

BAB III

METODOLOGI PENELITIAN

3.1. Gambaran Umum Objek Penelitian

Objek pada penelitian ini berfokus pada prediksi pergerakan harga saham BBKA, BBRI, dan BMRI di Indonesia. Objek pada penelitian ini adalah saham. Baru-baru ini di Indonesia, terjadi peningkatan volume transaksi saham dengan cukup signifikan dimana pada bulan Januari tahun 2020, jumlah transaksi volume 164.351 juta saham bertambah menjadi 373.334 juta saham pada bulan Januari tahun 2024. Angka ini naik sebesar 208.983 juta saham dan menjadi angka transaksi tertinggi selama periode 5 tahun lalu (*Badan Pusat Statistik, 2024*).

Berdasarkan data di atas, penulis ingin memprediksi harga saham BBKA, BBRI, dan BMRI di Indonesia pada tahun 2024 mendatang dengan melakukan *forecasting* atau peramalan. Sumber data yang digunakan pada penelitian ini bersumber dari Badan Pusat Statistik (BPS), yahoo finance, dan Bursa Efek Indonesia (BEI).



Gambar 2.1 Harga Saham BBKA, BBRI, dan BMRI tahun 2017 - 2021

Gambar 3.1 menjelaskan mengenai data berupa grafik dalam bentuk *line chart* yang bersumber dari website stockbit berupa data pergerakan harga saham BBKA, BBRI, dan BMRI dari tahun 2017 sampai dengan tahun 2021. Data yang tersedia berupa data berbentuk *time series* atau rentetan waktu.

3.2. Metode Penelitian

Penelitian ini menggunakan metode pemahaman data dan penyelesaian masalah dengan melakukan komparasi tiga *framework*. Untuk meningkatkan kualitas, efisiensi, dan kevalidan pada prediksi harga saham BBKA, BBRI, dan BMRI pada tahun 2024 mendatang. Perbandingan *framework* yang akan digunakan dijelaskan pada tabel 3.1 sebagai berikut:

Tabel 3.1 Perbandingan *framework* CRISP-DM, KDD, dan SEMMA
Sumber: [30]

Aspek	CRISP-DM	KDD	SEMMA
Singkatan	Cross-Industry Standard Process for Data Mining	Knowledge Discovery in Databases	Sample, Explore, Modify, Model, Assess
Tahap	6 (Pemahaman Bisnis, Pemahaman Data, Pemilihan Data, Pemodelan, Evaluasi, Implementasi)	9 (Pemahaman Domain Aplikasi, Pemilihan Data, Pemrosesan Data, Transformasi Data, Peminangan Data, Interpretasi/Evaluasi, Penggunaan, Penyimpanan Pengetahuan, Pemrosesan Pengetahuan)	5 (Pemilihan Sampel, Eksplorasi, Modifikasi, Pemodelan, Penilaian)
Pendekatan	Terstruktur dengan siklus iteratif	Tidak spesifik, berfokus pada ekstraksi informasi dari data yang besar	Berfokus pada pengembangan model dan analisis statistik
Kelemahan	Membutuhkan waktu yang cukup lama, terutama untuk memahami bisnis dan mengevaluasi model	Tidak memiliki proses terstruktur yang sama jelasnya seperti CRISP-DM	Mungkin kurang efektif untuk masalah yang memerlukan analisis yang mendalam
Kelebihan	Struktur yang jelas, dapat diterapkan di berbagai industri	Fleksibel, dapat menangani berbagai tipe data mining	Fokus pada analisis dan pemodelan secara langsung

Pada tabel 3.1, ketiga *framework* yang umumnya digunakan dalam *data mining* dibandingkan. Ketiga *framework* tersebut adalah CRISP-DM, KDD, dan SEMMA. Penelitian ini menggunakan *framework* CRISP-DM yang didasari dari penelitian terdahulu yang sudah dilakukan karena CRISP-DM memiliki tahapan yang lebih komprehensif [30].

