

Artificial Neural Network untuk Prediksi Pergerakan Harga Saham Sektor Keuangan dengan Melibatkan Data *Google trends*

Febry Triyadi¹, Deni Saepudin², Aniq Atiqi Rohmawati³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹triadifebry@students.telkomuniversity.ac.id, ²denisaepudin@telkomuniversity.ac.id,

³aniqatiqi@telkomuniversity.ac.id

Abstrak

Dalam dunia bisnis, memprediksi harga saham adalah salah satu tantangan yang sulit bagi para investor, sehingga banyak penelitian yang dilakukan untuk memprediksi harga saham. *Google trends* adalah grafik statistik pencarian web yang menampilkan informasi yang paling trend dan paling banyak menjadi perhatian orang menurut google pada kurun waktu tertentu. Banyak penelitian yang memanfaatkan data *Google trends* untuk memprediksi data *time series*, dikarenakan data *Google trends* itu selalu update dan mudah diakses. Pada penelitian ini, pergerakan harga saham dalam sektor keuangan dengan melibatkan data *Google trends* menggunakan metode *Artificial neural network* (ANN). Penelitian ini juga menyelidiki pengaruh *Google trends* terhadap data harga saham menggunakan metode *Cointegration test* dan *Granger causality analysis*. Data *Google trends* menjadi salah satu cara untuk mempertimbangkan hasil prediksi harga saham. Penelitian ini juga membandingkan dua prediksi menjadi tipe I dan tipe II diantaranya: tipe I prediksi pergerakan harga saham tanpa data *Google trends* dan tipe II prediksi pergerakan harga saham dengan data *Google trends*. Hasil menunjukkan bahwa prediksi dengan penambahan data *Google trends* memberikan dampak yang sedikit lebih signifikan terhadap hasil prediksi dibandingkan prediksi tanpa *Google trends*.

Kata kunci : *Google trends*, harga saham, *Artificial neural network*

Abstract

In the business world, predicting share prices is one of the difficult challenges for investors, so a lot of researches were done to predict share prices. *Google trends* is a graph of web search statistics that shows information which have the most popular information and attention of people according to Google at a certain time. Many studies use *Google trends* data to predict time series data, because *Google trends* data is always up to date and easily accessible. In this study, the movement of stock prices in the financial sector by involving *Google trends* data using the *Artificial neural network* (ANN) method. This study also investigates the effect of *Google trends* on stock price data using the *Cointegration test* and the *Granger causality analysis* methods. *Google trends* data is a way to consider the results of stock price predictions. This study also compares two predictions with Type I and Type II, which include: Type I is prediction of stock price movements without *Google trends* data and Type II is prediction of stock price movements with *Google trends* data. The results show that predictions with the addition of *Google trend* data have a slightly more significant impact on the results of the predictions than the predictions without *Google trends*.

Keywords: *Google trends*, stock price, *Artificial neural network*

1. Pendahuluan

Banyak penelitian untuk memprediksi harga saham dengan berbagai macam pendekatan diantaranya dengan data *time series* bisa menggunakan metode ARIMA atau metode *machine learning* seperti ANN, SVR, SVM, dll. Selain metode *machine learning* banyak penelitian yang memanfaatkan data *Google trends* untuk memprediksi data *time series* seperti prediksi *oil consumption* dan prediksi harga saham, dikarenakan data *Google trends* itu selalu update dan mudah diakses.

Berdasarkan banyaknya penelitian prediksi harga saham yang menggunakan metode ANN dengan hasil yang cukup akurat, maka penelitian ini menggunakan metode *Artificial neural network* (ANN) untuk memprediksi pergerakan harga saham pada sektor keuangan dengan melibatkan data *Google trends*. Data *Google trends* menjadi salah satu cara untuk mempertimbangkan hasil prediksi pergerakan harga saham.

Beberapa contoh penelitian yang berkaitan dengan metode ANN dan prediksi harga saham, seperti yang dilakukan oleh Qiu dan Song memprediksi harga indeks saham Jepang dengan metode ANN yang dioptimalkan. Hasil dari penelitian ini, memprediksi harga indeks saham Jepang 81.27%[1]. Selanjutnya, penelitian Zahedi dan Rounaghi memprediksi harga saham di Bursa Efek Teheran menggunakan metode ANN dan metode Analisis Komponen Utama. Hasil dari penelitian ini, memprediksi harga saham di Bursa Efek Teheran 0.0018[2].

