

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

Penelitian ini mengambil objek berupa saham dari perusahaan Intel Corporation (INTC) yang merupakan salah satu pemain utama dalam industri semikonduktor global. Intel dikenal sebagai produsen *micro processor* dan *chip* yang digunakan di berbagai perangkat elektronik mulai dari komputer pribadi, *server*, hingga sistem berbasis *artificial intelligence*. Sebagai perusahaan besar yang sahamnya diperdagangkan di bursa NASDAQ, pergerakan harga saham Intel sangat dipengaruhi oleh kondisi industri teknologi, persaingan pasar, hingga sentimen global.



Gambar 3. 1 Visualisasi Harga Saham Intel

Dapat dilihat pada Gambar 3.1, pergerakan harga saham Intel relatif tidak dapat diprediksi dengan mudah karena banyaknya faktor yang memengaruhi. Misalnya, peluncuran produk baru, kompetisi ketat dengan perusahaan seperti AMD dan Nvidia, penundaan inovasi teknologi fabrikasi, hingga krisis pasokan chip akibat gangguan rantai distribusi. Di sisi lain, langkah ekspansi Intel seperti membangun pabrik di Eropa dan melakukan restrukturisasi juga berdampak terhadap persepsi investor dan fluktuasi harga sahamnya.

Karena sifatnya yang fluktuatif dan non linier, data harga saham Intel menjadi sangat menarik untuk dianalisis secara prediktif. Dalam penelitian ini, data yang

digunakan adalah harga penutupan harian (*Close*) yang diambil dari situs *Yahoo Finance*. Data dikumpulkan sejak tahun 1980 hingga 2025, dengan total lebih dari 11.000 baris data. Data ini sudah cukup panjang untuk menangkap pola historis yang terjadi dalam jangka waktu panjang dan digunakan untuk membangun model prediksi berbasis *machine learning*, khususnya model *Hybrid* antara *Long Short-Term Memory* (LSTM) dan *Support Vector Machine* (SVM).

### 3.2 Metode Penelitian

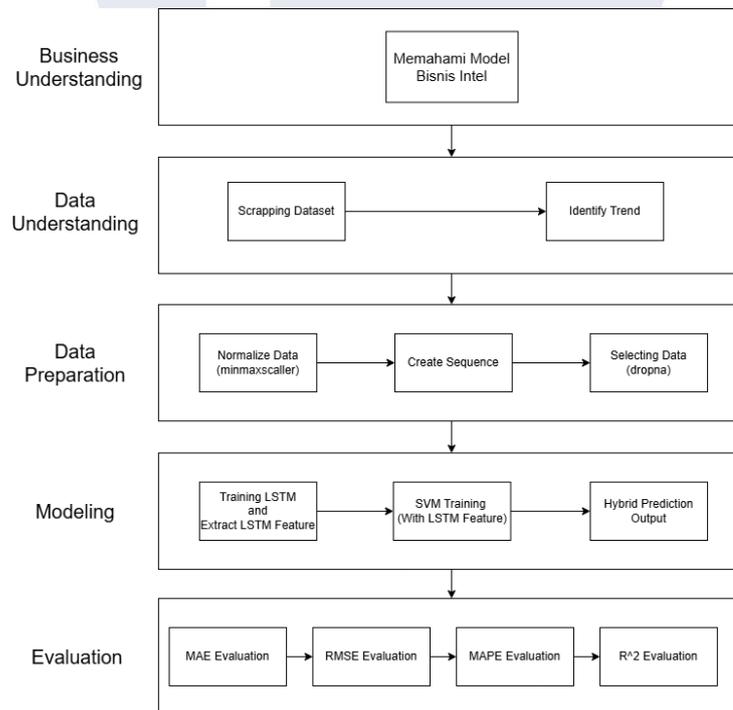
Penelitian ini menggunakan pendekatan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM). Metode ini dipilih karena menyediakan kerangka kerja yang sistematis dan fleksibel dalam menyusun alur proses data mining yang kompleks, termasuk dalam permasalahan prediksi harga saham berbasis *time series*. CRISP-DM dirancang untuk dapat digunakan dalam berbagai industri dan mendukung proses iteratif, sehingga cocok dalam menangani eksperimen model prediktif berbasis kecerdasan buatan.

Jika dibandingkan dengan metode lain seperti *SEMMA* (*Sample, Explore, Modify, Model, Assess*), CRISP-DM lebih menekankan pada pemahaman konteks bisnis secara menyeluruh sebelum dilakukan eksplorasi dan pemodelan data. *SEMMA* lebih terfokus pada eksplorasi dan pemodelan data secara teknis, dan umumnya digunakan dalam lingkungan perangkat lunak SAS. Sedangkan CRISP-DM lebih fleksibel diterapkan secara umum dengan platform open source dan memungkinkan pendekatan yang lebih menyeluruh dari awal hingga deployment model.

