

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini adalah untuk melakukan analisis prediksi harga komoditas emas dan perak dengan menggunakan algoritma LSTM, penggunaan LSTM tersebut dikembangkan dengan beberapa arsitektur LSTM, seperti vanilla LSTM, stacked LSTM, dan Bi-LSTM hal tersebut berguna untuk membandingkan arsitektur yang mana paling baik digunakan untuk melakukan prediksi harga emas dan perak. Kemudian juga dari penggunaan tersebut dikembangkan lagi dengan menambahkan *activation*, *optimizer techniques* dan melakukan tuning parameter berguna untuk melakukan uji model mana yang paling baik dalam melakukan prediksi harga emas dan perak. *Activation* yang digunakan yaitu Linear, ReLU, Tanh, dan Sigmoid. Sedangkan untuk *optimizer* yang digunakan yaitu Adam, AdaGrad, Nadam, RMSProp, AdaDelta, dan SGD. Penerapan arsitektur LSTM, *activation*, dan *optimizer* berguna untuk melakukan perbandingan tingkat *error rate* terendah dalam melakukan prediksi komoditas emas dan perak. Penelitian ini juga menggunakan beberapa evaluasi, seperti *mean squared error* (MSE), *root mean squared error* (RMSE), *mean absolute error* (MAE), *mean absolute percentage error* (MAPE), *r-squared*, dan untuk test normalisasi menggunakan *Kolmogorov-Smirnov*. Dalam melakukan analisis tersebut dibutuhkan *dataset* yang didapat melalui *website* Investing, di mana data tersebut diambil berdasarkan kode komoditasnya, seperti emas (XAU) dan perak (XAG) dalam hal ini pair yang digunakan yaitu USD. Data tersebut diambil selama jangka waktu 19 tahun terakhir, dari tanggal 3 Januari 2005 hingga 31 Desember 2024. Dengan menggunakan *time frame* 1 hari.