Beberapa penelitian lain ada yang melibatkan data *Google trends* untuk membantu memprediksi *oil consumption* dan memprediksi harga indeks saham, seperti penelitian yang dilakukan oleh Zhao, Yu, Tang, dan Yang memprediksi *oil consumption* menggunakan data *Google trends*. Pada penelitian ini terdapat dua jenis prediksi yaitu: prediksi dengan menggunakan *Google trends* dan prediksi tanpa menggunakan *Google trends*. Hasil dari prediksi ini menyatakan bahwa prediksi menggunakan *Google trends* meningkatkan teknik tradisional tanpa *Google trends* secara signifikan untuk prediksi arah dan level, Zhao dkk mengatakan bahwa hasil prediksinya akan berubah-ubah seiringnya dengan waktu[3]. Selanjutnya, penelitian Hongping, Li, dan Zhang juga memprediksi arah harga pembukaan saham pada indeks S&P 500 dan indeks rata-rata industri Dow Jones dengan data *Google trends* untuk meningkatkan prediksi harga indeks saham. Hasil dari penelitian ini, memprediksi indeks S&P 500 86,81% dan indeks Dow Jones 88,98% [4].

Berdasarkan penelitian sebelumnya yang dilakukan oleh Zhao tentang *oil consumption* dengan *Google trends* dan Hongping tentang prediksi harga index saham dengan *Google trends*, maka penelitian ini memprediksi pergerakan harga saham dalam sektor keuangan dengan melibatkan data *Google trends* menggunakan metode *Artificial neural network* (ANN), dan menyelidiki pengaruh *Google trends* terhadap data harga saham menggunakan metode *Cointegration test* dan *Granger causality analysis*, dan penelitian ini membandingkan dua prediksi yaitu, prediksi pergerakan harga saham tanpa data *Google trends* dan prediksi pergerakan harga saham dengan data *Google trends*.

Tujuan

Berdasarkan penelitian sebelumnya, tujuan dari penelitian ini adalah :

1. Bagaimana implementasi *Granger causality analysis* dan *Cointegration test* untuk mengetahui hubungan antara *Google trends* dengan harga saham.
2. Mengetahui pengaruh penambahan data *Google trends* terhadap prediksi harga saham sektor keuangan.

2. Penelitian Terkait

Penelitian yang dilakukan oleh Yakup Kara, Melek Acar Boyacioglu, dan Omer Kaan Baykan pada tahun 2011, yang berjudul “*Predicting direction of stock price index movement using Artificial neural network and Support vector machines: The sample of the Istanbul Stock Exchange*”. Penelitian ini mengembangkan dua model yang efisien dan membandingkan kinerja dalam memprediksi arah pergerakan di bursa 100 Nasional Bursa Efek Istanbul (ISE) harian. Metode yang digunakan adalah *Artificial neural network* (ANN) dan *Support vector machine* (SVM) [5].

Berikutnya penelitian yang dilakukan oleh Mingyue Qiu dan Yu Song pada tahun 2016, yang berjudul “*Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial neural network Model*”. Penelitian ini memprediksi pergerakan harga indeks saham menggunakan metode *Artificial neural network* (ANN) yang dioptimalkan. Penelitian ini memodifikasi metode ANN dengan metode Algoritma Genetika (GA) menjadi metode *hybrid GA-ANN* sebagai performansi dalam penelitian ini [1].

