

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian dilakukan dengan menganalisis prediksi harga saham pada tiga bank digital. Objek pada penelitian ini merupakan harga dari saham, terutama pada kenaikan dan penurunan pada harga saham Bank Digital dan harga jual saham kedepannya. Sekaligus juga membandingkan performa dalam melakukan prediksi dari algoritma atau dari kedua metode yang digunakan dalam analisis harga saham Bank Digital. Dalam memperoleh dataset untuk saham Bank Digital, data yang digunakan berasal dari situs *website*, yaitu Yahoo Finance, dengan mengambil data dari 3 pergerakan harga saham Bank Digital yang terdiri dari; PT Bank JAGO tbk (ARTOJK), PT Allo Bank Indonesia tbk (BBHI.JK) dan PT Bank Neo Commerce tbk (BBYB.JK). Pemilihan tiga dataset ini bertujuan untuk memungkinkan perbandingan yang fokus dan mendalam pada bank-bank yang murni digital. Praktik ini sejalan dengan penelitian terdahulu di mana perbandingan model menggunakan beberapa dataset representatif, seperti studi oleh Fan et al. [5] dan Sulthan Ali [10] yang juga menggunakan tiga dataset saham berbeda. Dari data bank digital yang didapatkan tersebut, penelitian ini menggunakan data dari kurun waktu 6 tahun terakhir, dari tanggal 01 Januari 2019 hingga 17 April 2025. Pada dataset tersebut, masing-masing datanya terdiri dari 7 atribut dari *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close* dan *Volume*.

Penelitian ini dilakukan, untuk dilakukan analisis dan membandingkan perilaku atau pergerakan pasar saham atau *stock market* pada Bank Digital. Dengan menggunakan algoritma *Long Short-Term Memory* (LSTM) dan *Extended Long Short-Term Memory* (xLSTM). LSTM merupakan salah satu jenis algoritma yang ada didalam pembelajaran mesin atau *Machine Learning* yang dimana algoritma ini cocok untuk digunakan dalam memodelkan data yang berupa deret waktu atau *time-series*, yang merupakan dasar pergerakan pada harga saham pada pasar keuangan. Dari data bank digital tersebut dikumpulkan kedalam file dengan format “.csv” lalu

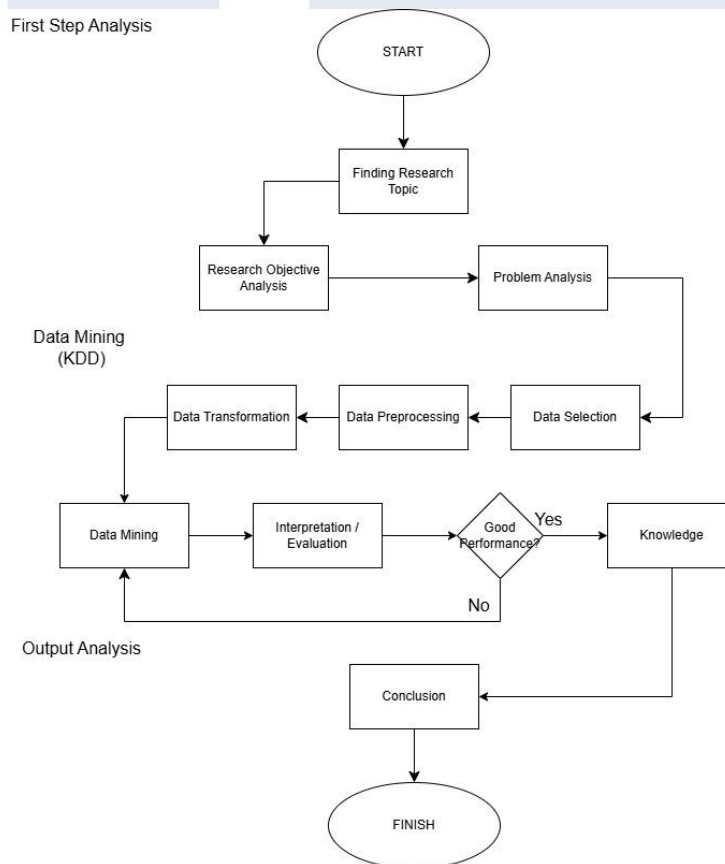
dari hasil data tersebut akan dilakukan penelitian atau analisis prediksi dan juga akan dilakukan pengujian untuk performa dan tingkat presisi dari metode atau algoritma yang digunakan didalam penelitian ini, yang bertujuan untuk menghasilkan suatu temuan empiris mengenai perbandingan kinerja kedua model tersebut.