3.3. Alur Penelitian

Pada penelitian ini digunakan *Cross Industry Standard Process for Data Mining* atau biasa dikenal sebagai CRISP-DM sebagai alur penelitian untuk prediksi nilai perusahaan melalui perbandingan harga saham dengan valuasi perusahaan pada tahun 2024. CRISP-DM cocok untuk digunakan pada penelitian yang menggunakan algoritma *Long Short-Term Memory (LSTM)* dan digunakan untuk prediksi. Tahapan dalam CRISP-DM dibagi menjadi 6 bagian yang akan diimplementasikan pada penelitian ini yang akan dijelaskan sebagai berikut:

1. *Business Understanding*

Tahap pertama dalam CRISP-DM adalah tahap *business understanding* dimana pada tahap ini, suatu organisasi atau perusahaan harus mengetahui tujuan dari bisnis atau kegiatan yang akan dijalankan. Pada Desember 2015, jumlah investor di pasar modal berjumlah sekitar 434.107 orang dan pada awal 2018, jumlah investor sudah meningkat menjadi sekitar 635.000. Angka jumlah investor di pasar modal Indonesia ini naik sebanyak 46 persen dalam skala nasional [23]. Dengan data angka jumlah investor yang meningkat, peneliti ingin memprediksi harga saham dan nilai valuasi perusahaan tersebut di tahun 2024 mendatang.

2. *Data Understanding*

Sumber data yang digunakan dalam penelitian ini bersumber dari Yahoo Finance (<https://finance.yahoo.com/>). Dataset yang diperoleh berupa csv yang bersumber dari Yahoo Finance berupa data historikal harga saham BBKA,

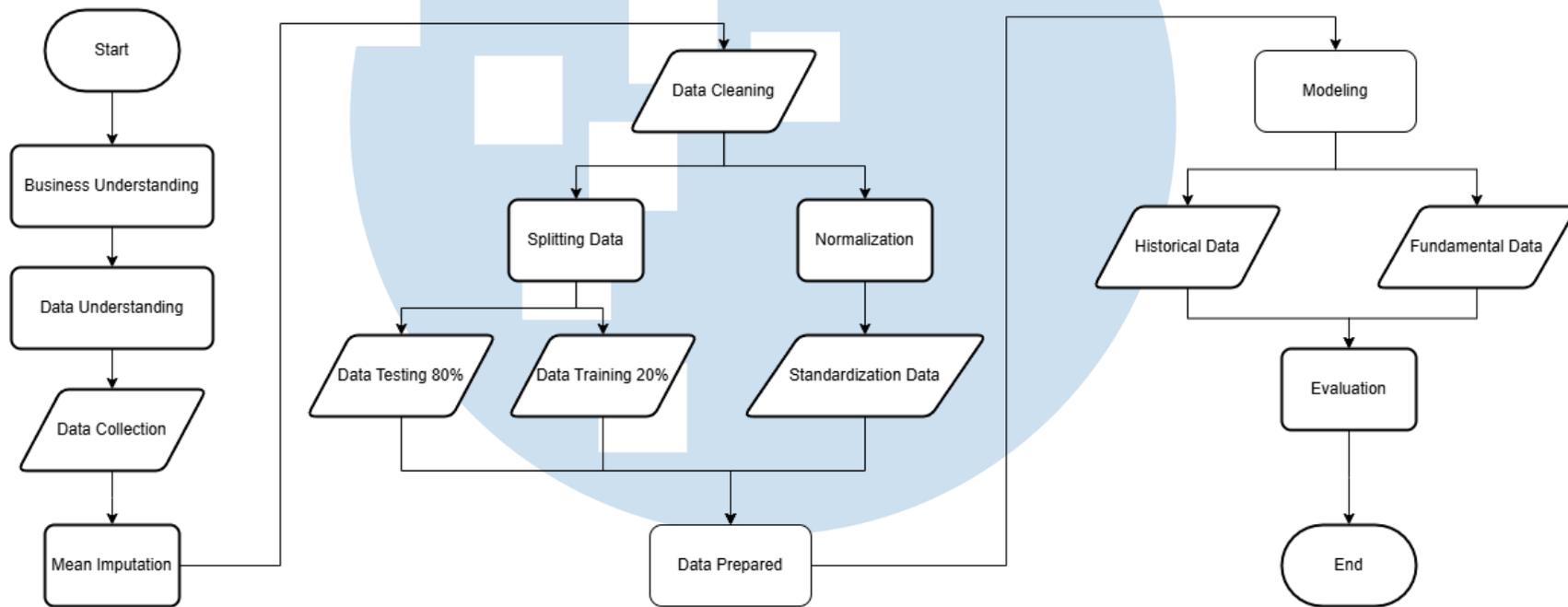
BBRI, dan BMRI dari tanggal 1 Januari 2018 sampai 31 Desember 2023. Bentuk data yang tersedia berupa data rentet waktu atau *time-series*. Data tersebut sesuai dengan penelitian ini karena adanya informasi terkait tanggal, harga pembukaan, harga tertinggi, harga terendah, harga penutupan, penyesuaian harga penutupan, volume transaksi, dan laporan keuangan dari ketiga perusahaan tersebut. Pengambilan data dari Yahoo finance didasari oleh sumber data yang dapat dipercaya yaitu dari Bursa Efek Indonesia sehingga dapat dibuktikan kredibilitas dan keandalannya [31].