3.2 Metode Penelitian

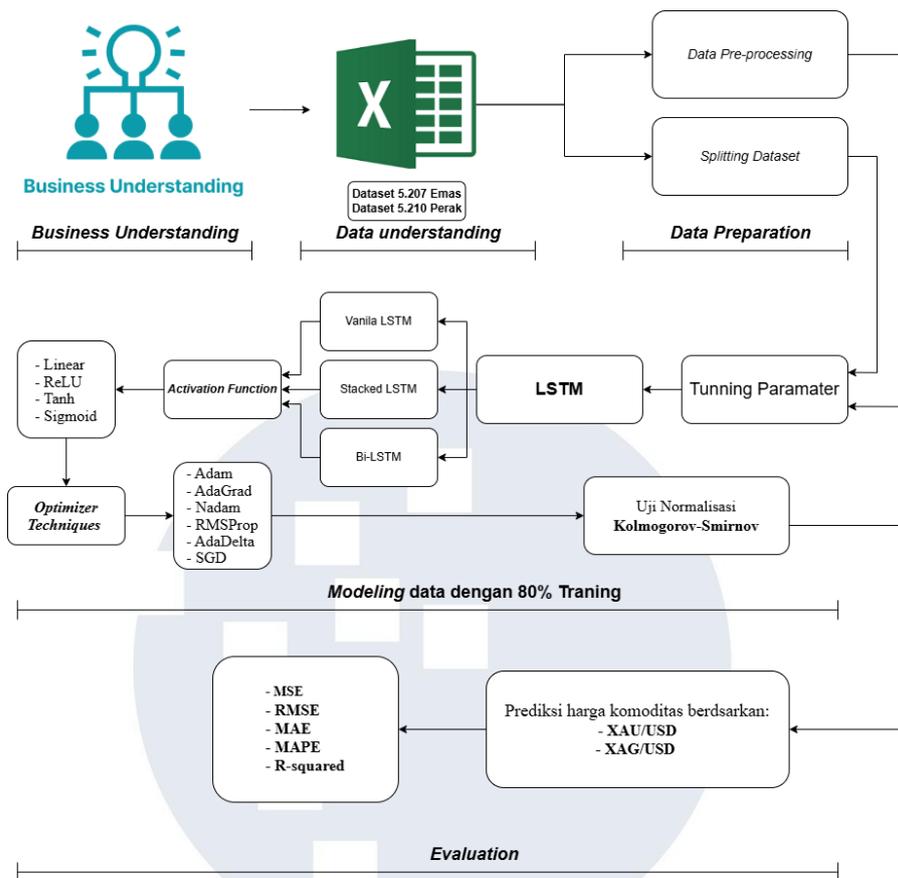
3.2.1 Alur penelitian

Dalam menentukan metodologi diperlukannya analisis untuk mendapatkan jenis *framework* yang sesuai dalam penelitian ini. Berikut merupakan tiga *framework* yang akan dibandingkan pada tabel 3.1 di bawah. Yang berisikan CRISP-DM, KDD, dan SEMMA.

Tabel 3. 1 Tabel perbandingan *framework* [51] , [52]

Indikator	CRISP-DM	KDD	SEMMA
Pengertian	CRISP-DM framework (<i>Cross-Industry Standard Process for Data Mining</i>) adalah sebuah teknik metodologi yang digunakan dalam proses data <i>mining</i> dan analisis data.	KDD (<i>Knowledge Discovery in Database</i>) adalah proses <i>framework</i> menemukan pengetahuan yang bermanfaat dari kumpulan data.	SEMMA adalah <i>framework</i> yang digunakan untuk membuat model prediktif. Pada penerapannya penggunaan SEMMA sering digunakan pada aplikasi sias dalam melakukan analisis data <i>mining</i> .
Langkah-langkah	<ol style="list-style-type: none"> 1. <i>Business Understanding.</i> 2. <i>Data understanding.</i> 3. <i>Data preparation.</i> 4. <i>Modeling.</i> 5. <i>Evaluation.</i> 6. <i>Deployment.</i> 	<ol style="list-style-type: none"> 1. <i>Selection</i> 2. <i>Preprocessing.</i> 3. <i>Transformation.</i> 4. <i>Data mining.</i> 5. <i>Evaluation.</i> 	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assess</i>
Kelebihan	<ol style="list-style-type: none"> 1. Sudah banyak digunakan 2. Memiliki struktur yang mudah dimengerti 3. Lebih menekankan pada kebutuhan bisnis 4. Bersifat fleksibel 	<ol style="list-style-type: none"> 1. Lebih fokus pada <i>knowledge discovery process</i> 2. Melingkupi seluruh proses <i>data mining</i> 3. Dapat dilakukan 	<ol style="list-style-type: none"> 1. Mudah untuk diterapkan 2. SEMMA memberikan pendekatan terstruktur

Indikator	CRISP-DM	KDD	SEMMA
		<p>secara berulang agar hasil lebih baik.</p> <p>4. Hasil eksplorasi data yang mendalam.</p>	<p>yang relatif sederhana dalam pengembangan model prediktif.</p> <p>3. Memiliki fleksibilitas dalam proses data <i>exploration</i>.</p>
Kekurangan	<p>1. Penerapannya kurang spesifik.</p> <p>2. Memerlukan jangka waktu yang cukup Panjang.</p>	<p>1. Proses KDD cenderung kompleks dan memerlukan pemahaman yang mendalam.</p> <p>2. Keberhasilan KDD sangat bergantung pada kualitas dan kuantitas data.</p>	<p>1. Tidak memiliki proses bisnis <i>understanding</i>.</p> <p>2. SEMMA mungkin kurang cocok untuk proyek-proyek yang kompleks.</p> <p>3. Memiliki kekurangan pada tahap evaluasi dan penerapannya.</p>



