari tingkat nominal (0,05) maka data tersebut tidak stationer

(Sherly Aktivani, 2021). Data yang tidak stasioner tersebut dapat dijadikan data stasioner dengan cara uji stasioneritas pada tingkat deferensi data atau uji derajat integrasi. Uji ini dilakukan untuk mengetahui pada derajat integrasi berapakah data tersebut stasioner.

```

C> Augmented Dickey-Fuller Test:
ADF test statistic      0.600578
p-value                  0.987630
# lags used              11.000000
# observations            382.000000
critical value (1%)     -3.447585
critical value (5%)      -2.869136
critical value (10%)     -2.570816
Weak evidence against the null hypothesis
Fail to reject the null hypothesis
Data has a unit root and is non-stationary

```

Gambar 3. 3 ADF Test

Pada gambar diatas diketahui jika data time series saham IHSG tidak stationer yang berarti harus melalui proses differencing data terlebih dahulu jika ingin mendapatkan performa yang optimal dari metode data mining yang digunakan.

### 3.5 Data Preparation

Pada tahap ketiga dilakukan penyiapan data awal yang digunakan pada fase berikutnya. Pada dataset yang digunakan dalam penelitian ini melalui tahap differencing data terlebih dahulu, berikut adalah penjelasan tahap yang dilakukan.

Tahap differencing, Penelitian kali ini menggunakan data deret waktu IHSG atau data time series. Setelah dilakukan pengujian ternyata data time series IHSG tidak stationer. Data time series yang tidak stationer adalah salah satu penyebab hasil prediksi pada model regresi meragukan dan kurang akurat(Mahfud Al et al., 2020). Data time series yang dikatakan tidak stationer jika varian rata rata dan kovarian lainnya terdapat lag pada rentang waktu yang sama. Ada beberapa metode dalam uji stasionaritas. Metode yang sering digunakan untuk menguji masalah stationer adalah dengan menggunakan uji akar unit. Lalu dilakukan differencing pada data

IHSG dengan cara menghitung perubahan atau selisih nilai observasi, untuk menstationerkan data dengan proses mencari selisih antara data dengan data sebelumnya. Nilai selisih yang diperoleh dan dicek lagi apakah stasioner atau tidak. Jika belum stasioner maka dilakukan differencing lagi sampai didapatkan nilai test ADF stationer.

### **3.6 Modeling**

Pada tahap modeling ini dilakukan menentukan dan menerapkan teknik pemodelan yang sesuai, mengkalibrasi aturan model untuk mengoptimalkan hasil, dan dapat kembali ke fase pengolahan data ke dalam bentuk kebutuhan tertentu. Modeling data yang digunakan pada penelitian ini adalah *Support Vector Machine*.

Data training nantinya digunakan untuk melatih algoritma dalam mencari model yang sesuai, sedangkan data testing dipakai untuk menguji dan mengetahui performa model yang didapatkan pada tahapan testing.

### **3.7 Prediction**

Hasil dari metode metode modelling menggunakan *Support Vector Machine* ini adalah berupa prediksi harga saham IHSG yang nantinya digunakan dalam proses pengetesan.

### **3.8 Evaluation**

Setelah hasil pengukuran diketahui, maka langkah selanjutnya adalah evaluasi menggunakan *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE).

## DAFTAR PUSTAKA

- Alfannisa Annurullah Fajrin. (2018). *PENERAPAN DATA MINING UNTUK ANALISIS POLA PEMBELIAN KONSUMEN DENGAN ALGORITMA FP-GROWTH PADA DATA TRANSAKSI PENJUALAN SPARE PART MOTOR.*
- Imelda A.Muis & Muhammad Affandes, M. T. (2015). Penerapan Metode Support Vector Machine ( SVM ) Menggunakan Kernel Radial Basis Function ( RBF ) Pada Klasifikasi Tweet. *Sains, Teknologi Dan Industri.UIN Sultan Syarif Kasim Riau*, 12(2), 189–197.
- Imron. (2019). *Analisa Pengaruh Kualitas Produk Terhadap Kepuasan Konsumen Menggunakan Metode Kuantitatif Pada CV.Meubele Berkah Tangerang.*
- Kafil, M. (2019). PENERAPAN METODE K-NEAREST NEIGHBORS UNTUK PREDIKSI PENJUALAN BERBASIS WEB PADA BOUTIQ DEALOVE BONDOWOSO. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 3, Issue 2).
- Kurniawan, A. S. (2018). Implementasi Metode Artificial Neural Network Dalam Memprediksi Hasil Ujian Kompetensi Kebidanan (Studi Kasus Di Akademi Kebidanan Dehasen Bengkulu). *Pseudocode*, 5(1), 37–44. <https://doi.org/10.33369/pseudocode.5.1.37-44>
- Lusiana, Martha, S., & Rizki, S. W. (2018). Simulasi Pergerakan Harga Saham Menggunakan Pendekatan Metode Monte Carlo. *Buletin Ilmiah Math.Stat. Dan Terapannya (Bimaster)*, 07(2), 119–126.
- Mahfud Al, A., Kurniasari, D., Mustofa Usman, dan, Matematika, J., Mipa, F., Lampung Jl Sumantri Brojonegoro No, U., Lampung, B., kunci, K., Spektral, A., Waktu, D., & Pesawat, P. (2020). Peramalan Data Time Series Seasonal Menggunakan Metode Analisis Spektral Berdasarkan data yang tersedia diperoleh model terbaik untuk peramalan penumpang