3. *Data Preparation*

Pada tahap *data preparation*, dilakukan proses *data collection*, *mean imputation*, *data cleaning* yang terbagi menjadi *split data* dan *normalization*. Data dibagi menjadi *data testing* dengan persentase 80% dan *data training* dengan persentase 20% dan proses *normalization* melakukan proses standarisasi data agar data lebih konsisten. Proses ini dilakukan dengan tujuan untuk memudahkan peneliti dalam melakukan penelitian agar mendapatkan data yang bersih dan meningkatkan hasil akurasi dalam membuat sebuah keputusan.

UMMN

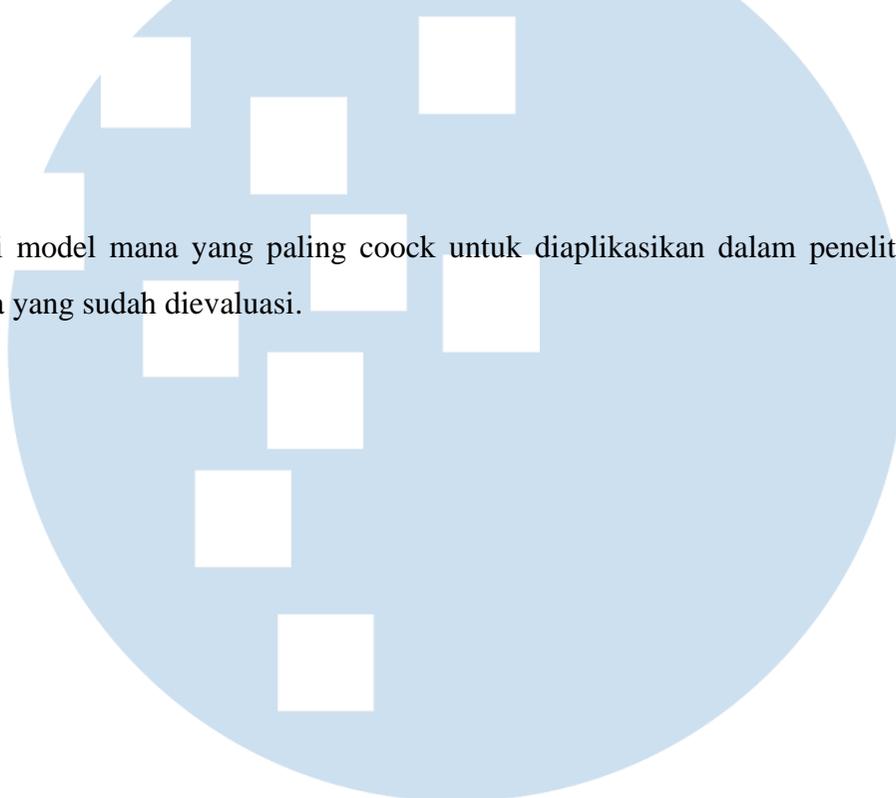
UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.2 Alur metode penelitian

Gambar 3.2 menunjukkan alur metode penelitian yang mengacu pada metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dimana proses metode penelitian mulai dengan pemahaman bisnis, pemahaman data, pengumpulan data yang dilakukan dengan split data 80% untuk data testing dan 20% untuk data training. Penggunaan pembagian data *testing* 80% dan *training* 20% didasari oleh penelitian terdahulu [9], [16]. Setelah dilakukan pembagian porsi data, dilakukan pembuatan model, disini peneliti menggunakan algoritma *Long Short-Term Memory (LSTM)* pada *historical data* dan *fundamental data* dan dilakukan

komparasi agar mengetahui model mana yang paling cocok untuk diaplikasikan dalam penelitian ini. *Output* dari penelitian ini berupa hasil pemodelan data yang sudah dievaluasi.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

4. Modeling

Tahap *Modeling* merupakan tahap pembuatan model dengan beberapa metode algoritma yang sesuai dan telah ditentukan sebelumnya untuk kebutuhan prediksi. Peneliti akan menggunakan bahasa pemrograman *Python* dengan *Integrated Development Environment (IDE) Jupyter Notebook* karena *Jupyter Notebook* mendukung library untuk menganalisis algoritma yang akan digunakan seperti *pandas*, *matplotlib*, dan *sklearn*, serta dapat memvisualisasikan data atau model yang telah dibuat. Dalam tahap ini, peneliti menggunakan 1 algoritma yang sudah dipilih berdasarkan penelitian terdahulu yaitu *Long Short-Term Memory (LSTM)* untuk membuat model prediksi dari objek penelitian. Setelah itu, akan dilakukan proses pengembangan pemodelan algoritma *Long Short Term-Memory* dengan beberapa *optimizer* yaitu *Stochastic Gradient Descent*, *Adagrad*, *Adam*, *Adadelta*, dan *RMSProp*.