CRISP-DM menjadi pilihan utama dalam penelitian ini karena tahapannya yang lebih cocok untuk menyusun dan mengevaluasi pendekatan *Hybrid* LSTM-SVM, mulai dari memahami konteks prediksi saham, mengolah data historis dalam bentuk *sequence*, merancang model *deep learning*, hingga menganalisis hasil prediksi secara menyeluruh. Pendekatan ini memungkinkan alur kerja yang sistematis dan terukur dari awal hingga akhir proses pengembangan model. Dengan demikian, CRISP-DM memberikan pondasi metodologis yang kuat untuk menghasilkan prediksi harga saham yang lebih akurat dan relevan secara kontekstual.

### 3.2.1 Alur Penelitian

Penelitian ini menggunakan kerangka kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai acuan utama dalam pengembangan dan implementasi model prediksi. CRISP-DM dipilih karena mampu memberikan alur kerja yang sistematis, iteratif, dan fleksibel dalam proses data mining, terutama dalam konteks permasalahan prediksi data time series seperti harga saham. Kerangka ini mencakup enam tahapan utama: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan deployment. Setiap tahapan saling berkesinambungan dan dapat diulang jika diperlukan, sehingga sangat cocok diterapkan dalam eksperimen berbasis kecerdasan buatan yang memerlukan evaluasi berkala dan pengujian model secara mendalam.



Gambar 3. 2 Alur Penelitian

Dalam konteks penelitian ini, tahapan-tahapan CRISP-DM yang diimplementasikan telah disesuaikan dengan kebutuhan spesifik dalam membangun dan menguji model *Hybrid LSTM-SVM* untuk prediksi harga

saham Intel. Setiap tahapan diadaptasi agar mendukung tujuan utama penelitian, mulai dari perumusan masalah bisnis yaitu kebutuhan prediksi saham yang akurat, hingga penerapan teknis berupa pemrosesan data historis saham dan pengembangan model *Hybrid* berbasis *deep learning* dan *machine learning*. Visualisasi alur tahapan penelitian ini ditampilkan pada Gambar 3.2, yang menggambarkan secara menyeluruh proses integrasi antara analisis data historis dan pembuatan model prediktif yang bertujuan untuk memberikan hasil prediksi yang presisi serta adaptif terhadap dinamika pasar saham yang fluktuatif. Berikut merupakan tahapan CRISP-DM yang diimplementasikan pada penelitian ini :

### **3.2.1.1 Business Understanding**

Tahapan Penelitian ini diawali dengan identifikasi permasalahan yang terjadi pada dunia pasar modal, khususnya pada aspek prediksi harga saham yang fluktuatif dan kompleks. Pergerakan harga saham Intel yang dinamis sering kali dipengaruhi oleh berbagai faktor eksternal dan internal, seperti kondisi pasar global, persaingan industri semikonduktor, hingga sentimen investor. Hal ini menjadikan pendekatan konvensional kurang mampu menangkap pola historis yang kompleks dan non-linier secara akurat. Oleh karena itu, dirancanglah pendekatan berbasis kecerdasan buatan dengan menggabungkan kekuatan dua algoritma yang memiliki karakteristik berbeda namun saling melengkapi, yaitu LSTM dan SVM. LSTM dipilih karena kemampuannya dalam mengenali pola urutan data jangka panjang yang terdapat dalam *time series*, sedangkan SVM dimanfaatkan sebagai pelengkap untuk meningkatkan akurasi regresi melalui klasifikasi non-linier.

### **3.2.1.2 Data Understanding**

Setelah permasalahan bisnis dipahami, tahap selanjutnya adalah memahami struktur dan karakteristik data yang akan digunakan. Data yang digunakan berasal dari situs *Yahoo Finance*, berupa data harga penutupan harian saham Intel dari tahun 1980 hingga 2025. Total data yang diperoleh berjumlah lebih dari 11.000 baris, yang mana cukup panjang untuk

digunakan sebagai dasar pelatihan model *deep learning*. Pada tahap ini, dilakukan eksplorasi awal seperti visualisasi tren harga saham secara historis, pemeriksaan adanya *missing value* atau duplikasi data. Proses konversi kolom *Date* menjadi format *datetime* dan menjadikannya sebagai *index data* juga dilakukan pada tahap ini, agar memudahkan proses manipulasi data sebagai *time series*. Hanya kolom *Close* yang digunakan dalam penelitian ini karena dianggap sebagai representasi utama dari nilai saham yang umum digunakan dalam analisis teknikal.

#### **3.2.1.3 Data Preparation**

Tahapan ini merupakan tahapan penting sebelum data dimasukkan ke dalam model. *Data preparation* diawali dengan melakukan transformasi nilai menggunakan *MinMaxScaler* untuk mengubah skala harga penutupan ke rentang 0 hingga 1, agar model LSTM dapat belajar dengan lebih optimal dan stabil. Setelah itu, dilakukan pembentukan data *sequence*, yaitu pemecahan data historis sepanjang 60 hari, yang akan menjadi input standar bagi model LSTM. Hal ini bertujuan agar LSTM dapat menangkap pola harga dalam 60 hari terakhir untuk memprediksi harga pada hari ke-61. *Dataset* kemudian dibagi menjadi data *training* (80%) dan data *validation* (20%) untuk memastikan model dapat dievaluasi dengan data yang belum pernah dilihat sebelumnya. Selanjutnya, data latih diubah ke dalam bentuk tiga dimensi dengan format [jumlah sampel, panjang *sequence*, jumlah fitur] yang merupakan bentuk input standar dalam arsitektur LSTM.

