

Perbandingan Prediksi Harga Saham Dengan Menggunakan LSTM GRU Dengan Transformer

Idham¹, Muhammad Ghudafa Taufik Akbar², Supriadi Panggabean³, Mohamad Noor⁴

Email : 14002600@nusamandiri.ac.id¹, 14002455@nusamandiri.ac.id², 14002471@nusamandiri.ac.id³,

14002440@nusamandiri.ac.id⁴

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri Jakarta^{1,2,3,4}

Abstrak

Saham adalah sebuah bukti kepemilikan nilai sebuah perusahaan, artinya pemilik saham adalah pemilik perusahaan. Semakin besar saham yang dimiliki, maka semakin besar kekuasaannya di perusahaan tersebut. Faktor yang terjadi sekarang dalam sektor pasar saham yaitu adanya dampak dari virus corona terhadap indeks harga saham dan arus dana asing ke pasar saham. Maka sangat perlu untuk dilakukan prediksi sentiment analysis pandemi corona terhadap sektor pasar saham untuk melihat bagaimana perbandingan pergerakan IHSG di Indonesia sebelum terjadi pandemi dan pada saat terjadi pandemi Covid-19. Metode yang digunakan untuk prediksi analysis sentimen dengan index harga saham Indonesia ini menggunakan transformers dengan fitur bag of word, TF-IDF dan word embedding. Dari hasil prediksi sebelum menggunakan metode transformers pada LSTM, dan GRU didapatkan rata-rata pada LSTM Performance akurasi 0,394 dan GRU 0,216[1]. Algoritma yang digunakan dalam model ini adalah Long short-term memory (LSTM), dan Gated Recurrent Unit (GRU), sedangkan untuk mendapatkan hasil word embedding menggunakan Vector space model. Terdapat 1989 baris data dan 27 atribut, sedangkan untuk akurasi yang dihasilkan setelah melakukan iterasi beberapa kali mendapatkan hasil yang signifikan, performance yang dihasilkan adalah semakin mendekati akurasi yang cukup tinggi. Berdasarkan hasil eksperimen perbandingan performance akurasi antara LSTM dan GRU terhadap penggunaan Transformers, maka terlihat lebih baik performance akurasinya setelah menggunakan transformers pada ketiga model tersebut.

Kata kunci : *Transformer, GRU, LSTM, TF-IDF, word embedding, bag of word*

1. Pendahuluan

Pasar modal/ saham yang ada di Indonesia merupakan pasar yang sedang berkembang, Sangat rentan terhadap ekonomi global dan pasar modal dunia. Dalam pasar modal terdapat istilah investasi. Investasi adalah suatu cara/ proses penundaan konsumsi dari saat ini untuk masa depan yang akan datang yang mana terdapat risiko ketidakpastian. (Dwi & Pengembangan, 2015) Saham merupakan surat berharga/ instrument pasar keuangan yang menjadi bukti kepemilikan seseorang atas suatu perusahaan (Ramadani, 2018). Semakin besar saham yang dimiliki, maka semakin besar kekuasaannya di perusahaan tersebut. Faktor yang terjadi sekarang dalam sektor pasar saham yaitu adanya dampak dari virus corona terhadap indeks harga saham dan arus dana asing ke pasar saham.

Untuk Metode Deep Learning yang digunakan menggunakan model Transformers dengan algoritma GRU, dan LSTM. sedangkan untuk mendapatkan hasil word embedding menggunakan Vector space model.

Analisa sentimen sering juga dikenal dengan opinion mining yaitu studi komputasi dari pendapat, sentimen, sikap, serta emosi yang disajikan dalam sebuah teks (Liu, 2012).

Deep learning adalah salah satu cabang dari ilmu machine learning yang terdiri algoritme pemodelan abstraksi tingkat tinggi pada data menggunakan sekumpulan fungsi transformasi non-linear yang ditata berlapis-lapis dan mendalam (Deng, 2014).

GRU adalah arsitektur yang diciptakan oleh Kyunghun Cho pada Tahun 2014 (Cho et al., 2014). Serupa dengan LSTM, GRU juga menggunakan sistem gerbang, Arsitektur GRU lebih sederhana daripada LSTM. GRU tidak menggunakan cell state, tetapi memanfaatkan hidden state untuk menyimpan informasi. Reset gate dalam GRU menentukan informasi baru harus dilupakan atau tidak, sedangkan update gate untuk mengingat.

LSTM adalah metode yang lebih baik dibandingkan dengan konvensional, metode LSTM sangat cocok untuk diterapkan pada sentiment analisis (Jelodar et al., 2020).

2. Metode Penelitian

Dalam penelitian ini melakukan prediksi harga saham dengan menggunakan metode transformer. Transformer adalah model transduksi pertama yang mengandalkan seluruhnya pada self-attention sendiri untuk mengetahui hasil representasi masukan maupun keluarannya. Pada penerapan ini, fitur yang digunakan adalah Bag of Words, Tf-IDF dan Word embedding.

Bag of Words (BOW) adalah algoritma yang menghitung berapa kali suatu kata muncul dalam suatu dokumen. Jumlah kata dalam dokumen tersebut memungkinkan untuk membandingkan dokumen dan mengukur kesamaannya untuk aplikasi seperti pencarian, klasifikasi dokumen, dan pemodelan topik.

Tf-Idf merupakan singkatan dari Term frequency-inverse document frequency. Jadi, dua hal: frekuensi istilah dan frekuensi dokumen terbalik. Term frequency (TF) pada dasarnya adalah keluaran dari model BoW. Untuk dokumen tertentu, ini menentukan seberapa penting sebuah kata dengan melihat seberapa sering kata itu muncul di dokumen. Frekuensi istilah mengukur pentingnya kata. Jika sebuah kata muncul berkali-kali, maka kata itu pasti penting. Misalnya, jika dokumen kita adalah “Saya seorang pecinta kucing. Saya punya kucing bernama Steve. Saya memberi makan kucing di luar kamar saya secara teratur”, kita melihat bahwa kata-kata dengan frekuensi tertinggi adalah saya, a, dan kucing. Ini sesuai dengan intuisi kami bahwa frekuensi jangka tinggi = kepentingan yang lebih tinggi.

$TF(t) = (\text{Jumlah istilah } t \text{ muncul dalam dokumen}) / (\text{Jumlah total istilah dalam dokumen})$

IDF digunakan untuk menghitung bobot kata-kata langka di semua dokumen. Kata-kata yang jarang muncul di corpus memiliki skor IDF yang tinggi. Namun, diketahui bahwa istilah-istilah tertentu, seperti “aku”, “a” mungkin sering muncul tetapi tidak begitu penting. Jadi kita perlu mempertimbangkan istilah yang sering digunakan sambil meningkatkan yang langka

$IDF(t) = \log_e(\text{Jumlah total dokumen} / \text{Jumlah dokumen dengan term } t \text{ di dalamnya})$.

Word embeddings adalah proses konversi kata yang berupa karakter

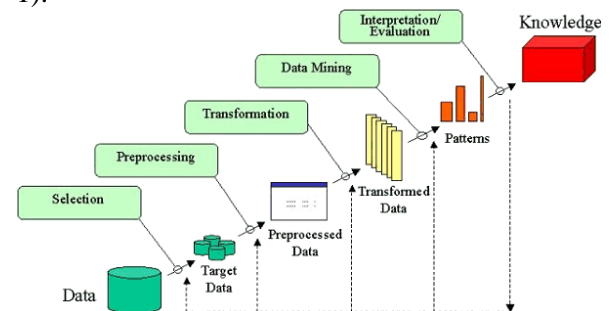
alphanumeric kedalam bentuk vector. Prediksi yang dilakukan diimplementasikan dengan menggunakan Word Embedding atau representasi numerik dari teks sehingga komputer dapat menanganinya.

Setelah menerapkan transformer dengan fitur Bag of Words, Tf-IDF dan Word embedding, kemudian dibandingkan dengan penerapan LSTM dan GRU.

Di dalam GRU, komponen pengatur alur informasi tersebut disebut sebagai gate dan GRU mempunyai 2 gate, yaitu reset gate dan update gate. Bila kita ingin membuat keputusan untuk makan seperti analogi diatas, reset gate pada GRU akan menentukan bagaimana untuk menggabungkan input baru dengan informasi masa lalu, dan update gate, akan menentukan berapa banyak informasi masa lalu yang harus tetap disimpan.

Dalam penelitian ini penulis melakukan eksperimen terhadap data public Jakarta Composite Index (JKSE) dan data Combine News DJIA.

Penelitian ini menggunakan metode Knowledge Discovery Database (KDD), ada 5 langkah yang harus dilakukan dalam metode ini, yaitu: Seleksi Data, Pra Pemrosesan, Transformasi, Penambangan Data, dan Evaluasi Interpretasi (lihat Gambar 1).



Gambar 1. Metode KDD

3. Hasil dan Pembahasan

Dalam penelitian ini melakukan prediksi harga saham dengan menggunakan metode transformer. Transformer adalah model transduksi pertama yang mengandalkan seluruhnya pada self-attention sendiri untuk mengetahui hasil representasi masukan maupun keluarannya. Pada penerapan ini, fitur yang digunakan adalah Bag of Words, Tf-IDF dan Word embedding.

Classification Score:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	737
1	1.00	1.00	1.00	873
accuracy			1.00	1610
macro avg	1.00	1.00	1.00	1610
weighted avg	1.00	1.00	1.00	1610

Confusion Matrix:
[[737 0]
[0 873]]
Accuracy Score: 0.47

Classification Score:

	precision	recall	f1-score	support
0	0.46	0.45	0.45	186
1	0.48	0.50	0.49	192
accuracy			0.47	378
macro avg	0.47	0.47	0.47	378
weighted avg	0.47	0.47	0.47	378

Confusion Matrix:
[[83 103]
[96 96]]

Gambar 2. Performance Deep model LSTM

Accuracy Score: 1.0

Classification Score:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	737
1	1.00	1.00	1.00	873
accuracy			1.00	1610
macro avg	1.00	1.00	1.00	1610
weighted avg	1.00	1.00	1.00	1610

Confusion Matrix:
[[737 0]
[0 873]]
Accuracy Score: 0.49

Classification Score:

	precision	recall	f1-score	support
0	0.48	0.44	0.46	186
1	0.50	0.54	0.52	192
accuracy			0.49	378
macro avg	0.49	0.49	0.49	378
weighted avg	0.49	0.49	0.49	378

Confusion Matrix:
[[82 104]
[89 103]]

Gambar 3. Performance Deep model GRU

Dari hasil eksperimen di dapatkan hasil dapat di lihat dari table berikut:

Tabel 1. Performance Akurasi

No	Model	Precision	Recall	F1-Score
1	LSTM	0,27	0,50	0,35
	Sequential1	0,29	0,54	0,38
2	LSTM	0,25	0,50	0,34
	Sequential2	0,26	0,51	0,34
3	GRU	0,27	0,50	0,35
	Sequential3	0,29	0,54	0,38

4	GRU	0,25	0,50	0,34
	Sequential4	0,26	0,51	0,34
5	LSTM		1,00	1,00
	(Drop Out)	1,00	1,00	1,00
	Bidirect model	1,00		
6	LSTM(Drop Out)		0,47	0,47
	Bidirect model	0,47		
7	GRU(Drop Out)		1,00	1,00
	Bidirect model	1,00		
8	GRU(Drop Out)		0,49	0,49
	Bidirect model	0,49		

Tabel 2. Performance Akurasi Transformers

No	Model	Loss	Accuracy	Val-Loss	Val-Accuracy
1	Transformers Epoch 1	0,39	0,82	0,34	0,86
2	Transformers Epoch 2	0,20	0,93	0,32	0,87

Akurasi yang dihasilkan setelah melakukan iterasi beberapa kali mendapatkan hasil yang signifikan. Terlihat performance dari data diatas semakin mendekati akurasi yang cukup tinggi.

Gerbang di LSTM adalah fungsi aktivasi sigmoid yaitu mereka menghasilkan nilai antara 0 atau 1 dan dalam sebagian besar kasus itu adalah 0 atau 1. kami menggunakan fungsi sigmoid untuk gerbang karena, kami ingin gerbang hanya memberikan nilai positif dan harus dapat memberi kami jawaban yang jelas apakah, kami perlu mempertahankan fitur tertentu atau kami perlu membuang fitur itu. Persamaan untuk gerbang LSTM adalah :

$$It = \sigma(Wi[ht - 1, xt] + bi) \quad (1)$$

$$ft = \sigma(Wf[ht - 1, xt] + bf) \quad (1)$$

Ada beberapa variasi pada unit gated penuh, dengan gating dilakukan menggunakan keadaan tersembunyi sebelumnya dan bias dalam berbagai kombinasi, dan bentuk yang disederhanakan disebut unit gated minimal. [8]

Awalnya, untuk t=0, vector keluarannya adalah h₀=0:

$$Zt = \sigma g(WzXt + UzHt - 1 + bz) \quad (1)$$

$$rt = \sigma g(WzXt + UrHt - 1 + br) \quad (2)$$

$$Ht = \phi g(WhXt + Uh(r \odot ht - 1) + bh) \quad (3)$$

$$Ht = (1 - zt) \phi ht - 1 + zt \odot ht \quad (4)$$

4. Kesimpulan

Hasil penelitian dengan menggunakan model algoritma Recurrent neural networks (RNN,) Long short-term memory (LSTM), dan Gated Recurrent Unit (GRU), sedangkan untuk mendapatkan hasil word embedding menggunakan Vector space model. Terdapat 1989 baris data dan 27 atribut, sedangkan untuk akurasi yang dihasilkan setelah melakukan iterasi beberapa kali mendapatkan hasil yang signifikan, performance yang dihasilkan adalah semakin mendekati akurasi yang cukup tinggi.

Berdasarkan hasil eksperimen perbandingan performance akurasi antara LSTM dan GRU terhadap penggunaan Transformers, maka terlihat lebih baik performance akurasinya setelah menggunakan transformers pada ketiga model tersebut.

Kesimpulan menjawab masalah penelitian tidak melampaui kapasitas temuan. Kesimpulan berbentuk narasi, logis, dan tepat guna. Kesimpulan tidak berupa poin-poin. Ukuran 11pt font Times New Roman.

5. Daftar Pustaka

- [1] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.
- [2] Deng, L. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, 7(3–4), 197–387. <https://doi.org/10.1561/20000000039>.
- [3] <https://www.akseleran.co.id>. (2021, October 22). Mengenal Evergrande, Raksasa Properti China yang Sedang Krisis. <https://www.akseleran.co.id/Blog/Mengenal-Evergrande-Raksasa-Properti-China-Yang-Sedang-Krisis/>.
- [4] Jelodar, H., Wang, Y., Orji, R., & Huang, S. (2020). Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *IEEE Journal of Biomedical and Health Informatics*, 24(10), 2733–2742. <https://doi.org/10.1109/JBHI.2020.3001216>.
- [5] Lien Minh, D., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. *IEEE Access*, 6, 55392–55404. <https://doi.org/10.1109/ACCESS.2018.2868970>.
- [6] Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>.
- [7] Liu, P., Qiu, X., & Huang, X. (2016). Recurrent Neural Network for Text Classification with Multi-Task Learning.
- [8] Ramadani, F. (2018). PENGARUH INFLASI, SUKU BUNGA DAN NILAI TUKAR RUPIAH TERHADAP HARGA SSAHAM PERUSAHAAN SEKTOR PROPERTI DAN REAL ESTATE YANG TERCATAT DI BURSA EFEK INDONESIA. *Manajemen Bisnis*, 6(1). <https://doi.org/10.22219/jmb.v6i1.5392>.