

BAB III

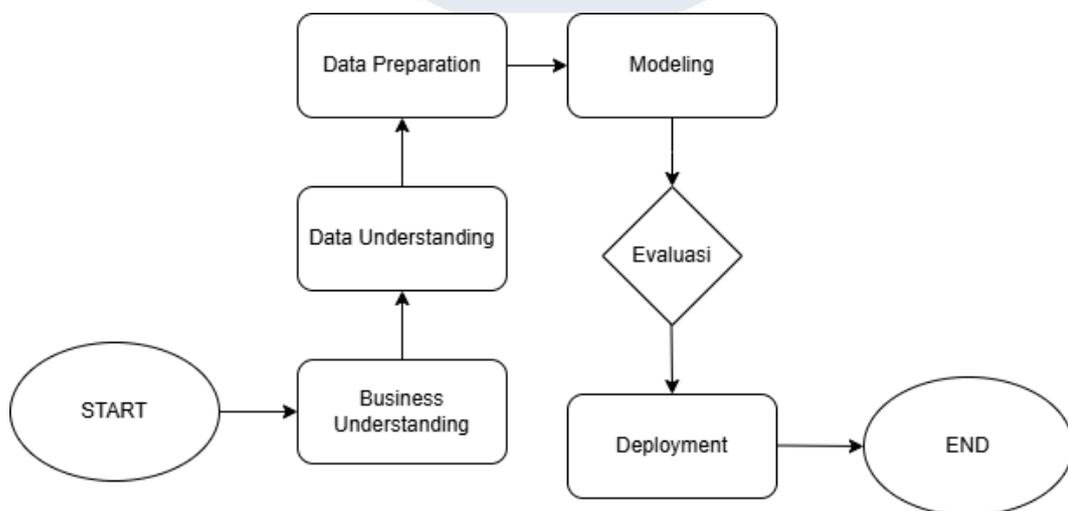
METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini melakukan prediksi pergerakan saham sektor energy minyak dan gas. Terdapat empat emiten yang dipilih untuk penelitian yang dilakukan ini yaitu ada PGAS, MEDC, AKRA, dan juga RAJA. Empat emiten yang dipilih tersebut berdasarkan kapitalisasi pasar. Berdasarkan gambar 1.1, PGAS merupakan kapitalisasi pasar terbesar yang dimana merupakan kapitalisasi saham yang paling berpengaruh pada sektor migas, dan yang menduduki posisi kedua adalah MEDCO lalu setelah itu, terdapat AKRA, RATU, dan RAJA. Emitmen RATU tidak di pilih dalam penelitian ini karena data yang di input pada website investing baru tahun ini saja yang dimana website tersebut adalah tempat peneliti memperoleh data dari keempat emiten tersebut.

3.2 Metode Penelitian

3.2.1 Alur Penelitian



Gambar 3. 1 Alur Penelitian

Seperti gambar 3.2, adalah alur penelitian dari penelitian ini dengan *framework* CRISP-DM yang diawali dengan *business understanding*, *Data*

Understanding, Data Preparation, Modeling, Evaluasi, hingga deployment.[41].

3.2.2 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode CRISP-DM, yang dimana terdapat 6 tahapan dalam metode tersebut diawali dengan, business understanding, data understanding, data preparation, data modeling, evaluation, dan deployment. Digunakannya metode ini dapat membantu mengatasi masalah dalam melakukan prediksi pergerakan saham. Berikut ini penjelasan detail terkait dengan 6 tahapan yang dilakukan penelitian ini:

1. Business Understanding

Tahap ini adalah tahap awal yang dimana merumuskan permasalahan yang ada dalam penelitian ini. Melakukan analisis terhadap pergerakan saham selama 10 tahun dari tahun 2014-2024 untuk memahami perubahan yang terjadi seperti naik turunnya pergerakan saham dari tahun ke tahun.

2. Data Understanding

Tahap ini adalah tahap proses analisis data yang dimana melakukan pemahaman lebih dalam terhadap data yang diambil untuk melakukan prediksi terhadap pergerakan saham, yang dimana pemahaman data yang sudah dimiliki untuk dikelola lebih lanjut.

3. Data Preparation

Tahap ini adalah tahap persiapan dan pembersihan data yang dilakukan setelah data understanding. Pada tahap ini peneliti melakukan ekstraksi data yang kemudian akan dilakukan pemilihan fitur-fitur yang dibutuhkan.

4. Data Modeling

Tahap ini adalah tahap setelah melakukan persiapan data. Pada tahap ini melakukan pemodelan data terhadap data yang sudah siap untuk dikelola lebih lanjut. Tujuan dari pemodelan tersebut yaitu untuk

menggapai pemahaman yang lebih baik, menggambarkan dengan jelas, serta mengorganisir data secara logis hingga datanya tersebut terstruktur hingga dapat dipahami dengan mudah.

5. *Evaluation*

Tahap ini adalah tahap penilaian dan pengujian data. Pada tahap ini, model yang sudah dipilih akan dilakukannya evaluasi untuk melihat performa dari model tersebut.

6. *Deployment*

Tahap ini adalah tahap akhir dalam metode CRISP-DM yang dimana bertujuan untuk menerapkan model yang telah dievaluasi ke dalam lingkungan nyata. Hasil dari model dapat digunakan untuk mendukung pengambilan keputusan, baik melalui laporan, visualisasi, maupun sistem otomatis yang dapat diakses pengguna.

Setelah itu, data mining juga memiliki dua framework lainnya selain CRISP-DM yang diantaranya terdapat SEMMA dan juga KDD. Berikut ini adalah tabel perbandingan dari ketiga framework tersebut:

Tabel 3. 1 Perbandingan *Framework*

Framework	CRISP-DM	SEMMA	KDD
Proses	<i>Business Understanding</i>	<i>Sample</i>	<i>Selection</i>
	<i>Data Understanding</i>	<i>Explore</i>	<i>Preprocessing</i>
	<i>Data Preparation</i>	<i>Modify</i>	<i>Transformation</i>
	<i>Data Modeling</i>	<i>Model</i>	<i>Data Mining</i>
	<i>Data Evaluation</i>	<i>Assess</i>	<i>Interpretation</i>
	<i>Data Deployment</i>	---	<i>Knowledge</i>
Kelebihan	Kemampuan yang fleksibel.	Membantu memahami data yang lebih efisien	Pemahaman yang lebih mendalam sehingga mudah dipelajari.
	Memiliki siklus yang terstruktur	Membantu mendeteksi pola untuk pemodelan data.	Kemampuan untuk mendeteksi pola yang kompleks.

	Membantu membuat hasil yang diinginkan.	Meningkatkan kualitas dari model yang dikembangkan.	Dapat menangani data yang beragam.
Kekurangan	Kesulitan dalam mengidentifikasi kebutuhan yang pasti.	Kurangnya pemahaman tentang bisnis.	Proses yang terlalu Panjang.
	Struktur yang dapat dikatakan kaku walaupun terstruktur terutama apabila data mining membutuhkan evolusi yang cepat.	Eksplorasi data yang terbatas .	Adanya kesulitan terhadap penanganan data yang besar.
	Memiliki sifat yang sekuensial.	Pemodelan yang terbatas.	Campur aduk manusia yang terlalu tinggi sehingga menjadi tantangan bagi proyekproyek yang memerlukan otomatisasi lebih lanjut.

Sesuai tabel 3.1 yang ada diatas ini, setiap framework memiliki kelebihan dan kekurangannya masing-masing tergantung dengan tujuan dan juga jenis data apa yang dihadapi. Setelah itu, peneliti menggunakan metode CRISP-DM karena, framework ini lebih mengutamakan pemahaman bisnis pada awal proses yang dimana, dengan diutamakan pemahaman bisnis, analisis yang dilakukan dapat diidentifikasi sehingga dapat memberikan hasil yang maksimal sesuai dengan tujuan yang dimiliki si peneliti.

3.2.3 Metode Pengolahan Data

Tabel 3. 2 Perbandingan Algoritma

	Algoritma
--	------------------

Indikator	LSTM (Long Short-Term Memory)	GRU (Gated Recurrent Unit)	1D-CNN (1-Dimensional Convolutional Neural Network)
Tujuan	Mengatasi vanishing gradient dalam RNN dan mampu menangani urutan panjang.	Menyediakan alternatif yang lebih sederhana dan lebih cepat dibandingkan LSTM dengan performa yang kompetitif.	Mengekstraksi fitur spasial dari data deret waktu menggunakan filter konvolusi. Cocok untuk mendeteksi pola lokal.
Kelebihan	Mampu menangani dependensi jangka panjang dengan baik.	Arsitektur lebih sederhana dibandingkan LSTM, sehingga lebih cepat untuk dilatih dan digunakan.	Lebih cepat dalam pelatihan dibanding RNN.
	Lebih fleksibel dalam menangani berbagai jenis data sekuensial.	performa yang mendekati LSTM	Mampu mengekstraksi fitur lokal yang kuat.
	Cocok untuk tugas yang membutuhkan memori jangka Panjang.	Lebih sedikit parameter.	Tidak rentan terhadap vanishing gradient.
Kekurangan	Memiliki lebih banyak parameter.	Kurang efektif untuk menangani urutan yang Panjang.	Tidak memiliki mekanisme memori eksplisit.
	Mebutuhkan lebih banyak daya komputasi dibandingkan GRU.	Kurang fleksibel dalam menangani data dengan dependensi jangka panjang	Kurang baik dalam menangani dependensi jangka panjang.
	Mebutuhkan lebih banyak daya komputasi dibandingkan GRU.	Kurang detail menangani data.	-
	Memiliki tiga gerbang utama: Forget Gate, Input	Menggunakan dua gerbang utama: Reset Gate dan Update Gate.	Menggunakan filter konvolusi 1 dimensi untuk mengekstraksi pola dari urutan data.

Proses	Gate, dan Output Gate.		
	Forget Gate menentukan informasi mana yang harus dibuang.	Reset Gate menentukan seberapa banyak informasi dari masa lalu yang harus dilupakan.	Pooling layer mengurangi dimensi.
	Input Gate memperbarui informasi yang akan disimpan.	Update Gate mengontrol seberapa banyak informasi baru yang akan disimpan.	Tidak ada memori eksplisit seperti RNN/LSTM.
	Output Gate menghasilkan output berdasarkan sel memori.	Tidak memiliki sel memori eksplisit seperti LSTM; memori langsung dikelola oleh unit tersembunyi.	-

Berdasarkan tabel 3.2 yang ada diatas ini, peneliti membandingkan dua algoritma yaitu LSTM dan GRU dalam melakukan prediksi pergerakan saham. Dari masing-masing algoritma tersebut memiliki kelebihan dan juga kekurangannya, dimulai dari LSTM mengatasi vanishing gradient dalam RNN dan mampu menangani urutan Panjang dan GRU yang menyediakan alternatif yang lebih sederhana dan lebih cepat dibandingkan LSTM dengan performa yang kompetitif dan lain-lainnya. Kemudian dari segi kekurangan LSTM yang dimana Memiliki lebih banyak parameter sehingga lambat dalam pelatihan dan inferensi, dan untuk kelemahan GRU itu sendiri kurang efektif untuk menangani urutan yang Panjang. Setelah itu untuk prosesnya, LSTM memiliki tiga gerbang utama yaitu Forget Gate, Input Gate, dan Output Gate, dan untuk GRU hanya memiliki dua gerbang utama saja yaitu, Reset Gate dan Update Gate.

3.3 Teknik Pengumpulan Data

Data yang dipakai dalam penelitian ini adalah data sekunder yang dimana data diambil langsung dari website investing (<https://www.investing.com/>).

Terdapat empat emitmen saham yang diambil untuk penelitian ini sebagai berikut.

1. PT Perusahaan Gas Negara Tbk (PGAS)
2. PT Medco Energi Internasional Tbk (MEDC)
3. PT AKR Corporindo Tbk (AKRA)
4. PT Rukun Raharja Tbk (RAJA)

Dari keempat emitmen yang ada diatas ini, isi dari data tersebut berupa, tanggal, harga pembukaan, harga terendah, dan harga penutupan setiap harinya dari tanggal 1 januari 2010 - 31 desember 2024. Pengambilan data dapat dilakukan di investing, namun, karena peneliti hanya mengambil data dari investing yang dimana data tersebut belum tersedia secara langsung dalam bentuk comma separated values (csv) sehingga data harus dikelola lebih lanjut.

3.4 Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah teknik analisis deskriptif yang dimana teknik ini adalah metode analisis data yang digunakan untuk memberikan gambaran-gambaran dengan jelas terhadap karakteristik data atau sampel yang diteliti oleh si peneliti. Teknik ini diawali dengan membuang semua informasi yang tidak diperlukan pada data yang digunakan hingga sesuai kebutuhan si peneliti.

Tabel 3. 3 Perbandingan Bahasa Pemograman

Indikator	Bahasa Pemograman	
	Python	R
Tujuan	Melakukan pengembangan web, analisis data, dan data science.	Melakukan visualisasi data dan analisis statistik.
Tipe Pemograman	<i>General Purpose</i>	<i>General Purpose</i>
Kelebihan	<ol style="list-style-type: none"> 1. Mudah untuk dimengerti. 2. Memiliki banyak library 3. Pemrosesan teks dan juga 	<ol style="list-style-type: none"> 1. Dibuat khusus untuk analisis statistik, 2. Memiliki banyak Tools visualisasi 3. Memiliki sifat yang fleksibel.
kekurangan	<ol style="list-style-type: none"> 1. kecepatan eksekusi menjadi lambat. 2. Pengolaan memori yang kurang efisien. 3. terbatasnya eksekusi kode python 	<ol style="list-style-type: none"> 1. Pemrosesan data yang lambat.

		2.Kurang cocok untuk pengembangan perangkat lunak. 3.Pengolahan data yang terbatas
--	--	---

Seperti tabel 3.3 yang dilihat diatas ini, terdapat tujuan, tipe pemograman, dan kelebihan masing-masing dari setiap bahasa pemograman yang dimana python adalah bahasa pemograman yang mudah dipahami untuk para pemula. Setelah itu penggunaan python juga memiliki banyak library dan juga framework sehingga dapat memenuhi kebutuhan pengguna, setelah itu juga dapat memproses data dengan baik. Dan juga R juga memiliki kelebihan yang dimana dibuat khusus untuk analisis statistik, tools yang beragam, dan bersifat fleksibel.

Selain adanya kelebihan dari kedua bahasa pemograman tersebut, terdapatnya juga kekurangannya yang dimana python bersifat interpretatif sehingga kecepatan eksekusi menjadi lambat, pengolaan memori yang kurang efisien, dan juga thread yang terbatas. Begitu juga dengan R yang dimana pemrosesan data yang lambat, kurang cocok digunakan untuk perangkat lunak, dan pengolahan data yang terbatas. Secara singkat, python dan juga R dapat memberikan keringanan dalam penggunaannya yang dimana penggunaan untuk memproses data. Dari kedua bahasa pemograman tersebut, memiliki kelebihan dan juga kekurangan masing-masing.