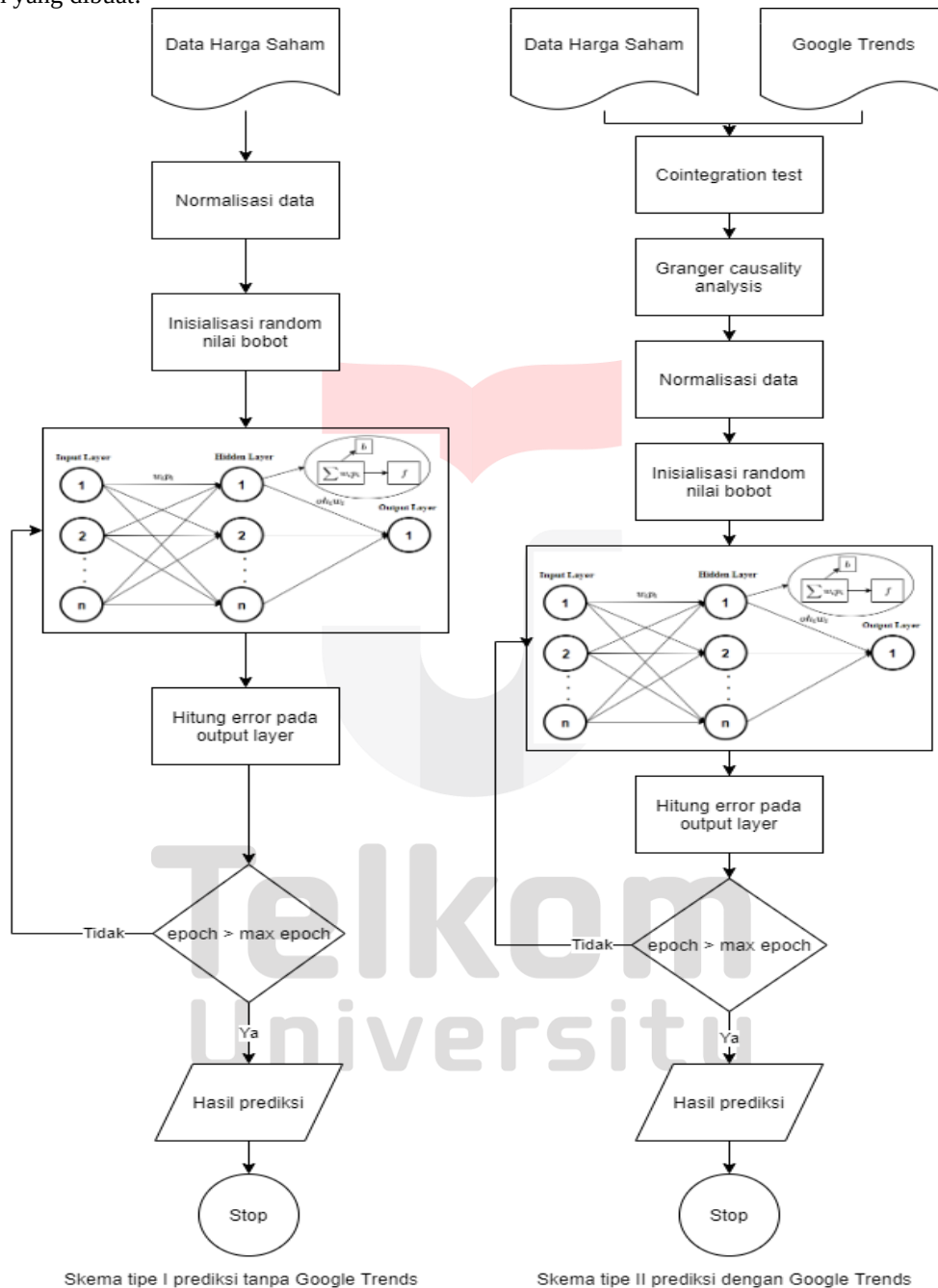
Selanjutnya, penelitian yang dilakukan oleh Tobias Preis, Helen Susannah Moat, dan H. Eugene Stanley pada tahun 2013, yang berjudul “*Quantifying Historical Behavior in Financial Markets Using Google trends*”. Penelitian ini menganalisis perubahan volume kueri Google untuk data pencarian yang berkaitan dengan keuangan. Pada penelitian ini, menghasilkan gambaran potensi yang menggabungkan data set perilaku yang luas menawarkan untuk pemahaman yang lebih baik dari perilaku manusia kolektif. Dengan menggunakan strategi dimana mereka mengambil posisi user dari jauh dan dekatnya menggunakan data volume pencarian Amerika Serikat [6].

Penelitian yang dilakukan oleh Hongping Hu, Li Tang, Shuhua Zhang, dan Haiyan Wang pada tahun 2018, yang berjudul “*Predicting the direction of stock markets using optimized neural networks with Google trends*”. Penelitian ini memprediksi arah harga pembukaan saham untuk indeks S&P 500 dan indeks rata-rata industri Dow Jones, lalu dengan data *Google trends* untuk meningkatkan prediksi harga saham. Penelitian ini menggunakan metode gabungan dari *Improved Sine Cosine Algorithm* (ISCA) dengan *Backpropagation Neural Networks* (BPNN) menciptakan jaringan baru ISCA-BPNN. *Hit ratio* penelitian ini mencapai 86,81% untuk indeks S&P 500 dan 88,98% untuk indeks DJIA. Hasil menunjukkan bahwa ISCA-BPNN mampu memprediksi harga saham, dan data *Google trends* dapat membantu meningkatkan nilai prediksi harga saham [4].

Selanjutnya, penelitian yang dilakukan oleh Lean Yu, Yaqing Zhao, Ling Tang, dan Zebin Yang pada tahun 2019, yang berjudul “*Online big data-driven oil consumption forecasting with Google trends*”. Pada penelitian ini, memprediksi *global oil consumption* dengan data *Google trends* menggunakan hubungan investigasi *Cointegration test*, *Granger causality test*, dan *Google trends* dengan hubungan yang signifikan. Penelitian ini pada metode prediksi menggunakan banyak algoritma yang telah dicoba seperti SVR, LR, BPNN, dan ELM. Lalu terdapat 2 model dalam memprediksi *oil consumption* sebagai berikut: prediksi model tanpa *Google trends*, dan prediksi model dengan *Google trends*. Hasil prediksi model dengan *Google trends* relative terhadap model tanpa *Google trends* sekitar 1,54% dan 1.8637. Sedangkan, model tanpa Google sekitar 1.58% dan 1.8939. Hasil menunjukkan bahwa *Google trends* meningkatkan prediksi tanpa *Google trends* secara signifikan [3].