Gambar 3. 1 Alur kerja penelitian dengan CRISP-DM

Berdasarkan kelebihan dan kekurangan yang didapatkan dapat disimpulkan bahwa dalam pada penelitian ini akan menerapkan *framework* sesuai untuk yang digunakan yaitu CRISP-DM *framework* Cross-Industry Standard Process for Data Mining adalah sebuah teknik metodologi yang digunakan dalam proses data *mining* dan analisis data. Alasan kuat pada penggunaan CRISP-DM pada penelitian ini karena menekankan pada hasil evaluasi. Pada penelitian ini CRISP-DM dapat memberikan *framework* yang lebih terstruktur dalam melakukan prediksi harga emas dan perak. Berikut merupakan tahapan dalam menerapkan CRISP-DM, yaitu:

1. **Business understanding:** Pada tahap ini, melakukan analisis bisnis untuk menemukan atau menghasilkan gambaran yang dibutuhkan secara keseluruhan. Pada tahap ini juga melakukan analisis sumber data yang dapat diperoleh dalam menunjang penelitian ini dan melakukan analisis tujuan dari data *mining*. Pada gambar 3.1 merupakan alur kerja yang telah dirancang untuk melaksanakan penelitian ini.

2. **Data understanding:** Pada tahap kedua, melakukan pengumpulan data dari sumber situs *website* Investing. Kemudian hasil dari data yang diperoleh akan dilakukan eksplorasi data, pemahaman data, pengecekan kualitas data, dan melakukan analisis data. Data yang diperoleh yaitu emas untuk XAU dan perak untuk XAG. Data yang digunakan dengan jangka waktu 19 tahun terakhir, dari tanggal 3 Januari 2005 hingga 31 Desember 2024. Data tersebut berisikan 7 kolom yaitu tanggal, terakhir, pembukaan, tertinggi, terendah, vol., dan perubahan%.

Tabel 3. 2 Jumlah *dataset* XAU dan XAG

Dataset	Jumlah kolom	Jumlah baris
XAU/USD	7	5.207
XAG/USD	7	5.210

3. **Data preparation:** Pada tahap ketiga, melakukan seleksi data dengan menentukan data yang akan digunakan dan data yang akan dibuang. Pada tahap ini juga akan melakukan pengecekan *missing* data dan melakukan pembersihan data untuk menghasilkan data yang *valid* yang akan digunakan selama penelitian ini. Pada tabel 3.3 merupakan data *missing value* di mana dapat kita lihat bahwa data emas dan perak Vol. tidak ada maka dari itu akan dihapus. Maka dari itu data yang akan disiapkan, yaitu tanggal, terakhir, pembukaan, tertinggi, terendah, dan perubahan%. Data yang sudah diproses kemudian diolah dengan cara, mengganti koma menjadi titik pada data terakhir, pembukaan, tertinggi, dan terendah untuk data di jadikan *float*. Kemudian mengubah format tanggal menjadi *date time*, setelah tanggal diubah lalu diurutkan dari bawah ke atas atau *ascending*. Setelah selesai membersihkan data kemudian masuk ke tahap selanjutnya yaitu pembagian untuk testing dan *training* pada penelitian ini menggunakan dua cara untuk melakukan pengetestan testing *training* yaitu pada *base* model LSTM ada 80% *training* dan 20% *testing* serta 70% *training* dan 30% *testing* yang kemudian hasil terbaiknya di bandingkan, yang untuk hasil terbaiknya kan dilanjutkan untuk *performance*

model LSTM. Setelah semua data selesai di proses masuk ke tahap selanjutnya yaitu modeling.

Tabel 3. 3 Data *missing value* XAU dan XAG

Data	Emas (XAU)	Perak (XAG)
Tanggal	0	0
Terakhir	0	0
Pembukaan	0	0
Tertinggi	0	0
Terendah	0	0
Vol	5.206	5.210
Perubahan	0	0