#### **3.2.1.4 Modeling**

Pada tahap pemodelan, proses dibagi menjadi dua bagian utama. Pertama, dibangun arsitektur model LSTM dengan jumlah unit memori sebanyak 100, diikuti oleh *layer dropout* sebesar 0.3 untuk mengurangi risiko *overfitting*. Model ini dilatih dengan data *sequence* yang telah disiapkan sebelumnya. Setelah model LSTM terlatih, dilakukan proses ekstraksi fitur dari *hidden state* terakhir LSTM. Fitur-fitur ini merupakan representasi kompleks dari pola historis yang telah dipelajari oleh LSTM, dan kemudian digunakan sebagai input untuk membangun model SVM.

Algoritma SVM dilatih untuk melakukan regresi terhadap *output* yang telah ditransformasikan oleh LSTM, menggunakan kernel RBF yang efektif untuk menangani data non-linier. Selanjutnya, dilakukan proses penggabungan hasil prediksi dari LSTM dan SVM menggunakan teknik *weighted sum*. Pengujian dilakukan dengan beberapa konfigurasi rasio, mulai dari 90:10, 80:20, hingga 50:50 untuk mengetahui sejauh mana kontribusi dari masing-masing model mempengaruhi hasil akhir prediksi..

#### 3.2.1.5 Evaluation

Model *Hybrid* yang telah dibangun akan dievaluasi menggunakan empat metrik utama: RMSE, MAE, MAPE, dan  $R^2$ . Selain itu, hasil prediksi juga dibandingkan dengan data aktual melalui grafik visual agar bisa dilihat kesesuaiannya secara langsung. Evaluasi dilakukan pada data uji yang tidak dilibatkan dalam pelatihan model untuk menilai performa generalisasi.

### 3.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan pendekatan data sekunder, yakni dengan mengunduh data historis harga saham Intel *Corporation* dari situs *Yahoo Finance* (<https://finance.yahoo.com/quote/INTC/>) dalam format CSV. Pemilihan situs ini didasarkan pada pertimbangan bahwa Yahoo Finance merupakan sumber data pasar saham yang kredibel, bersifat terbuka, mudah diakses, serta telah banyak dijadikan referensi dalam penelitian-penelitian terdahulu. *Dataset* yang diperoleh terdiri atas lebih dari 11.000 baris data, mencakup rentang waktu dari tahun 1980 hingga 2025. Rentang waktu sepanjang 45 tahun dipilih untuk memastikan model memiliki cakupan historis yang memadai dalam mempelajari pola pergerakan harga, sekaligus memenuhi ketentuan akademik yang mensyaratkan minimal 10.000 data observasi.

Jenis data yang digunakan dalam penelitian ini adalah harga penutupan harian (*closing price*), mengingat kolom ini cukup lumrah dijadikan sebagai indikator utama dalam analisis teknikal oleh para analis dan investor. Sementara itu, atribut lain seperti harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), dan volume perdagangan tidak disertakan guna menjaga fokus dan kesederhanaan

model prediksi yang dibangun. Setelah proses pengunduhan data selesai, dilakukan tahapan *preprocessing* yang mencakup konversi data ke dalam format deret waktu (*time series*), pengecekan dan penanganan nilai hilang (*missing values*), serta normalisasi data agar sesuai dengan karakteristik input model *machine learning* yang akan digunakan.

### 3.4 Teknik Analisis Data

Analisis data dalam penelitian ini dilakukan setelah seluruh model selesai dibangun dan dilatih. Model LSTM dilatih menggunakan data sekuensial harga saham untuk mempelajari pola historis pergerakan harga. Hasil pelatihan LSTM tidak langsung digunakan untuk prediksi, melainkan diekstrak fitur internal, tepatnya dari output *hidden state* sebelum *layer Dense*, untuk digunakan sebagai input ke dalam model SVM. Model SVM dilatih secara terpisah menggunakan fitur dari LSTM tersebut, dengan parameter *kernel* RBF dan nilai C sebesar 100, serta epsilon 0.01. Tujuannya adalah agar SVM dapat menangkap representasi fitur yang telah dipelajari oleh LSTM dan menghasilkan prediksi akhir dengan pendekatan klasifikasi atau regresi berbasis *margin*. Setelah kedua model selesai dibangun, hasil prediksi LSTM dan SVM di lakukan evaluasi akhir dengan empat metrik utama: MAE, RMSE, MAPE, dan R<sup>2</sup>. Visualisasi hasil prediksi terhadap data aktual dalam bentuk grafik garis juga dilakukan untuk menunjukkan seberapa dekat model mengikuti pola harga saham asli.