3. Sistem yang Dibangun

Pada penelitian ini, menggunakan metode *Artificial neural network* (ANN) untuk memprediksi pergerakan harga saham dalam sektor keuangan yang ada di Indonesia dengan data tambahan *Google trends* untuk mempertimbangkan hasil akurasi. Penelitian ini memiliki dua skema umum, ada yang skema untuk tipe I tanpa *Google trends* dan skema tipe II dengan *Google trends*. Kedua skema ini dibandingkan performansinya, berikut skema sistem yang dibuat:



Gambar 3. 1 Skema Umum Prediksi Harga Saham

Pada Gambar 3.1 terdapat 2 tipe skema yaitu: skema tipe I prediksi tanpa *Google trends* dan skema tipe II prediksi dengan *Google trends*. Tipe I untuk prediksi tanpa *Google trends* tidak menggunakan metode *Cointegration test* dan *Granger causality analysis*. Sedangkan, tipe II prediksi dengan *Google trends* menggunakan metode tersebut. Pada tipe I hanya data set harga saham yang dijadikan *input*, tipe II terdapat 2 input data yaitu data set harga saham dan data *Google trends*. Untuk proses tipe II lebih banyak dibandingkan tipe I dikarenakan tipe II memiliki proses *Cointegration test* dan *Granger causality analysis* setelah itu proses *Normalisasi data* set pengskalaan data agar data memiliki nilai yang berbeda-beda, lalu inisialisasi bobot dan *bias* secara random *upper* dan *lower bound*, dan masuk ke dalam model ANN mendapatkan nilai error dan

akurasi dari *output layer* berdasarkan data training, kemudian pengecekan apakah nilai *epoch* sudah lebih dari *max epoch* (*max epoch* = 500), jika belum proses kembali ke model ANN dan jika sudah maka muncul hasil prediksi.

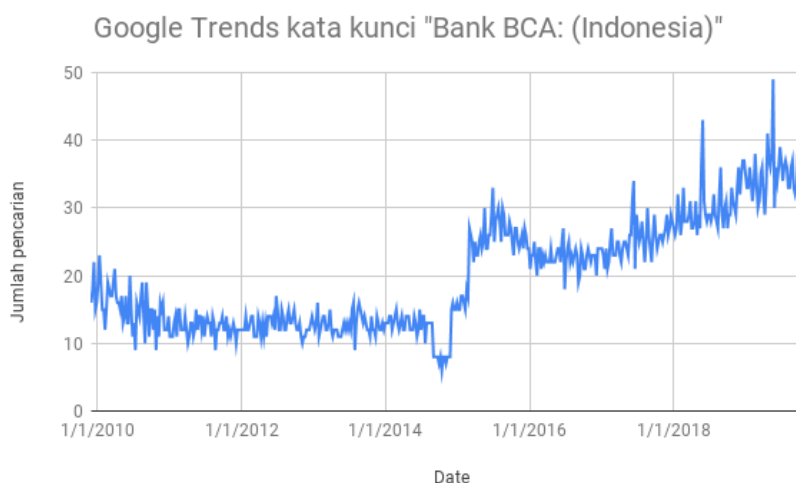
Pertama, penelitian ini menggunakan data set harga saham dalam sektor keuangan berdasarkan nilai kapital market tertinggi, dan diambil 3 tinggi yaitu bank BRI, bank BCA, dan bank Mandiri, diambil dari website Yahoo Finance (<https://www.finance.yahoo.com>) dan website PT Bursa Efek Indonesia (<https://www.idx.co.id>). Data yang diolah merupakan data harga saham mingguan periode 2009-2019, dengan total 522 data harga saham. Kedua, menambahkan data *Google trends* (<https://www.google.com/trends>) sebagai data pencarian pada google berdasarkan user [7], dengan 12 kata kunci pencarian yaitu: “Bank BRI: (Indonesia)”, “Bank BCA: (Indonesia)”, “Bank Mandiri: (Indonesia)”, “BBCA: (Indonesia)”, “BBRI: (Indonesia)”, “BMRI: (Indonesia)”, “Bank BRI: (Worldwide)”, “Bank BCA: (Worldwide)”, “Bank Mandiri: (Worldwide)”, “BBCA: (Worldwide)”, “BBRI: (Worldwide)”, “BMRI: (Worldwide)” dengan masing-masing memiliki jumlah data sebanyak 522.

Dalam pemodelan, sampel data perlu dibagi menjadi dua *subset* yaitu data *training* dan data *testing*, dimana jumlah data *training* harus lebih besar dari jumlah data *testing* untuk menghindari adanya suatu nilai dari data *training* yang tidak bisa dihasilkan oleh model *time series* dan *Artificial neural network* karena kurangnya data *training* [8].