4. **Modeling:** pada tahap keempat, yaitu pemodelan data pada fase ini data yang telah disiapkan akan dilakukan analisis dan menentukan teknik pemodelan yang sesuai, melakukan testing data, melihat tingkat *error rate* data, dan melakukan pembuatan model. Pada penelitian ini pemodelan yang digunakan untuk memprediksi harga emas dan perak yaitu dengan menggunakan algoritma *Long Short-Term Memory*, dipilih sebagai metode pemodelan karena algoritma LSTM dirancang khusus untuk menangani data *sequential* dan memiliki kemampuan mengingat pola jangka panjang. Hal tersebut yang membuat algoritma LSTM sangat relevan untuk prediksi harga emas dan perak sebagai data *time series*. Pada penelitian ini memang di fokuskan untuk menggunakan algoritma LSTM, maka kemudian ditambahkan juga dengan bantuan arsitektur LSTM, seperti Vanilla LSTM, Stacked LSTM, dan Bi-LSTM. Kemudian setelah menggunakan 3 arsitektur tersebut ditambahkan *Activation* antara lain, Linear, ReLU, Tanh, dan Sigmoid. Setelah menambahkan *Activation* disempurnakan dengan *optimizer* untuk meningkatkan hasil prediksi, yaitu digunakan yaitu Adam, AdaGrad, Nadam, RMSProp, AdaDelta, dan SGD. Setelah selesai melakukan pemodelan tahap selanjutnya yaitu melakukan testing terhadap normalisasi dengan menggunakan *Kolmogorov-Smirnov*. Setelah selesai melakukan pemodelan data masuk ke tahap selanjutnya yaitu *evaluation*.

5. **Evaluation:** pada tahap kelima, yaitu evaluasi melakukan pengecekan dari hasil pemodelan data yang akan diperiksa dan dievaluasi untuk mengetahui apakah data yang telah dibuat sesuai dengan tujuan dari penelitian ini. Tahap evaluasi merupakan tahap yang sangat penting pada penelitian ini dikarenakan untuk mengetahui bahwa algoritma LSTM lebih baik diterapkan antara emas dan perak. Pada tahap evaluasi melakukan perbandingan hasil dengan menggunakan *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *r-squared*,. Hasil prediksi dan pemodelan tersebut akan dibandingkan untuk mengetahui komoditas mana yang lebih cocok menggunakan algoritma LSTM untuk memprediksinya.

3.3 Teknik Pengumpulan Data

Pada penelitian ini menggunakan *dataset* komoditas mineral dari situs *website* Investing dengan mengambil dua data yaitu, emas dan perak. Dengan mengambil data dari pergerakan kedua harga emas (XAUUSD) dan perak (XAGUSD). Pada penelitian ini data yang digunakan yaitu dengan *time frame* 1 hari. Dari data yang didapatkan, pada penelitian ini menggunakan data dari kurun waktu 19 tahun terakhir, dari tanggal 3 Januari 2005 hingga 31 Desember 2024. Periode waktu dalam satu hari mencakup 4 sesi, yaitu sesi Asia, Eropa, Amerika, dan Pasifik hal tersebut membuat dalam satu hari perdagangan mencapai 23 jam yaitu buka pasar pada jam 05.00 pada sesi Pasifik dan tutup pasar pada sesi Amerika jam 04.00. Perdagangan juga berlangsung selama lima hari, yaitu Senin, Selasa, Rabu, Kamis, dan Jumat. Pemilihan *dataset* emas dan perak sejak tahun 2005 didasarkan, karena sudah mencerminkan data yang panjang untuk melakukan tren analisis jangka panjang, sehingga mendapatkan pola siklus pasar. Dengan total masing-masing data yaitu 5.207 emas dan 5.210 perak, berikut merupakan penjelasan dari masing-masing *dataset*, yaitu:

Tanggal : Tanggal perdagangan.

Terakhir : Harga penutupan pasar.

Pembuka : Harga pembukaan pasar.

Tertinggi : Harga tertinggi pasar di hari itu.

Terendah : Harga terendah pasar di hari itu.

Vol. : Volume transaksi di hari itu.

Perubahan% : Perubahan harga berdasarkan persentase di hari itu.

Tabel 3. 4 Sumber data XAU dan XAG

Komoditas	Kode Komoditas	Sumber
Emas	XAU/USD [53]	https://www.investing.com/currencies/xau-usd
Perak	XAG/USD [54]	https://www.investing.com/currencies/xag-usd