3.1.1 Metode Penelitian

Penelitian ilmiah yang dilakukan termasuk kedalam jenis penelitian kuantitatif, dikarenakan penelitian ini melibatkan data berbentuk angka dan perhitungan atau rumus serta model matematis. Pembuatan model pada penelitian ini didasarkan pada penggunaan data variabel "*Close*". Pemilihan variabel harga penutupan "*Close*" sebagai target prediksi utama merupakan praktik yang telah dilakukan pada penelitian terdahulu yang sejenis, seperti yang dilakukan oleh Mutmainah et al [2]. dan Wathani et al [9]. Variabel yang digunakan memiliki data dalam bentuk numerik atau angka dan melibatkan perhitungan rumus dalam algoritma *machine learning*, yaitu algoritma xLSTM dan LSTM yang dimana ditujukan untuk menghasilkan nilai prediksi, juga nilai ukuran pada model dan performa prediksi pada data dari model yang dibangun dalam penelitian ini. Evaluasi model dilakukan berdasarkan perhitungan rumus, dan rumus tersebut termasuk pada proses jenis penelitian kuantitatif.

3.1.2 Alur Penelitian

Pada gambar 3.1 Merupakan alur penelitian dengan serangkaian proses yang dimana akan dilaksanakan pada penelitian selama penelitian itu berlangsung dan juga merupakan sebagai acuan dari penelitian agar dapat dilaksanakan dengan terstruktur yang bertujuan agar mendapatkan informasi yang sesuai dengan tujuan dari penelitian.



Gambar 3.1 Alur penelitian

Proses dari KDD bertujuan untuk melakukan identifikasi pola yang valid, baru, memiliki potensi yang berguna dan dapat dipahami dari data [17]. Dengan demikian, penelitian ini secara khusus berfokus pada perbandingan performa algoritma xLSTM dan LSTM dalam melakukan prediksi terhadap harga saham, yang dimana perihal pemilihan metode KDD ini digunakan karena xLSTM merupakan algoritma yang dapat dikatakan relatif baru dan masih didalam tahap eksploratif di dunia akademik ataupun industri. Oleh karena itu, pendekatan dan

penggunaan metode KDD dianggap yang paling memadai dan relevan, karena metode tersebut menekankan pada proses penemuan pola dan evaluasi hasil model secara menyeluruh, yang paling penting adalah dalam konteks pembuktian efektivitas algoritma xLSTM. Dengan menggunakan *framework* KDD, penelitian diharapkan dapat berjalan secara sistematis sekaligus menghasilkan pengetahuan yang dapat dimanfaatkan secara praktis oleh para investor atau pemangku kepentingan dalam pengambilan keputusan dalam berbisnis dan investasi yang berbasis data.

Dengan demikian, tahapan proses dari metode KDD dalam penelitian ini diuraikan secara rinci pada poin-poin berikut.

1.) Pemilihan Data (*Data Selection*)

Tahap pertama, Memahami bisnis dan tujuan analitik yang ingin dicapai pada penelitian. Oleh karena itu, perlu dilakukan proses pemilahan, identifikasi masalah atau peluang bisnis yang ingin dipecahkan berbasis data, seperti meningkatkan penjualan, mengoptimalkan persediaan, atau memahami perilaku pelanggan, dan memperoleh pemahaman terhadap tujuan dari penelitian. Sebagaimana yang sudah ditentukan pada batasan masalah serta tujuan dari penelitian. Tujuan penelitian ini adalah agar membantu para investor sekaligus pemegang saham untuk mengambil keputusan dalam berbisnis pada bidang saham Bank Digital.

	Date	Open	High	Low	Close	Adj Close	Volume
0	17/04/2025	189	192	187	190	190	14623500
1	16/04/2025	192	198	186	189	189	14893500
2	15/04/2025	195	198	190	191	191	23649100
3	14/04/2025	171	200	171	195	195	46175100
4	11/04/2025	174	178	171	171	171	6855300

Gambar 3.2 Data Hasil *Scrapping*

Pada penelitian ini, dilakukan pemilahan dataset yang digunakan untuk memprediksi harga saham yang ditunjukkan pada gambar 3.2, data historis saham

yang digunakan adalah data dari PT Bank JAGO tbk (ARTO.JK), PT Allo Bank Indonesia tbk (BBHI.JK) dan PT Bank Neo Commerce (BBYB.JK). Pada data yang digunakan ditentukan juga tanggal, bulan dan tahun, yaitu mulai dari 01 Januari 2019 hingga 17 April 2025. Kemudian dipilih berdasarkan relevansi variabel yang digunakan untuk model prediksi adalah harga penutupan “*close price*”.

2.) Preprocessing

Pada tahap ini dilakukan eksplorasi dataset yang akan digunakan dalam analitik. Identifikasi sumber data, pemahaman struktur data, dan konten data yang relevan untuk analisis bisnis yang diinginkan. Pada penelitian ini, data yang digunakan adalah data saham Bank Digital, dan data yang digunakan adalah data dari kurun waktu 5 tahun terakhir, dari tanggal 01 Januari 2019 hingga 17 April 2025. Data di ambil dengan menggunakan *tag* HTML to *Dataframe* dikarenakan jika ingin melakukan *download* secara langsung pada *website* Yahoo finance, dibutuhkan fitur berlangganan.

```
# Load the uploaded CSV file
file_path = f'/content/BBYB (Bank NEO).csv'
df = pd.read_csv(file_path, delimiter=';')

# Convert 'Close' column to numeric, handling errors
df['Close'] = pd.to_numeric(df['Close'], errors='coerce')

# Drop rows with NaN values in 'Close' column, which result from the 'coerce' option
df.dropna(subset=['Close'], inplace=True)

df.head()
```

Gambar 3.3 Proses *Missing Value*

Pada gambar 3.3 dilakukan proses pembersihan dan persiapan data, sebelum data digunakan untuk pelatihan model, pada tahapan ini termasuk juga penanganan pada data “NaN” (*missing value*), data yang tidak relevan ataupun *outlier* yang dapat mengganggu hasil analisis.

3.) Transformasi Data (*Transformation Data*)

Pada tahap ini dilakukan perubahan atau modifikasi data dalam format tertentu untuk agar format data siap untuk diproses oleh model pembelajaran atau *machine learning*. Transformasi ini mencakup berbagai teknik yang

digunakan untuk memperbaiki kualitas data sekaligus menyelaraskan data agar lebih efektif dan efisien pada saat analisis atau dalam pelatihan model. Tahap ini melibatkan perubahan atau normalisasi data, supaya data lebih mudah diproses oleh model. Proses ini mencakup perubahan skala data (penggunaan *scaling* atau normalisasi), *encoding* variabel kategorikal maupun perubahan struktur data.

Pada penelitian ini, penting dilakukan proses transformasi data, yang dijelaskan sebagai berikut:

1) Normalization

Normalisasi pada penelitian ini merupakan langkah untuk memastikan bahwa model xLSTM dan LSTM dapat bekerja dengan efisien, dikarenakan harga penutupan saham (*close*) memiliki rentang yang besar.

2) MinMaxScaler

MinMaxScaler digunakan untuk menormalkan harga saham ke dalam rentang “0” hingga “1”. Jadi normalisasi ini membantu untuk menghindari bias atau prasangka yang disebabkan oleh fitur atau fungsi yang memiliki rentang nilai yang berbeda.

4.) Data Mining

	Date	Open	High	Low	Close	Adj Close	Volume
0	2019-01-02	238.50	260.18	229.46	247.53	247.21	1725380.0
1	2019-01-03	247.53	252.95	245.72	249.34	249.02	833301.0
2	2019-01-04	251.14	251.14	243.92	247.53	247.21	27452.0
3	2019-01-07	251.14	251.14	243.92	249.34	249.02	34315.0
4	2019-01-08	251.14	251.14	243.92	249.34	249.02	77817.0

Gambar 3.4 Hasil Data Transformation

Data mining merupakan tahapan dimana model pembelajaran (*machine learning*) diterapkan pada data untuk melakukan ekstraksi pola atau pengetahuan. Pada tahap ini, model dilatih untuk mempelajari pola dalam data sekaligus menggunakannya untuk membuat model prediksi ataupun klasifikasi.