Berdasarkan data set *Google trends* dan data set harga saham diproses melalui ANN menghasilkan nilai prediksi dengan mempertimbangkan data *training* dan data *testing*. Data *training* diambil sebanyak 80% dari data set dan data *testing* 20% dari data set. Data *training* digunakan untuk membangun model ANN dan data *testing* digunakan mengetahui akurasi model ANN. Pada data *training* juga bisa menghasilkan akurasi, tetapi akurasi belum bisa dipakai sebagai acuan, jika belum dilakukan *testing* model dengan menggunakan data *testing*. Sehingga, yang dicari pasti akurasi *testing* bukan *training*.



Gambar 3. 2 Contoh grafik data set harga saham BBKA



Gambar 3. 3 Contoh grafik data *Google trends* dengan kata kunci “Bank BCA: (Indonesia)”

Pada Gambar 3.2 adalah contoh grafik data set harga saham pada perusahaan bank BCA dengan kode BBKA. Sedangkan, pada Gambar 3.3 adalah contoh grafik data *Google trends* menggunakan data pencarian user dengan

kata kunci “Bank BCA: (Indonesia)”. Data ini digunakan sebagai pertimbangan tingkat akurasi prediksi harga saham.

Proses pertama yang dilakukan adalah *Cointegration test* pada data set harga saham dan data *Google trends*. *Cointegration test* adalah suatu hubungan jangka panjang antara peubah-peubah yang meskipun secara individual tidak stasioner, tetapi kombinasi linier antara peubah tersebut dapat menjadi stasioner[9]. *Cointegration test* dapat digunakan untuk mengetahui data harga saham dan data *Google trends* memiliki hubungan keseimbangan jangka panjang atau tidak. Apabila data telah stasioner artinya antara data tersebut terkointegrasi atau memiliki hubungan jangka panjang. *Cointegration test* ini menyelidiki data *Google trends* yang memiliki pengaruh terhadap data harga saham berdasarkan data harga *close*, dan menguji secara statistik akurasi hasil prediksi *Google trends* untuk harga saham. Dalam *Cointegration test* dua *time series* X_t dan Y_t terkointegrasi jika keduanya stasioner pada urutan yang berbeda dan residu regresi linier u_t juga stasioner. Setelah *Cointegration test* mendapatkan nilai stasioner dari X_t dan Y_t . Selanjutnya, melakukan tes *Augmented dickey-fuller* (ADF) untuk menguji stasioneritas. Berikut persamaan *Cointegration test* dari seri stasioner X_t dan Y_t [3]:

$$Y_t = a_0 + a_1 X_t + u_t \quad (1)$$

dimana a_0 adalah nilai konstan, u_t adalah deret residu yang akan diuji stasioner, X_t adalah data harga saham pada waktu t menggunakan ADF test untuk mengetahui apakah residu u_t bernilai stasioner. Berikut persamaan ADF dari variabel dependen ke model yang ada [10]:

$$\Delta Y_t = \alpha Y_{t-1} + \sum_{j=1}^p \beta_j \Delta Y_{t-j} + u_t \quad (2)$$

dimana p adalah tingkat augmentasi, β adalah nilai konstan pada tren waktu, ΔY_t adalah level lag, dan tes ADF yang menarik untuk $H_0: \alpha = 0$ terhadap $H_0: \alpha < 0$. Jika hasil residu u_t adalah stasioner, maka pemodelan X_t dan Y_t pada persamaan (1) dapat diterapkan.

Selanjutnya, proses *Granger causality analysis* untuk mencari data *Google trends* yang lebih berpengaruh terhadap data harga saham atau tidak berpengaruh sama sekali. Tujuan *Granger causality* adalah meneliti apakah X_t mendahului Y_t , ataukah Y_t mendahului X_t , ataukah hubungan antara X_t dan Y_t timbal balik. Uji *Granger causality* bertujuan untuk melihat pengaruh masa lalu dari suatu variabel terhadap kondisi variabel lain pada masa sekarang. Dengan kata lain, uji *Granger causality* dapat digunakan untuk melihat apakah peramalan dengan *Google trends* dapat lebih akurat dengan memasukan lag. Proses *Granger causality analysis* ini mendeteksi efek dari data *Google trends* terhadap data harga saham berdasarkan nilai stasioner.

Granger causality untuk kasus dua seri waktu bernilai stasioner X_t dan Y_t didefinisikan sebagai berikut: misalkan $F(X_t | I_{t-1})$ adalah distribusi peluang bersyarat dari X_t yang bila diberikan informasi bivariat $I_{t-1} = \{X_{t-Lx}^{Lx}, Y_{t-Ly}^{Ly}\}$, dimana $X_{t-Lx}^{Lx} = \{X_{t-Lx}, \dots, X_{t-1}\}$, dan $Y_{t-Ly}^{Ly} = \{Y_{t-Ly}, \dots, Y_{t-1}\}$. Jika *time series* Y_t tidak dapat membantu prediksi X_t maka Lx dan Ly diberikan lag [11]

$$F(X_t | I_{t-1}) = F(X_t | (I_{t-1} - Y_{t-Ly}^{Ly})), \text{ dengan } t = 1, 2, \dots, T. \quad (3)$$

Jika persamaan (3) tidak benar, maka nilai Y_t di masa lalu dapat membantu untuk memprediksi nilai X_t , dan nilai Y_t . Kekurangan *Granger causality* dari Y_t ke X_t terjadi jika persamaan seperti berikut

$$F(X_t | I_{t-1}) = F(X_t | (I_{t-1} + Y_t)) \quad (4)$$

dimana informasi bivariat dimasukkan ke nilai Y_t saat ini. Jika persamaan (4) tidak berlaku, maka Y_t dikatakan *Granger* menyebabkan X_t . Terdapat model tambahan dalam model *Granger causality* yaitu *Vector autoregression* (VAR), model ini menggunakan rumus.

$$X_t = a_0 + a_1 X_{t-1} + \dots + a_{Lx} X_{t-Lx} + b_1 Y_{t-1} + \dots + b_{Ly} Y_{t-Ly} + u_t, \quad (5)$$

$$Y_t = a'_0 + a'_1 Y_{t-1} + \dots + a'_{Ly} Y_{t-Ly} + b'_1 X_{t-1} + \dots + b'_{Lx} X_{t-Lx} + v_t \quad (6)$$

dimana u dan v adalah error yang saling independen dan individual distribusi, dengan nol yang berarti nilai awal dan nilai konstan. *Standard test* gabungan dilakukan untuk menguji signifikan koefisien $b_i (i = 1, \dots, Ly)$ dan $b'_i (i = 1, \dots, Lx)$ secara individual. Jika koefisien yang didapat menyimpang dari nol, *Granger causality* diproses mulai dari Y_t ke X_t atau X_t ke Y_t dapat dibuktikan berdasarkan persamaan (5) dan (6).

Pada proses *Granger causality analysis* didapatkan hasil hubungan antara data harga saham dengan data *Google trends* yang lebih dominan untuk diprediksi berdasarkan dari lag 1 sampai lag 6, nilai yang mencapai

tingkat signifikan level 5%, maka diambil data kata kunci *Google trends* untuk memprediksi pergerakan harga saham menggunakan metode ANN.

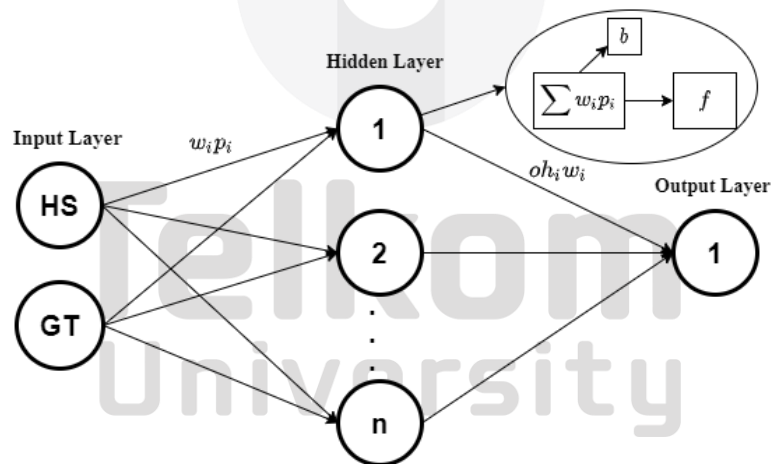
Setelah ditentukan data *Google trends* yang lebih dominan dengan tingkat signifikan level 5%, proses selanjutnya adalah normalisasi data. Pada metode normalisasi nilai data atribut dikurangi nilai data atribut *minimum* dan dibagi dengan nilai atribut *maximum* sebelumnya yang telah dikurangi data *minimum*. Tujuan dari normalisasi data ini untuk mengubah nilai atribut data yang memiliki rentang 0 sampai 1 dengan nilai yang berbeda-beda sehingga semua nilai atribut data memiliki nilai bobot yang sama. Berikut rumus normalisasi data yang terdapat pada persamaan:

$$\text{Normalization data} = \frac{\text{data} - \text{data minimum}}{\text{data maximum} - \text{data minimum}} \quad (7)$$

Selanjutnya, masuk kedalam tahap proses *Artificial neural network*. Pada proses ini dilakukan *forward pass* mulai dari *input layer* yang berisikan data set dan data *Google trends* menuju *hidden layer* dengan nilai bobot yang berbeda-beda yang sudah diinisialisasikan pada proses sebelumnya, diproses dengan nilai *bias* pada setiap neuron menghasilkan nilai *output*, dan menuju *output layer* dengan membawa nilai *output* tadi dan nilai bobot yang baru, diproses pada *output layer* menghasilkan nilai *output*.

Pada ANN terdapat beberapa *node* yang saling memiliki keterkaitan antara satu dengan yang lain dimana *node* tersebut biasa dikenal dengan istilah neuron. ANN adalah bentuk representasi dari neuron yang saling terhubung selayaknya cara kerja otak manusia. Berdasarkan prinsip dasar ANN, ANN setidaknya memiliki 3 *layer* yaitu : *input layer*, *hidden layer*, dan *output layer*. *Input layer* adalah *node* yang memiliki atribut dan nilai yang dijadikan masukan ke sistem *neural network*, yang biasanya berhubungan dengan variabel independen pada metode regresi linier. *Hidden layer* berfungsi untuk melakukan transformasi non linier terhadap atribut dan nilai dari *input*. *Output layer* menghitung jumlah bobot yang dihasilkan dari *hidden layer* [12]. Pada setiap koneksi antara *layer* mulai dari *input layer* sampai *output layer* memiliki nilai bobot yang nilainya berbeda-beda. *Hidden layer* dan *output layer* memiliki nilai tambahan yaitu *bias*, *bias* ini untuk menjaga nilai *output* agar tidak bernilai 0.

Fungsi aktivasi hanya dimiliki oleh *hidden layer* dan *output layer*, sedangkan *input layer* tidak memiliki fungsi aktivasi. Pada proses fungsi aktivasi sebuah neuron menerima *input*, mengaktifkan neuron tersebut berdasarkan kekuatan sinyal *input*.



Gambar 3. 4 Arsitektur Artificial neural network.

Pada Gambar 3.4 adalah arsitektur ANN pada umumnya. Penelitian ini memiliki 2 *input layer* yang terdiri dari data harga saham dan data *Google trends*, dari *input layer* terdapat nilai bobot yang diinisialisasikan sebagai w_i dan p_i nilai vektor *input*. Lalu menuju *hidden layer* untuk proses pada fungsi aktivasi, pada *hidden layer* ini terdapat nilai *bias* agar hasil *output* oh_i tidak bernilai 0. Ketika menuju *output layer*, oh_i dihitung dengan nilai bobot w_i setelah itu *output layer* dihitung *output error* yang nantinya dibandingkan nilainya dengan nilai target error.

Pada dasarnya fungsi aktivasi ini memiliki dua jenis yaitu: linear dan non linier. Berikut beberapa fungsi aktivasi non linier yang diimplementasikan oleh *perceptron* [13]:

1. Fungsi aktivasi sigmoid

$$f(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Persamaan (8) adalah fungsi aktivasi sigmoid yang mempunyai rentang antara 0 hingga 1.

2. Fungsi aktivasi tanh

$$f(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (9)$$

Persamaan (9) adalah fungsi aktivasi tanh yang mempunyai rentang antara -1 hingga 1.

3. Fungsi aktivasi ReLU

$$R(z) = \max(0, z) \quad (10)$$

Persamaan (10) adalah fungsi aktivasi ReLU yang melakukan *threshold* dari 0 hingga tak terhingga.

Setelah proses fungsi aktivasi diberikan nilai bobot dan kemudian mendistribusikan ke neuron yang ada pada *layer* selanjutnya. Setiap neuron memiliki nilai *threshold* menjadi nilai *minimum* yang harus dicapai hingga neuron aktif. Proses untuk menghasilkan *output* pada *layer* mempunyai persamaan sebagai berikut [14]:

$$oh = f\left(\sum_{i=1}^n w_i p_i + b\right) \quad \mathbf{p} = (p_1, p_2, \dots, p_n) \quad (11)$$

dimana w adalah bobot, $\mathbf{p}$ adalah vektor input, f adalah fungsi aktivasi, i adalah panjang data, dan b adalah bias.

Pada penelitian ini menggunakan library dari keras (<https://keras.io/>) untuk membuat model arsitektur ANN. Berdasarkan penelitian yang dilakukan oleh Hillier untuk proses inisialisasi pada *neural network* terdiri dari [15]:

1. Inisial bobot menggunakan *random upper* dan *lower bounds*.
2. Fungsi aktivasi, untuk *hidden layer* menggunakan fungsi aktivasi sigmoid, dan untuk *output layer* menggunakan fungsi aktivasi linear.
3. Jumlah *epoch*, pada penelitiannya jumlah *epoch* yang memiliki nilai optimasi yaitu *epoch* 500.
4. Fungsi error yang digunakan adalah *mean squared error* (MSE).

Mean Square Error (MSE) adalah metode untuk mengukur tingkat keakuratan. Jika nilai MSE semakin kecil, lebih konsisten membuat nilai *output* yang memiliki nilai error rendah berapapun nilai yang diinputkan. Berikut rumus umum MSE pada metode ANN untuk memprediksi harga saham.