5. Evaluation

Pada tahap *evaluation*, proses validasi model yang telah dibuat akan dilakukan dengan melihat tingkat akurasi *p-value*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Nilai *p-value* akan dikatakan baik jika nilai *p-value* lebih besar dari 0.05 (> 0.05), nilai *Root Mean Squared Error* apabila mendekati angka 0 (nol), dan *Mean Absolute Percentage Error (MAPE)* yang mengukur tingkat kesalahan relatif antara prediksi dan nilai sebenarnya dalam bentuk persentase.

3.4. Teknik Pengumpulan Data

Data yang dipakai merupakan data sekunder yang diambil langsung dari website Yahoo Finance (<https://www.finance.yahoo.com/>) berupa data-data sebagai berikut:

1. Data historis saham BBKA, BBRI, dan BMRI berupa tanggal, harga pembukaan, harga terendah, harga penutupan setiap harinya pada tanggal 1 Januari 2018 – 31 Desember 2023.

2. Volume transaksi saham BBKA, BBRI, dan BMRI setiap harinya pada tanggal 1 Januari 2018 – 31 Desember 2023.

3. Laporan keuangan saham BBKA, BBRI, dan BMRI pada tanggal 1 Januari 2018 – 31 Desember 2023.

Pengambilan data dapat dilakukan pada website Yahoo Finance dan Bursa Efek Indonesia, akan tetapi peneliti hanya menggunakan data dari Yahoo Finance karena data yang tersedia pada Bursa Efek Indonesia hanya selama 60 hari terakhir dan belum tersedia secara langsung dalam bentuk *comma seperated values* (csv) sehingga data akan sulit untuk diolah [16], [32].

3.5. Teknik Pengambilan Sampel

Penelitian ini menggunakan metode pengambilan sampel *purposive sampling*. *Purposive sampling* merupakan teknik penentuan sampel dengan pertimbangan tertentu (Sugiyono, 2016). Alasan peneliti menggunakan *purposive sampling* adalah karena metode ini cocok untuk digunakan dengan data kuantitatif dimana data yang digunakan dalam penelitian ini merupakan data kuantitatif. Penelitian ini membagi jumlah sampel menjadi 2 bagian yaitu 80% untuk data *training* (pelatihan) dan 20% untuk data *testing* (pengujian). Penelitian yang dilakukan oleh Wilda dan Versiandika mendapat kesimpulan bahwa pembagian data *testing* 80% dan data *training* 20% memberikan akurasi model yang lebih tinggi [16].

3.6. Teknik Analisis Data

Tabel 3.2 Komparasi Tools

Tools	Kelebihan	Kekurangan
RStudio	Memiliki library-library dengan berbagai fungsi yang dapat digunakan dalam R, data maupun kode-kode mudah untuk dikompilasi.	<i>User Interface</i> yang kurang bersahabat, dan penggunaannya yang terlalu kompleks.
Jupyter Notebook	Dapat melakukan visualisasi data di notebook, berbagi kode, dan interaksi langsung dengan kode.	Sulit untuk di debug karena jika ada kesalahan pada kode, sulit untuk mengidentifikasi bagian

		mana yang menjadi sumber masalah.
SAS Studio	Mempermudah perhitungan statistika, hasil akurat dan reliabel, dan tidak membutuhkan spesifikasi yang tinggi.	Terbatas untuk konsep statistika sosial dan belum terintegrasi dengan program database.

Tabel 3.2 berisikan tools yang diimplementasikan oleh penulis pada penelitian ini. Jupyter Notebook dipilih untuk menjadi teknik analisa data pada penelitian ini karena dapat melakukan pemodelan prediksi, proses *cleaning data*, dapat memvisualisasikan model prediksi, dan merupakan gabungan dari berbagai bahasa pemograman sehingga membuat data lebih mudah untuk dibaca dengan *output* yang interaktif [29], [33].

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA