123123

by OHX B14-81

**Submission date:** 28-Jun-2022 08:41PM (UTC-0700)

**Submission ID:** 1862777294

**File name:** TB2\_DL\_41519310029.docx (887.47K)

Word count: 2396

**Character count:** 14564

# Evaluasi Prediksi Harga Saham Menggunakan Metode Deep Learning dan Machine Learning

Harpan Budi Santoso Universitas Mercu Buana Jakarta, Indonesia panoet.h@gmail.com 12 Ilham Hanif Universitas Mercu Buana Jakarta, Indonesia ilham.hanif@gmail.com

Abstrak — Analisis prediksi harga saham telah banyak dilakukan saat ini dengan menggunakan metode machine learning maupun deep learning. Penelitian ini akan melakukan evaluasi terhadap 2 metode machine learning dan 2 metode deep learning dengan membandingkan metrics Mean Squared Error, Root Mean Squared Error, dan Mean Absolute Error. Dataset menggunakan harga saham GOOGLE dalam kurun waktu tahun 2012 sampai dengan 2016. Penelitian dilakukan dengan melakukan uji coba terhadap keempat algoritma tersebut pada environment yang berbeda-beda untuk mendapatkan hasil yang optimal. Pada akhirnya laporan ini diharapkan dapat menjadi acuan untuk penelitian-penelitian yang akan datang.

Kata kunci—prediksi, saham, long short term memory, recurrent neural network, k-nearest neighbour, support vector regression

### I. PENDAHULUAN

Saham adalah tanda partisipasi atau kepemilikan investor individual maupun institusional atau trader atas investasi mereka atau sejumlah dana yang ditanamkan dalam sebuah perusahaan. [1] Adapun tujuan dari investor pada dasarnya adalah mendapatkan keuntungan baik dari dividen maupun capital gain. Investasi saham pada umumnya diperdagangkan melaui pasar modal. Pasar modal adalah suatu tempat untuk memfasilitasi kegiatan jual beli khususnya untuk surat berharga. [2]

Untuk mendapatkan keuntungan dari dividen, investor cukup membeli saham sebelum *cum date* lalu menunggu waktu pembagian dividen. Sedangkan *capital gain* diperoleh dari selisih harga beli dan harga jual atas transaksi saham yang telah dilakukan oleh investor.

Untuk mendapatkan keuntungan dari *capital gain* tersebut, seorang investor harus dapat melakukan prediksi terhadap pergerakan harga saham yang akan ditransasikan olehnya. Telah diketahui secara umum terdapat 2 metode analisis pergerakan harga saham, yaitu teknikal dan fundamental. [3]

Analisis teknikal digunakan untuk mengidentifikasi pola dan tren harga saham lalu bereksperimen dengan pola tersebut untuk menentukan pola berikutnya. [4] Sedangkan analisis fundamental berkutat pada aspek ekonomi, industri, kondisi perusahaan, bahkan mungkin politik secara umum. Hal ini mungkin terjadi karena harga saham merepresentasikan kekuatan permintaan dibandingkan dengan penawaran yang secara langsung juga berkorelasi dengan kondisi ekonomi, politik, dan sosial secara umum. [5]

Untuk mendapatkan analisis teknikal yang lebih baik lagi perlu ada bantuan prediksi data saham berdasarkan *time series*. Hal ini dapat meyakinkan para investor ataupun trader dalam melakukan transaksi. [6]

Untuk membantu mengidentifikasi pola dan memprediksi pergerakan harga saham, pendekatan Machine Learning dan Deep Learning banyak diteliti belakangan ini, salah satunya adalah yang menggunakan pendekatan Deep Learning. Sebelumnya penelitian menggunakan Support Vector Machine untuk memprediksi pergerakan Harga Saham.

Penelitian ini bertujuan untuk mengetahui algoritma yang terbaik dalam melakukan analisis time series terhadap harga saham GOOGLE. Dalam penelitian ini menggunakan dua pendekatan yaitu Machine Learning dan Deep Learning dengan masing-masing 2 algoritma yaitu K-Nearest Neighbour (KNN) dan Support Vector Regression (SVR) pada Machine Learning dan Long Short Term Memory (LSTM) dan Recurrent Neural Network (RNN) pada Deep Learning, dengan terlebih dahulu data training dilakukan beberapa proses awal yaitu cleansing, transformasi, smoothing dan differencing. Dengan harapan dari kombinasi data preprocessing tersebut dapat mendapatkan hasil evaluasi yang baik. [7]

Data yang digunakan adalah data dari pergerakan saham GOOGLE dari 1 Januari 2012 sampai dengan 30 Desember 2016, dengan banyaknya data sebanyak 1258 data.

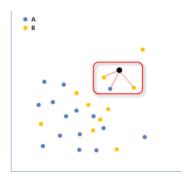
## II. TEORI

### A. Pasar Modal

Pasar modal adalah pasar yang digunakan untuk berbagai instrumen keuangan jangka panjang yang dapat diperdagangkan. [8] Instrumen-instrumen tersebut antara lain adalah saham, oblogasi, reksa dana, instrumen derivatif maupun instrumen lainnya. Selain sebagai tempat berinvestasi bagi investor, pasar modal merupakan tempat industri atau pemerintahan dalam hal ini pelaku usaha mencari modal. [9]

## B. K-Nearest Neighbor

KNN merupakan salah satu algoritma yang terbukti cukup baik digunakan untuk menyelesaikan permasalahan klasifikasi. [10] Secara umum cara kerja KNN adalah dengan (1) menentukan data tetangga dari kelas yang akan digunakan untuk pertimbangan penentuan kelas, (2) menghitung jarak antar masing-masing data dan (3) mengambil sejumlah data dengan jarak kelas lalu kemudian menentukan hasil dari data tersebut menjadi kelas. [11]

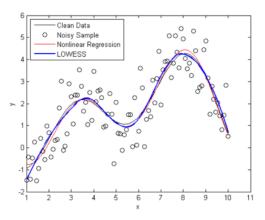


Gambar 1 - Ilustrasi KNN

Algoritma KNN memiliki beberapa keunggulan yaitu memiliki kekuatan terhadap training data yang mengandung banyak noise dan efektif jika training data-nya besar. Sedangkan, kelemahan dari KNN adalah KNN harus menentukan nilai dari data berdasarkan parameter K, trainingnya tidak berjalan dengan baik karena jarak ke data tidak jelas tentang jarak yang digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil yang terbaik, dan biaya penggunaan yang relatif tinggi karena jarak dari tiap individu dalam data training perlu diperhitungkan. [12]

## C. Support Vector Regression

SVR merupakan salah satu pengembangan dari Support Vector Machine (SVM) untuk keperluan regresi. [13] Pada SVR, garis lurus yang digunakan untuk mencocokan data disebut dengan hyperplane. Konsep dasar dari SVR adalah menemukan garis batas yang paling baik yang berarti adalah hyperplane yang mempunyai jumlah titik terbanyak. [14]

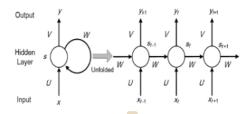


Gambar 2 - Ilustrasi Hyperplane pada plot data

# D. Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) merupakan salah satu algoritma dari jaringan syaraf tiruan yang sudah dirancang khusus untuk memproses data secara terus menerus. Algoritma RNN akan memproses data historis yang sudah terjadi di masa lampau dan evaluasi untuk pembelajarannya. Algoritma ini memproses input secara sekuensial, sampel per sampel. Pada setiap proses, cara yang dilakukan RNN untuk

menyimpan pembelajaran adalah dengan melakukan perulangan di dalam arsitektur algoritmanya, yang otomatis akan membuat sebuah model dari data masa lalu. [15]



Gambar 3 - Arsitektur Recurrent Neural Network

RNN merupakan salah satu metode yang dirancang khusus untuk menyelesaikan tugas yang berkaitan dengan data time series, misalnya pada data saham. RNN dapat melakukan prediksi jika terdapat data input pada waktu sebelumnya dan juga input data sebelumnya. Sehingga RNN memiliki memori yang berisikan data yang pembelajaran yang dihasilkan oleh data sebelumnya. [16]

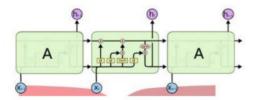
Berdasarkan arsitekturnya, RNN dapat dikatakan bisa menangani data yang jangka panjang. Tetapi dalam proses trainingnya, RNN tidak mampu dalam menangani pembelajaran jangka panjang dengan *gradient descent*, hal ini terjadi dikarenakan RNN tidak dapat menyimpan informasi sebelumnya dengan baik yang disebabkan oleh masalah gradien yang menghilang. [17]

#### E. Long Short Term Memory (LSTM)

LSTM merupakan jenis dari Recurrent Neural Network. RNN adalah jenis jaringan saraf tiruan yang kuat yang dapat mempertahankan memori input. LSTM sebagai salah satu pengembangan dari arsitektur RNN, bekerja dengan menambahkan sebuah proses dalam setiap cell yang menyeleksi informasi mana yang layak diteruskan, mana yang seharusnya dihapus. Selain itu, LSTM juga menjadi solusi atas masalah vanishing gradient yang terjadi pada RNN. [18]

LSTM secara umum terdiri dari beberapa arsitektur, yaitu memory cell, input gate, output gate, dan forget gate. LSTM cell akan mengambil data input dan menyimpannya sementara, input gate mengontrol nilai baru apa yang harus dilanjutkan kedalam cell, forget gate mengontrol berapa banyak nilai yang tersisa di cell, dan output gate mengontrol berapa banyak nilai didalam cell akan terus digunakan untuk menghitung aktivasi output dari unit LSTM. [19]

Pada tahun 1997 Hochreiter dan Schmidhuber memperkenalkan pertama kalinya Algoritma LSTM. Arsitektur Long Short Term Memory (LSTM) adalah evolusi atau perubahan dari Recurrent Neural Network (RNN). Arsitektur LSTM ini terus dikembangkan oleh banyak peneliti dibidang forecasting dan speech recognition. [20]

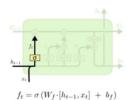


Gambar 4 - arsitektur LSTM

Melalui gambar di atas proses gerbang dibagi menjadi 4 bagian memory cell,input gate, forget gate, dan output gate.

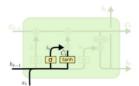
Berdasarkan prosesnya LSTM terdapat fungsi-fungsi yang digunakan pada setiap proses gerbang, yaitu:

 Forget gate adalah Gerbang yang digunakan untuk memutuskan apakah input dan output harus diteruskan ke cell state atau tidak.



Gambar 5 - forget gate

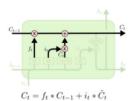
 Input Gate adalah Gerbang input yang memiliki dua fungsi aktivasi (sigmoid dan tanh), untuk menentukan bagian mana yang akan diperbaharui datanya.



$$i_t = \sigma \left( W_i \cdot [h_{t-1}, x_t] + b_i \right)$$
  
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

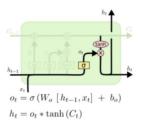
Gambar 6 - input gate

 Cell State adalah Gerbang yang digunakan untuk mengupdate input (nilai lama) dengan nilai input yang baru.



Gambar 7 - cell state

 Output Gate adalah Gerbang output dimana terdapat kesepakatan anatara nilai lama dan nilai baru



Gambar 8 - output gate

#### III. PEMBAHASAN

#### A. Metode Evaluasi

Evaluasi dilakukan dengan menggunakan *metrics Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Error (MAE)*. Semakin rendah nilai pada ketiga *metrics* tersebut menunjukkan performa model yang lebih baik.

## B. Dataset

Dataset menggunakan data open source historis harga saham Google dari tahun 2012 s.d. 2016 diambil dari laman Yahoo Finance (https://finance.yahoo.com/quote/GOOG/history?p=GOOG).

Gambar 9 - dataset harga saham google



Gambar 10 - plot dataset

## C. Penyiapan Dataset

Atribut yang digunakan untuk keperluan penelitian ini adalah atribut OPEN. Dataset *training* diambil dari 50 data terakhir dataset. Selanjutnya terhadap dataset tersebut dilakukan *scaling* menggunakan *MinMaxScaler*.

```
scaler = MinMaxScaler(feature_range= (0,1))
train_scaled = scaler.fit_transform(train)
```

Gambar 12 - scaling datasei

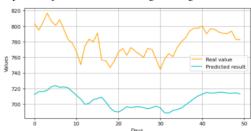
Setelah dilakukan *preprocessing* tersebut, dihasilkan data *training* sebanyak 1158 baris dan data *testing* 50 baris.

### D. Implementasi dan Evaluasi

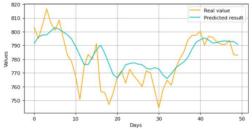
Terhadap dataset dilakukan pemodelan menggunakan 4 algoritma yaitu RNN, LSTM, SVR, dan KNN.

## 1. Recurrent Neural Network (RNN)

Implementasi RNN terhadap dataset dilakukan menggunakan beberapa konfigurasi untuk mencari model yang terbaik. Pemilihan *activation* yang digunakan sangat mempengaruhi performa model, dimana RELU tidak dapat bekerja lebih baik dibiandingkan dengan TANH.



Gambar 13 - RNN menggunakan activation RELU



Gambar 14 - RNN menggunakan activation TANH

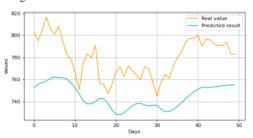
Empat buah layer dengan activation TANH, optimizer ADAM, epoch 150 dan batch 50 memberikan performa model terbaik dalam penelitian ini.



Gambar 15 - original, train, dan test predicted RNN

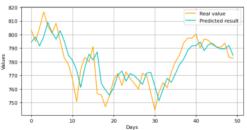
## 2. Long Short Term Memory (LSTM)

Serupa pada RNN, activation RELU juga tidak memberikan performa model lebih baik dibandingkan dengan TANH.



Gambar 16 - LSTM menggunakan activation RELU

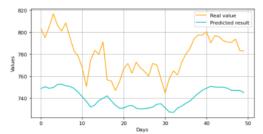
Pada LSTM, pemilihan *epoch* dan *batch* juga sangat berpengaruh terhadap performa model. LSTM dengan 4 layer, *activation* TANH, *optimizer* ADAM, *epoch* 100, *batch* 32 menghasilkan performa model terbaik dalam penelitian ini.



Gambar 17 - TANH, ADAM, EPOCH 100, dan BATCH 32

# 3. Support Vector Regression (SVR)

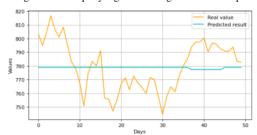
SVR menghasilkan model yang tidak lebih baik dibandingkan metode deep learning RNN dan LSTM.



Gambar 18 - prediksi menggunakan SVR

## 4. KNN (K-Nearest Neighbour)

KNN menghasilkan model yang tidak berhasil menghasilkan output yang sesuai dengan dinamika input.



Gambar 19 - prediksi menggunakan KNN

## E. EVALUASI

Metrics evaluasi dari metode-metode implementasi di atas adalah sebagai berikut:

	RMSE Train	RMSE Test	Test / Train
RNN 1	15.1688095	23.34869239	54%
RNN 2	36.62473703	75.14149081	105%
RNN 3	15.1025773	12.89372579	-15%
RNN 4	16.99174685	15.32312294	-10%
LSTM 1	23.77646076	19.31726138	-19%
LSTM 2	19.9430728	36.91553414	85%
LSTM 3	13.41989722	15.81871125	18%
LSTM 4	10.4924742	9.964675909	-5%
SVR	29.8484927	41.15996651	38%
KNN	13.97757655	18.10889904	30%

Tabel 1 - metrics RMSE

LSTM (4 layer, TANH, ADAM, EPOCH100, dan BATCH32) menghasilkan *metric* RMSE *loss* 10.49 dan *validation loss* 9.96, terbaik dibandingkan dengan metode lainnya.

	MSE Train	MSE Test	Test / Train
RNN 1	230.0927818	545.1614362	137%
RNN 2	1341.371363	5646.243642	321%
RNN 3	228.087841	166.2481647	-27%
RNN 4	288.7194612	234.7980966	-19%
LSTM 1	565.3200863	373.1565872	-34%
LSTM 2	397.7261529	1362.756661	243%
LSTM 3	180.0936414	250.2316255	39%

LSTM 4	110.0920149	99.29476597	-10%
SVR	890.9325167	1694.142843	90%
KNN	195.3726462	327.9322245	68%

Tabel 2 - metrics MSE

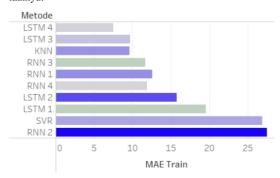
LSTM (4 layer, TANH, ADAM, EPOCH100, dan BATCH32) juga menghasilkan *metrics* MSE yang baik dengan *loss* 110.09 dan *validatin loss* 99.29. Hasil ini juga merupakan yang terbaik dibandingkan dengan yang lainnya.

	MAE Train	MAE Test	Test / Train
RNN 1	12.57008183	20.96826733	67%
RNN 2	27.47860361	74.15307153	170%
RNN 3	11.67144139	9.428771191	-19%
RNN 4	11.85448605	12.40207642	5%
LSTM 1	19.50564789	15.85227056	-19%
LSTM 2	15.73940325	34.66259912	120%
LSTM 3	9.664837741	11.74802993	22%
LSTM 4	7.513649118	8.054801807	7%
SVR	26.85965333	39.31915638	46%
KNN	9.623985147	15.72652	63%

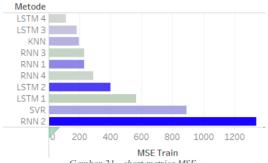
Tabel 3 - metrics MAE

## IV. SIMPULAN

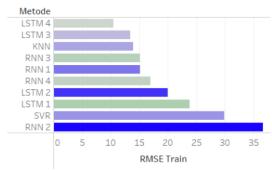
Metode deep learning LSTM (4 layer, TANH, ADAM, RMSE, EPOCH100, BATCH32) menghasilkan model dengan performa lebih baik dibandingkan dengan metode lainnya.



Gambar 20 - chart metrics MAE



Gambar 21 - chart metrics MSE

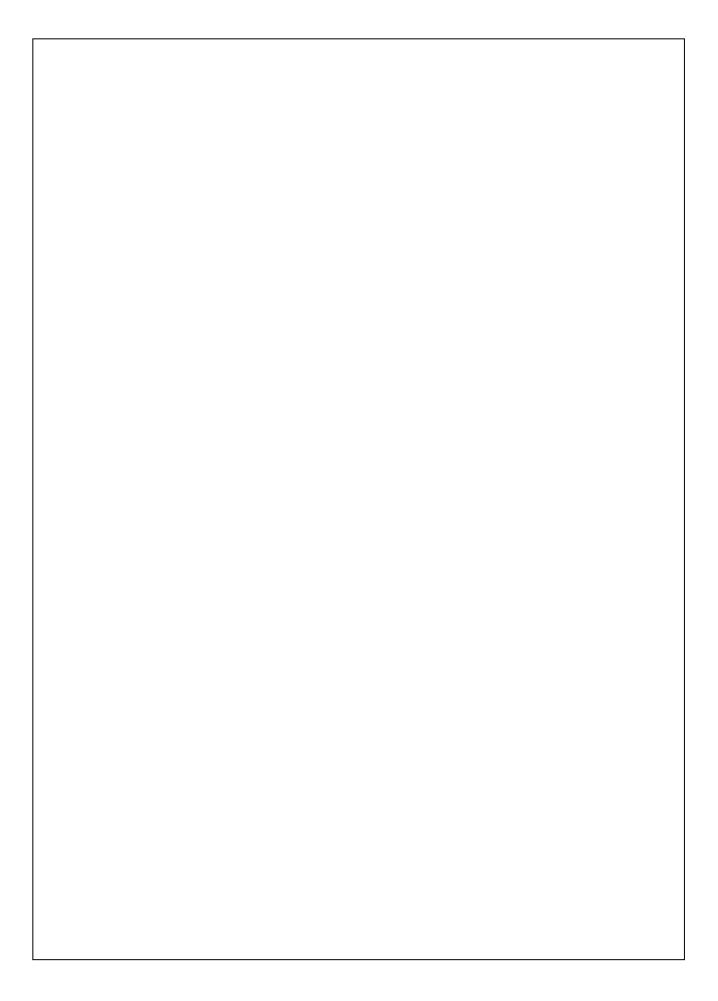


Gambar 22 - chart metrics RMSE

## DAFTAR PUSTAKA

- [1] A. M. Anwar, "PENGARUH CURRENT RATIO (CR), DEBT TO EQUITY RATIO (DER), RETURN ON ASSET (ROA) TERHADAP HARGA SAHAM (Studi kasus pada perusahaan sektor makanan dan minuman yang terdaftar di BEI tahun 2017-2019)," JIMA Jurnal Ilmiah Mahasiswa Akuntansi, pp. 146-157, 2021.
- [2] N. P. L. Narayanti and G., "Pengaruh Kebijakan Dividen Dan Profitabilitas Terhadap Harga Saham Emiten LQ 45 Tahun 2009-2018," Akuntansi, 2020.
- [3] D. Kusumawati and M. Safiq, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI INVESTMENT OPPORTUNITY SET DAN IMPLIKASINYA TERHADAP RETURN SAHAM".
- [4] D. R. Artha, N. A. Achsani and H. Sasongko, "ANALISIS FUNDAMENTAL, TEKNIKAL DAN MAKROEKONOMI HARGA SAHAM SEKTOR PERTANIAN," *JMK*, vol. Vol. 16, no. No. 2, pp. 175-184, 2014.
- [5] A. B. Untoro, "Prediksi Harga Saham Dengan Menggunakan Jaringan Syaraf Tiruan," *Jurnal Teknologi Informatika dan Komputer MH Thamrin*, vol. 06, no. 02, pp. 103-111, 2020.
- [6] E. Patriya, "MPLEMENTASI SUPPORT VECTOR MACHINE PADA PREDIKSI HARGA SAHAM GABUNGAN (IHSG)," Jurnal Ilmiah Teknologi dan Rekayasa, vol. 25, no. 1, pp. 24-38, 2020.
- [7] I. Suryani, "Penerapan Exponential Smoothing untuk Transformasi Data dalam Meningkatkan Akurasi Neural Network pada Prediksi Harga Emas," *Intelegent Systems*, vol. 1, 2015.
- [8] A. Zabdi and D. Pandu, "Tinjauan Pasar Modal," in Pusat Edukasi dan Informasi Pasar Modal di Yogyakarta, p. 8.
- [9] F. S. Mar'ati, "MENGENAL PASAR MODAL," 2020.
- [10] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation," *Journal of Engineering and Applied Technology*, vol. 2, no. 2, pp. 86-91, 2021.

- [11] W. Yustanti, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," *Matematika*, Statistika, Komputasi, vol. 1, 2022.
- [12] W. Yustanti, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," *Jurnal Matematika*, Statistika, dan Komputasi, vol. 9, no. 1, pp. 57-68, 2012.
- [13] R. Amanda, H. Yasin and A. Prahutama, "ANALISIS SUPPORT VECTOR REGRESSION (SVR) DALAM MEMPREDIKSI KURS RUPIAH TERHADAP DOLLAR AMERIKA SERIKAT," Jurnal Gaussian, vol. 3, no. 4, pp. 849-857, 2014.
- [14] D. I. Purnama and S. Setianingsih, "Support Vector Regression (SVR) Model for Forecasting Number of Passengers on Domestic Flights at Sultan Hasanudin Airport Makassar," *Matematika, Statistika, Komputasi*, vol. 16, 2020.
- [15] M. R. Firmansyah, R. Ilyas and F. Kasyidi, "Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network," in *Prosiding The 11th Industrial Research Workshop and National Seminar*, Bandung, 2020.
- [16] E. D. Tarkus, S. R. U. A. Sompie and A. Jacobus, "Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh," *Jurnal Teknik Informatika*, vol. 15, no. 2, pp. 137-144, 2020.
- [17] A. I. Caniago, W. Kaswidjanti and J. Juwairiah, Recurrent Neural Network With Gate Recurrent Unit For Stock Price Prediction, 2019.
- [18] J. Venskus, P. Treigys and J. Markevi ci ut e, "Unsupervised marine vessel trajectory prediction using LSTM network and wild bootstrapping techniques," *Nonlinear Analysis: Modelling and Control*, vol. 26, no. 4, pp. 718-737, 2021.
- [19] S. T. Aini and A. Wiguna, "IMPLEMENTASI ALGORITMA LONG SHORT-TERM MEMORY (LSTM)UNTUK MENDETEKSI UJARAN KEBENCIAN (HATE SPEECH)PADA KASUS PILPRES 2019," *Matrik*, vol. 19, 2019.
- [20] S. R. S. A. J. Exel Defrisco Tarkus, "Implementasi Metode Recurrent Neural Network pada Pengklasifikasian Kualitas Telur Puyuh," *Teknik Informatika*, vol. 15, 2020.



ORIGINALITY REPORT			
20% SIMILARITY INDEX	18% INTERNET SOURCES	10% PUBLICATIONS	6% STUDENT PAPERS
PRIMARY SOURCES			
1 123do Internet So			1 %
2 kc.umi Internet So			1 %
Data T Mesin Memo	Satyo Bayangka ime Series Sahan Belajar LSTM (Lo ry)", Journal of In ation Security, 20	n Bank BRI De ng ShortTerm formatic and	0//
4 github Internet So			1 %
5 Submi	tted to Universita	ıs Dian Nuswa	ntoro 1 %
6 eprints	s2.undip.ac.id		1 %
7 daniels	stephanus.wordp	ress.com	1 %
repo.ia	ain-tulungagung.a	ac.id	

repo.iain-tulungagung.ac.id

		1 %
9	jurnalmahasiswa.stiesia.ac.id Internet Source	1%
10	publikasi.dinus.ac.id Internet Source	1%
11	inmarketing.id Internet Source	1%
12	Mudrik Alaydrus. "Analysis of low pass filter using nonhomogeneous transmission lines", 2014 2nd International Conference on Information and Communication Technology (ICoICT), 2014 Publication	1%
13	ionides.github.io Internet Source	1 %
14	publikasi.mercubuana.ac.id Internet Source	1%
15	repository.teknokrat.ac.id Internet Source	1%
16	jurnal.fikom.umi.ac.id Internet Source	1%
17	repository.unib.ac.id Internet Source	1%

www.scribd.com

Exclude quotes On Exclude bibliography On

Exclude matches

Off

# 

PAGE 1	
PAGE 2	
PAGE 3	
PAGE 4	
PAGE 5	
PAGE 6	
PAGE 7	