pesawat di Bandar Udara Raden Intan II adalah Seasonal ARIMA (0. In *Jurnal Siger Matematika* (Vol. 01, Issue 01).

Maulita, D., & Arifin, M. (2018). Pengaruh Return On Investment (ROI) Dan Earning Per Share (EPS) Terhadpa Return Saham Syariah (Studi Kasus Pada Perusahaan Manufaktur Sub Sektor Makanan dan Minuman Yang Terdaftar Pada Indeks Saham Syariah Indonesia Periode 2012-2016). *Jurnal Manajemen*, 8, 10–19.

Monika Parapat, I., & Tanzil Furqon, M. (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3163–3169. <http://j-ptiik.ub.ac.id>

Pramansah, V. V., Iskandar Mulyana, D., & Silfia, T. (2022). *Analisis Perbandingan Algoritma SVM Dan KNN Untuk Klasifikasi Anime Bergenre Drama.* 03(01), 49–55. <https://ejurnalunsam.id/index.php/jicom/>

Purnama, I. (2020). *IMPLEMENTASI DATA MINING MENGGUNAKAN CRISP-DM PADA SISTEM INFORMASI EKSEKUTIF DINAS KELAUTAN DAN PERIKANAN PROVINSI JAWA TENGAH.* <https://doi.org/10.3997/2214-4609-pdb.319.35>

Purwanti, A. P. J. (2017). Integrasi Bursa Saham Indonesia Dengan Bursa Saham Global. *Jurnal Ilmiah Mahasiswa FEB Universitas Brawijaya*, 5(2). <https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/3749>

Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 72–77. <https://doi.org/10.31539/intecoms.v1i1.141>

Putri, E. S., & Sadikin, M. (1165). Prediksi Penjualan Produk Untuk Mengestimasi Kebutuhan Bahan Baku Menggunakan Perbandingan

- Algoritma LSTM dan ARIMA. In *Universitas Mercu Buana Alamat Jl. Raya.*
- Rosyida, H., Firmansyah, A., & Wicaksono, S. B. (2020). Volatilitas Harga Saham: Leverage, Ukuran Perusahaan, Pertumbuhan Aset. *JAS (Jurnal Akuntansi Syariah)*, 4(2), 196–208. <https://doi.org/10.46367/jas.v4i2.256>
- Sherly Aktivani. (2021). *UJI STASIONERITAS DATA INFLASI KOTA PADANG PERIODE 2014-2019.*
- Sheykhou, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (Vol. 13, pp. 6308–6325). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/JSTARS.2020.3026724>
- Shleifer, Andrei. (2000). *Inefficient markets : an introduction to behavioral finance*. Oxford University Press.
- Sugandhi, Y. P., Warsito, B., & Hakim, A. R. (2019). Prediksi Harga Saham Harian Menggunakan Cascade Forward Neural Network (CFNN) Dengan Particle Swarm Optimization (PSO). *STATISTIKA Journal of Theoretical Statistics and Its Applications*, 19(2), 71–82. <https://doi.org/10.29313/jstat.v19i2.4878>
- Suyanto, S., & Nursanti, A. (2017). Faktor-Faktor Yang Memengaruhi Harga Saham Perusahaan Food and Beverage Yang Terdaftar Di Bursa Efek Indonesia. *Jurnal Neraca: Jurnal Pendidikan Dan Ilmu Ekonomi Akuntansi*, 1(1), 18–31. <https://doi.org/10.31851/neraca.v1i1.1170>
- Wibowo, E. (2011). Analisis Penentuan Saham yang akan Dibeli Suatu Tinjauan Umum. *Jurnal Ekonomi Dan Kewirausahaan*, 11(1), 151–158.

Yuliyanti, R., Matematika, P., Uny, F., & Arlian, E. (2022). Peramalan jumlah penduduk menggunakan model arima Forecasting the number of population using the arima. In *Jurnal Kajian dan Terapan Matematika* (Vol. 8, Issue 2). <http://journal.student.uny.ac.id/ojs/index.php/jktm>: