

# BAB 1

## PENDAHULUAN

### 1.1. Latar Belakang

Saham dapat didefinisikan sebagai tanda penyertaan modal seseorang atau pihak (badan usaha) dalam suatu perusahaan atau perseroan terbatas. Dengan menyertakan modal tersebut, maka pihak tersebut memiliki klaim atas pendapatan perusahaan, klaim atas aset perusahaan, dan berhak hadir dalam Rapat Umum Pemegang Saham (RUPS) (IDX, 2021). Saham dibeli dan dijual terutama di bursa saham, meskipun bisa juga ada penjualan pribadi, dan merupakan dasar dari banyak portofolio investor individu. Transaksi-transaksi ini harus dilakukan sesuai dengan berbagai peraturan pemerintah yang dibuat untuk melindungi investor dari praktik penipuan (Hayes, 2021).

Mengingat keuntungan besar yang dapat dihasilkan dari pasar saham, banyak upaya telah dilakukan untuk memprediksi pergerakan pasar. Menurut Hall dkk (1996) dalam Eapen dkk (2019), perkembangan di bidang pembelajaran mesin memungkinkan penggunaan komputer dalam tugas memprediksi pasar keuangan dengan mengotomatiskan metode analisis data statistik yang sudah ada.

Data yang seringkali tersedia dalam pasar saham adalah data *time series*. Menurut Frank dkk (2001), *time series* adalah sebuah urutan vektor-vektor,  $\mathbf{x}(t)$  dimana  $t = 0, 1, \dots$ , dimana  $t$  mewakili waktu yang telah berlalu. Dalam teori, nilai  $\mathbf{x}(t)$  dapat berupa nilai apapun yang berubah secara kontinu terhadap  $t$ , seperti suhu. Dalam praktek, untuk setiap sistem fisik tertentu, nilai  $\mathbf{x}(t)$  akan diambil sampelnya sedemikian rupa untuk memberikan serangkaian titik data yang diskrit, dengan jarak waktu yang sama.

Analisis *time series* seringkali digunakan untuk berbagai macam tujuan seperti: prediksi ekonomi, prediksi penjualan, analisis anggaran, analisis pasar saham, dan lain-lain (NIST/SEMATECH, 2022). Analisis *time series* memiliki tiga tujuan, yaitu: peramalan (*predicting*), pemodelan (*modelling*), dan karakterisasi (*characterization*). Tujuan *predicting* adalah untuk memprediksi secara akurat evolusi jangka pendek dari sistem; tujuan dari *modelling* adalah untuk mencari sebuah deskripsi yang secara akurat menangkap fitur-fitur dari perilaku jangka panjang sistem, dan tujuan dari *characterization* adalah untuk menentukan, dengan

sedikit informasi atau pengetahuan sebelumnya (*a priori*), sifat dasar sistem, seperti bilangan derajat kebebasan sistem atau jumlah keacakan yang terdapat didalam sistem (Weigend, 2018).

Akhir-akhir ini, metode pembelajaran mesin yang lumayan populer yang bernama *Neural Networks (NN)* atau *Artificial Neural Networks (ANN)* digunakan untuk analisa tren serta data *time series* musiman (Mahalakshmi dkk, 2016). Menurut Zhang dkk (1997), *ANN* adalah metode berbasis data yang beradaptasi dan belajar dari contoh-contoh serta menangkap hubungan fungsional yang halus di antara data, bahkan jika hubungan tersebut mendasar ini tidak diketahui atau sulit untuk dijelaskan. Hal ini dimungkinkan dengan penggunaan algoritma *backpropagation*, dimana bobot sistem diubah secara proporsional dengan turunan *error* sehubungan dengan bobot (Rumelhat dkk, 1994).

*ANN* juga memiliki kemampuan untuk menggeneralisir dan melihat melalui kebisingan dan distorsi, untuk meng-abstrakkan karakteristik penting dari data yang tidak relevan (Lippman, 1989). Setelah mempelajari data yang diberikan (sebuah sampel), *ANN* seringkali dapat dengan tepat memperlihatkan bagian populasi yang tidak terlihat bahkan jika data sampel mengandung informasi yang bising (*noisy*). (Zhang dkk, 1997). Salah satu subset dari *ANN*, *Deep Learning*, mampu mempelajari fitur-fitur yang terdapat didalam data tanpa ada arahan dari manusia. (Han dkk, 2019).

Dalam *Deep Learning*, arsitektur *Recurrent Neural Network (RNN)*, yang dikembangkan berdasarkan penelitian Rumelhart (1986), memiliki kemampuan untuk menangani data berurut dan mempelajari pola-pola yang muncul dari dalam data (Bhattacharjee & Tollner, 2016). Perbedaan *RNN* dengan arsitektur sebelumnya adalah penggunaan keluaran *layer* sebelumnya kedalam perhitungan keluaran *layer* untuk *time step* seterusnya dalam proses pelatihan jaringan.

Satu kekurangan dari penggunaan *RNN* adalah ada kecenderungan untuk informasi lama untuk hilang dan digantikan oleh informasi baru seiring proses pelatihan berjalan. *RNN* juga menghadapi masalah mempelajari pola baru, sehingga jaringan terjebak dalam minima lokal. Masalah ini dikenal dengan nama *vanishing/exploding gradient problem* (Bengio dkk, 1994; Bengio dkk 2003;

Graves, 2008), dan disebabkan oleh penggunaan algoritma *gradient descent* untuk proses pencarian bobot jaringan yang optimal.

Salah satu solusi untuk masalah ini adalah dengan penambahan sebuah “*constant error carousel*” kedalam *RNN* yang memungkinkan *error* dari *time step* sebelumnya untuk terjaga dan memastikan pengaruh data *input* masa lalu masih relevan untuk proses pelatihan jaringan di masa depan, serta sebuah gerbang multiplikatif yang menentukan apakah *error* dari *time step* tertentu relevan atau tidak untuk representasi internal jaringan. Pendekatan model arsitektur *RNN* ini dikenal dengan nama *Long Short Term Memory (LSTM)* dan pertama dikembangkan oleh Hochreiter dan Schmidhuber (1997).

*LSTM* telah dipakai di bidang pasar saham untuk menangani berbagai macam masalah. Kim & Won (2018) mengusulkan sebuah model *hybrid* yang mengintegrasikan *LSTM* dengan beberapa model tipe *GARCH* untuk melakukan peramalan volatilitas indeks harga saham. Zhao, dkk., (2017) mengusulkan sebuah model *LSTM* yang menggunakan pembobotan berdasarkan waktu (*time-weighted*) serta penggunaan pelabelan dengan definisi baru untuk memprediksi tren saham. Li, Shen & Zhu, (2018) mengusulkan sebuah model *LSTM* yang mampu mengesktrak informasi berharga dari faktor-faktor yang memiliki korelasi rendah serta menghilangkan *noise* dari faktor-faktor tersebut dengan penggunaan *input gate* tambahan yang dikendalikan oleh faktor-faktor meyakinkan yang diistilahkan sebagai *mainstream*.

Dalam sebagian besar penelitian yang disebutkan, fokusnya adalah pada dampak arsitektur jaringan terhadap performa model, dengan penggunaan ukuran dataset pelatihan yang statis. Belum ada penelitian yang dilakukan tentang pengaruh ukuran dataset pelatihan terhadap performa unit *LSTM*, terutama untuk masalah prediksi harga pasar saham. Penelitian ini bertujuan untuk memenuhi kesenjangan penelitian tersebut.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan maka peneliti akan mengangkat permasalahan sebagai berikut:

1. Adakah pengaruh perubahan ukuran dataset pelatihan terhadap performa *LSTM Network*?

2. Berapa nilai *Root Mean Square Error (RMSE)* yang dihasilkan dari perubahan ukuran dataset pelatihan terhadap hasil prediksi *LSTM Network*?

### 1.3 Batasan Masalah

Beberapa hal yang menjadi batasan dalam penelitian ini adalah sebagai berikut:

1. Data saham historis perusahaan berasal dari sektor finansial perbankan, khususnya Bank Rakyat Indonesia (BRI), Bank Central Asia (BCA), Bank Negara Indonesia (BNI), dan Bank Mandiri.
2. Data saham historis yang digunakan untuk Bank Rakyat Indonesia (BRI) dari tanggal 10 November 2003 sampai tanggal 26 November 2021.
3. Data saham historis yang digunakan untuk Bank Central Asia (BCA) dari tanggal 8 Juni 2004 sampai tanggal 26 November 2021.
4. Data saham historis yang digunakan untuk Bank Negara Indonesia (BNI) dari tanggal 23 Desember 2003 sampai tanggal 26 November 2021.
5. Data saham historis yang digunakan untuk Bank Mandiri dari tanggal 14 Juli 2003 sampai tanggal 26 November 2021.
6. Varian unit *LSTM* yang dipakai merupakan varian unit *LSTM* yang menggunakan *forget gate*.
7. *LSTM Network* menggunakan satu hidden layer dan satu output layer.
8. Aplikasi sistem dibuat dengan menggunakan NumPy, Pandas, Sklearn, Matplotlib, Tkinter, dan Pillow.

### 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan adalah:

1. Mengetahui pengaruh dari ukuran dataset pelatihan terhadap performa *LSTM Network*.
2. Mengetahui nilai performa yang dihasilkan dari perubahan ukuran dataset pelatihan terhadap hasil prediksi *LSTM Network*.

### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sistem yang dibangun dapat membantu pengguna umum khususnya mahasiswa rekayasa sistem komputer untuk lebih memahami cara kerja *LSTM Network* dan pemrosesan data *timeseries*.

Manfaat teoritis yang didapatkan dari penelitian ini adalah penambahan pengetahuan mengenai pemanfaatan model *LSTM* dalam dunia perbankan. Manfaat praktis yang didapatkan dari penelitian ini adalah terbukanya peluang bagi para *trader* di pasar saham untuk memprediksi harga saham masa depan dengan menggunakan data yang lebih sedikit tanpa mengurangi keakurannya.

## 1.6 Sistematika Penulisan

Sistematika penulisan di penelitian ini terdiri dari 5 bab, yang disusun sebagai berikut:

- BAB 1 Menyajikan latar belakang penelitian, permasalahan masalah, tujuan penelitian, dan manfaat penelitian dalam menentukan dampak ukuran dataset pelatihan terhadap performa *LSTM Network*.
- BAB 2 Menyajikan tinjauan pustaka yang berkaitan dengan penelitian yang dilakukan. Beberapa teori-teori yang terkait pada penelitian ini berupa kajian penelitian terdahulu yang terkait, teori dasar jaringan saraf tiruan, dan algoritma *Long Short Term Memory (LSTM)*.
- BAB 3 Berisikan metodologi penelitian yang digunakan dalam penelitian ini meliputi studi literatur, metode pengumpulan data, analisis kebutuhan perangkat keras dan perangkat lunak, perancangan sistem, implementasi, dan pengujian.
- BAB 4 Berisikan penjelasan mengenai gambaran umum rancangan perangkat lunak yang dibuat. Rancangan terdiri dari perancangan lunak, perancangan antar muka dan perancangan pengujian *black box*. Dalam perancangan perangkat lunak menjelaskan mengenai diagram alir sistem dan *Data Flow Diagram (DFD)* sistem.
- BAB 5 Berisikan hasil dari implementasi rancangan yang telah dibuat, kode program, hasil pengujian sistem menggunakan *black-box* serta hasil pembahasan dari sistem yang telah dibuat.
- BAB 6 Merupakan bagian akhir yang berisikan kesimpulan dari penelitian yang telah dilakukan disertai saran/rekomendasi untuk perbaikan, pengembangan atau kesempurnaan penelitian yang telah dilakukan.