

1. BAB I PENDAHULUAN

1.1 Latar Belakang

Prediksi harga saham adalah suatu tindakan untuk mencoba menentukan nilai masa depan saham suatu perusahaan atau instrumen keuangan lain yang diperdagangkan pada pasar saham. Keberhasilan memprediksi harga saham dapat memaksimalkan keuntungan investor (Hegazy, Soliman and Salam, 2014). Pasar saham memiliki karakteristik data dengan fluktuasi besar, dimensi tinggi, dan *non-linear* sehingga sulit diprediksi (Jia *et al.*, 2019). Perubahan harga saham dapat dipengaruhi oleh berbagai faktor seperti politik, ekonomi, pasar, teknologi, serta perilaku investor (Jin, Yang and Liu, 2020). Adapun model prediksi harga saham terbaru dengan akurasi yang lebih baik adalah kombinasi dari tiga elemen, yakni teknik pemrosesan sinyal, model *machine learning* dan model stokastik tradisional. Teknik pemrosesan sinyal berfungsi untuk mengontrol ketidakteraturan (*noise*) data dan model *machine learning* untuk mempelajari pola rangkaian waktu dengan kompleksitas tinggi, sedangkan model stokastik tradisional untuk mempelajari pola rangkaian waktu dengan kompleksitas rendah. Kombinasi ketiga elemen ini mampu membangun model prospektif sehubungan dengan data *backward* dan data *forward* (De Luca Avila and De Bona, 2020).

Flandrin, Torres and Colominas, (2011) memperkenalkan teknik pemrosesan sinyal menggunakan *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN). CEEMDAN mampu menganalisis dan memproses sinyal *non-linear* dan *non-stasioner* dengan waktu proses lebih cepat dibandingkan metode *Ensemble Empirical Mode Decomposition* (EEMD). Roondiwala, Patel and Varma, (2015) melakukan prediksi harga saham menggunakan model *Long Short Term Memory* (LSTM). Namun, model LSTM memiliki keterbatasan hanya menggunakan ketergantungan pada data *forward* dan tidak secara efisien mempertimbangkan ketergantungan pada data *backward* untuk memperoleh informasi yang berguna (Cui *et al.*, 2018).

Cao et al., (2019) mengusulkan model CEEMDAN-LSTM yang mampu menghasilkan prediksi harga saham dengan tingkat *error* rendah dibandingkan model LSTM, SVM, CEEMDAN-SVM, CEEMDAN-MLP, dan EMD-LSTM.

Xuan, Yu and Wu, (2020) mengusulkan model *Empirical Mode Decomposition - Long Short Term Memory - Cubic Spline Interpolation* (EMD-LSTM-CSI) yang memiliki waktu proses lebih cepat dan stabilitas yang kuat dalam memprediksi harga saham. EMD digunakan untuk menguraikan rangkaian harga saham menjadi stabil dan mengekstrak informasi fitur lokal dari seri aslinya. Model LSTM berfungsi sebagai *input* multi variabel dan model CSI untuk menghindari masalah *overfitting*. Model EMD-LSTM-CSI memberikan nilai prediksi lebih baik dibandingkan model *Support Vector Machine* (SVM), LSTM, *Attention based LSTM* (LSTM-ATTE), EMD-LSTM, dan EMD-LSTM-ATTE. Teja, Tiwari, dan Mohanty (2020) melakukan perbandingan metode pemrosesan sinyal EMD, EEMD dan CEEMDAN dengan hasil bahwa metode CEEMDAN memperoleh nilai statistik yang lebih tinggi dibandingkan model lainnya.

De Luca Avila & De Bona, (2020) mengusulkan model XCEEMDAN-LSTM-Spline yang mampu meningkatkan akurasi prediksi harga saham dari model CEEMDAN-LSTM. Namun, penelitian De Luca Avila & De Bona, (2020) masih belum mencoba melihat keterbatasan model LSTM seperti pada penelitian Cui et al., (2018). Untuk mengatasi masalah tersebut, Siami-Namini et al., (2019) dan Jia et al., (2019) mengusulkan model *Bidirectional LSTM* yang dapat memberikan prediksi lebih baik dibandingkan dengan model LSTM. *Bidirectional LSTM* dapat melintasi data *input* dua kali yaitu dari data *backward* ke data *forward* dan sebaliknya. Hasil penelitian menunjukkan bahwa penggunaan lapisan tambahan yang melintasi data *input* pada *Bidirectional LSTM* membantu dalam meningkatkan akurasi karena menangkap informasi tambahan berupa informasi yang berhubungan dengan data *backward* ke data *forward* dan sebaliknya.

Dari permasalahan dan kajian literatur yang dilakukan maka diusulkan sebuah model menggunakan XCEEMDAN-*Bidirectional LSTM-Spline*. Data

prediksi akan melalui normalisasi dan *preprocessing* XCEEMDAN untuk mengatasi karakteristik data *chaotic* dengan mengontrol *noise*. Setelah itu diterapkan lapisan tambahan *Bidirectional LSTM* untuk menangkap informasi tambahan serta model *Spline* untuk menghindari masalah *overfitting* yang mendukung hasil prediksi harga saham. Berdasarkan latar belakang tersebut, maka dilakukan penelitian dengan mengambil judul “**Prediksi Harga Saham Menggunakan XCEEMDAN-Bidirectional LSTM-Spline**”.

1.2 Masalah Penelitian

Berdasarkan uraian pada latar belakang di atas, maka masalah dalam penelitian ini dapat diuraikan menjadi dua bagian yaitu identifikasi masalah dan rumusan masalah.

1.2.1 Identifikasi Masalah

Identifikasi masalah pada latar belakang di atas adalah sebagai berikut:

1. Karakteristik data dengan fluktuasi besar, dimensi tinggi, dan *non-linear* menyebabkan sulitnya memprediksi harga saham.
2. Kombinasi teknik pemrosesan sinyal, model *machine learning*, dan model *spline* memungkinkan model memiliki kemampuan mengingat data *backward* dan data *forward* serta sebaliknya untuk menangkap informasi tambahan yang mendukung hasil prediksi harga saham.

1.2.2 Rumusan Masalah

Rumusan masalah dalam penelitian tesis ini adalah adanya keterbatasan informasi yang diperoleh apabila tidak menerapkan *Bidirectional LSTM* pada XCEEMDAN-LSTM-Spline. Pengontrolan *noise* yang tepat pada tahapan *preprocessing* XCEEMDAN dapat mempengaruhi hasil prediksi.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah menghasilkan model yang dapat memprediksi harga saham yang lebih akurat menggunakan XCEEMDAN-*Bidirectional LSTM-Spline*.

Manfaat dari penelitian ini adalah sebagai berikut.

1. Model dapat membantu memberikan informasi yang lebih akurat kepada investor dibandingkan model XCEEMDAN-LSTM-*Spline*.
2. Model yang diusulkan dapat menjadi pertimbangan untuk menambah sarana ilmu pengetahuan serta sebagai referensi tambahan dalam menghasilkan model prediksi harga saham.

1.4 Batasan Masalah

Adapun batasan dari penelitian ini adalah sebagai berikut:

1. Dataset dapat diakses secara umum pada link <https://finance.yahoo.com/>, dengan membuang celah hari yang disebabkan oleh akhir pekan dan hari libur. Rentang waktu yang digunakan dari tanggal 1 Januari 2010 hingga 31 Desember 2020, diantaranya sepuluh likuiditas teratas pasar saham Brazil sesuai penelitian De Luca Avila & De Bona, (2020) yaitu PETR4, VALE3, BOVA11, ITUB4, BBDC4, B3SA3, BBAS3, ABEV3, MGLU3, VVAR3.
2. Dataset mempertimbangkan jam perdagangan reguler dan penyesuaian acara perusahaan. Atribut *close* sebagai seri harga penutupan yang harus diprediksi dan atribut yang lain merupakan fitur eksogen.
3. Pemodelan menggunakan bantuan library De Luca Avila & De Bona, (2020) yang dapat diakses pada link <https://github.com/avilarenan/xlstmceemdan>.
4. *Threshold* IMF untuk membatasi sinyal frekuensi tinggi dengan sinyal frekuensi rendah antara lain IMF 2 dan IMF 3.
5. Batasan pemberian nilai *random* pada *noise seed* berada diantara 1 hingga 2^{32} .

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada tahapan ini dilakukan pengumpulan referensi yang berhubungan dengan cara kerja *machine learning*, proses prediksi pada beberapa penelitian sebelumnya, dan pengumpulan data saham untuk digunakan dalam penelitian.

2. Analisis Masalah

Pada tahapan ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan, masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah dan tantangan yang ada. Proses dalam *machine learning* akan digambarkan dalam bentuk *flowchart*.

3. Perancangan Model

Pada tahapan ini dilakukan perancangan model dimulai dari mengumpulkan *dataset* saham, menerapkan normalisasi dan *preprocessing* data menggunakan XCEEMDAN dan menerapkan *Bidirectional LSTM-Spline* untuk memprediksi harga saham.

4. Pengujian

Membandingkan metode XCEEMDAN-*Bidirectional LSTM-Spline* dengan XCEEMDAN-*LSTM-Spline* untuk melihat pengaruh *Bidirectional LSTM* dalam menghasilkan model yang terhindar dari *overfitting*. Pengujian ini dilihat dari perhitungan nilai RMSE.

5. Menarik kesimpulan dari hasil pengujian

6. Menyusun laporan Tesis

1.6 Sistematika Penulisan

Sistematika penulisan terdiri dari 5 bab, dimana secara garis besar masing-masing bab membahas hal – hal berikut. Bab 1 Pendahuluan, berisi penjelasan umum, masalah dan solusi yang sudah ada dan akan dilakukan. Bab 2 Kajian Literatur, berisi tinjauan pustaka, kerangka pikir pemecahan masalah dan hipotesis penelitian yang akan dilakukan. Bab 3 Metodologi Penelitian, berisi analisis masalah, model, data, metode, alat penelitian dan teknik analisis data. Bab

4 Hasil dan Pengujian, berisi hasil dan pengujian yang dilakukan pada model yang dibangun. Bab 5 Kesimpulan dan Saran, berisi kesimpulan yang diperoleh dari hasil dan pengujian penelitian dan saran untuk penelitian selanjutnya.



UNIVERSITAS MIKROSKIL