Pada gambar 3.3 merupakan hasil data yang sudah mengalami perubahan, normalisasi, *cleaning* dan perubahan format pada tanggal yang sudah siap untuk dimasukkan sebagai input dalam *machine learning*. Pada penelitian ini, metode xLSTM dan LSTM digunakan untuk memprediksi naik dan turunnya harga saham pada beberapa hari kedepan (*Day+1* sampai *Day+5*). Model pembelajaran ini dilatih menggunakan data pelatihan (*data train*) dan diuji pada data pengujian (*data test*). Selama pelatihan, model belajar pola temporal didalam harga saham yang terdapat pada data historis saham.

5.) Interpretation / Evaluation

Setelah model dibangun, pada tahap ini melibatkan evaluasi terhadap performa model yang dimana tahap ini merupakan penilaian terhadap kinerja model yang didasari oleh metrik yang digunakan. Evaluasi merupakan tahapan yang penting pada proses penelitian ini, dikarenakan tahap ini memiliki fungsi untuk mengetahui seberapa baik performa pada model xLSTM dan LSTM pada saat sudah dikembangkan dalam menghasilkan prediksi dan nilai ukuran yang dihasilkan pada harga saham Bank Digital. Kemudian metrik evaluasi yang digunakan pada proses evaluasi ini adalah: RMSE, MSE, MAE, MAPE dan R^2 .

6.) Pengetahuan (Knowledge)

Tahap *knowledge* ini bertujuan untuk menilai atau mengukur performa model dengan menggunakan nilai ukuran atau metrik evaluasi yang sudah ditetapkan dan disesuaikan sebelumnya. Pada penelitian ini dilakukan evaluasi dari hasil prediksi xLSTM dan LSTM dengan menggunakan beberapa metrik utama, yaitu:

- a. RMSE (*Root Mean Square Error*) dan MAE (*Mean Absolute Error*) merupakan metrik utama untuk mengukur kesalahan pada prediksi (*error rate*). Ukuran RMSE memberikan penalti lebih besar untuk kesalahan besar, sementara itu, MAE mengukur kesalahan rata-rata tanpa memperhatikan outliers.

- b. MAPE (*Mean Absolute Error*) digunakan untuk mengukur kesalahan relatif antara prediksi dan nilai aktual, yang menunjukkan seberapa jauh model atau metode yang digunakan melakukan kesalahan dalam memprediksi naik atau turunnya harga saham.

3.2 Variabel Penelitian

variabel penelitian yang terdapat pada masing-masing data saham Bank Digital yang diperoleh dari *website* Yahoo Finance. Berikut ini merupakan variabel dari penelitian:

- 1) **Variabel Independen:** Pada penelitian ini terdapat variabel independen dari masing-masing *dataset* tersebut, yaitu atribut “*Close*” sebagai variabel independen.
- 2) **Variabel Dependen:** Pada penelitian ini variabel yang menjadi target prediksi adalah harga penutupan saham harian, yaitu variabel “*Close*” yang sudah diprediksi menggunakan metode xLSTM dan LSTM.

3.3 Teknik Pengumpulan Data

Data yang digunakan pada penelitian ilmiah ini, yaitu data saham bank digital yang diambil dari *website* Yahoo finance yang terdiri dari tiga dataset pergerakan harga saham bank digital yang terdiri dari; Bank JAGO tbk (ARTO.JK), Bank Allo Indonesia tbk (BBHL.JK) dan Bank Neo *Commerce* tbk (BBYB.JK). Dari data yang didapatkan tersebut, penelitian ini menggunakan data dari kurun waktu 6 tahun terakhir, dari tanggal 01 Januari 2019 hingga 17 April 2025 dengan total 4,610 baris data yang dimana masing-masing data tersebut terdiri dari kurang lebih 1537 baris data.

Berikut ini merupakan penjelasan variabel data pada masing-masing *dataset* yang dapat dilihat pada tabel 3.2.

Tabel 3.1 Atribut dataset

Variabel	Penjelasan
<i>Date</i>	Tanggal

Variabel	Penjelasan
<i>Open</i>	Harga pembukaan saham harian
<i>High</i>	Harga saham tertinggi di hari tersebut
<i>Low</i>	Harga saham terendah di hari tersebut
<i>Close</i>	Harga penutupan saham harian
<i>Adj close</i>	Harga saham penutup yang disesuaikan ketika terjadi aksi korporasi perusahaan
<i>Volume</i>	Jumlah transaksi saham yang diperdagangkan

3.4 Visualisasi

Visualisasi merupakan representasi grafis yang berasal dari hasil informasi pada data, dengan menggunakan elemen visual seperti grafik, peta, tabel dan bagan. Visualisasi bertujuan untuk menentukan informasi apa yang akan divisualisasikan sesuai dengan tujuan masalah dari suatu penelitian. Langkah visualisasi dimulai dari, jenis proyek yang akan digunakan, melakukan identifikasi masalah, memilih data yang akan digunakan untuk dianalisis, mengubah data dengan mengganti tipe data [24] dan memastikan data sudah siap untuk dilakukan pemrosesan visualisasi. Pada penelitian ini, hasil prediksi yang dilakukan oleh metode xLSTM dan LSTM divisualisasikan dalam bentuk grafik dengan tujuan untuk membandingkan antara harga prediksi (*predicted price*) dan harga aktual (*actual price*).

3.5 Teknik Analisis Data

Penelitian dilakukan dengan landasan kerangka kerja KDD dengan algoritma *deep learning* dan sub-bidang *Machine Learning*. Metode *Deep Learning* yang digunakan adalah metode *Extended Long Short-Term Memory* (xLSTM) dan *Long Short-Term Memory* (LSTM) yang dibantu oleh sub-bidang dari metode *Machine Learning* yaitu metode *Forecasting*. xLSTM merupakan algoritma yang dikembangkan dari LSTM yang pada umumnya digunakan untuk melakukan prediksi atau analisis terhadap penelitian yang bersifat deret waktu atau *time-series* yang dimana kedua algoritma tersebut digunakan untuk menghasilkan model prediksi pada tingkat performa yang baik dan tingkat presisi pada saham Bank

Digital didalam penelitian ini [25]. Proses analisa dan proses pengolahan data, menggunakan bantuan bahasa pemrograman Python yang dijalankan menggunakan Google Colab.

Setelah dilakukan proses pelatihan pada model *machine learning*, metode *forecasting* akan diterapkan untuk menggambarkan hasil prediksi pada harga saham Bank Digital untuk kedepannya [26], hal ini dapat membantu para pemegang saham ataupun para investor yang berinvestasi ke perusahaan Bank Digital dalam melakukan pengambilan keputusan untuk menjual saham tersebut dan mendapatkan keuntungan ataupun menahan untuk menjual *lot* sahamnya dikarenakan mereka mendapat bantuan atau bisa membaca dalam perihal memprediksi harga *lot* saham tersebut kedepannya. Sehingga hal tersebut dimungkinkan karena adanya prediksi dari model yang menghasilkan tingkat performa dan tingkat presisi yang tinggi [22].

Sumber: [10]

Tabel 3.2 Tabel parameter

No	Parameter	Keterangan
1	<i>Optimizer</i>	ADAM
2	<i>Batch Size</i>	32
3	<i>Loss Function</i>	MSE
4	<i>Learning Rate</i>	0.001 (<i>Default</i>)
5	<i>Training Epoch</i>	50

Pada tabel 3.3 merupakan parameter yang digunakan pada penelitian ini, parameter tersebut digunakan dalam pelatihan model xLSTM dan LSTM. Dengan menggunakan ADAM sebagai *optimizer*, ukuran batch (*batch size*) 32 dengan tujuan penyeimbangan efisiensi dan performa, MSE sebagai *loss function*, *learning rate* 0.001 yang dimana merupakan *default learning rate* dari ADAM, pelatihan model dilakukan sebanyak 50 *epoch*.