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2 \quad (12)$$

dimana n adalah jumlah data, y_t adalah nilai sebenarnya dan f_t adalah hasil prediksi.

Root mean square error (RMSE) merupakan parameter yang digunakan untuk mengevaluasi nilai hasil dari pengukuran terhadap nilai sebenarnya atau nilai dianggap benar. Semakin kecil nilai RMSE, maka pengklasteran data semakin mendekati benar.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2} \quad (13)$$

dimana n adalah jumlah data, y_t adalah nilai sebenarnya dan f_t adalah hasil prediksi.

Autocorrelations antara dua variable X_t dan Y_t didefinisikan sebagai berikut:

$$\rho = \frac{E[(X_t - \mu_1)(Y_t - \mu_2)]}{\sigma_1 \sigma_2} \quad (14)$$

dimana E adalah operator ekspektasi, μ_1 dan μ_2 adalah sarana masing-masing untuk X_t dan Y_t , dan σ_1, σ_2 adalah standar deviasi.

Partial autocorrelations adalah mengukur ketergantungan linear dari satu variabel setelah menghilangkan efek variabel lain yang mempengaruhi kedua variabel. Contoh, *Partial autocorrelations* urutan mengukur efek (ketergantungan linear) dari y_{t-2} pada y_t setelah menghilangkan efek y_{t-1} pada y_t dan y_{t-2} . Setiap *Partial autocorrelations* dapat diperoleh sebagai serangkaian regresi formulir [16]:

$$\tilde{y}_t = \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + e_t \quad (15)$$

di mana $\tilde{y}_t$ adalah seri asli dikurangi rata-rata sampel, $y_t - \bar{y}$. Estimasi ϕ_{22} akan memberikan nilai *Partial autocorrelations* dari dua pesanan. Memperluas regresi dengan k keterlambatan tambahan, estimasi istilah terakhir akan memberikan *Partial autocorrelations* urutan k . Pada penelitian ini untuk memprediksi data y_t menggunakan data y_{t-1} dan y_{t-2} .

4. Evaluasi dan Analisis

Pada penelitian ini, menggunakan data set harga saham dan data set *Google trends* untuk mengetahui apakah *Google trends* dan harga saham memiliki hubungan satu sama lain dalam *time series* menggunakan metode *Cointegration test* dan *Granger causality analysis*, dan memprediksi harga saham dalam sektor keuangan dengan data tambahan *Google trends* menggunakan metode *Artificial neural network*.

Langkah pertama, dari skema model yang telah dibuat melakukan proses *Cointegration test* dan *Granger causality analysis* untuk memilih *Google trends* yang efektif, dalam hal memiliki hubungan yang signifikan dengan harga saham.

Hasil pengujian stasioner metode ADF

Table 1 Hasil pengujian stasioner dalam t-statistik(p-value)

<i>Time series</i>	<i>Data original</i>	<i>First-order differencing</i>
Harga saham bank BCA	2.0106 (0.9987)	-6.0064 (0.0000)
Harga saham bank BRI	-0.2130 (0.9370)	-25.0835 (0.0000)
Harga saham bank Mandiri	-1.2925 (0.6326)	-27.6236 (0.0000)
Kata kunci <i>Google trends</i> "BBCA: (Indonesia)	1.1123 (0.9953)	-5.4026 (0.0000)
Kata kunci <i>Google trends</i> "Bank BCA: (Indonesia)	-0.2834 (0.9278)	-6.8539 (0.0000)
Kata kunci <i>Google trends</i> "BBCA: (Worldwide)	-1.6111 (0.4774)	-10.3256 (0.0000)
Kata kunci <i>Google trends</i> "Bank BCA: (Worldwide)	-0.4840 (0.8951)	-9.4789 (0.0000)
Kata kunci <i>Google trends</i> "BBRI: (Indonesia)	1.5060 (0.9975)	-17.727 (0.0000)
Kata kunci <i>Google trends</i> "Bank BRI: (Indonesia)	0.4478 (0.9832)	-6.4878 (0.0000)
Kata kunci <i>Google trends</i> "BBRI: (Worldwide)	3.3214 (1.0000)	-6.6703 (0.0000)
Kata kunci <i>Google trends</i> "Bank BRI: (Worldwide)	0.0289 (0.9609)	-7.2237 (0.0000)
Kata kunci <i>Google trends</i> "BMRI: (Indonesia)	-2.0567 (0.2623)	-8.4271 (0.0000)
Kata kunci <i>Google trends</i> "Bank Mandiri: (Indonesia)	-2.5141 (0.1121)	-7.2296 (0.0000)
Kata kunci <i>Google trends</i> "BMRI: (Worldwide)	0.0352 (0.9613)	-9.1725 (0.0000)
Kata kunci <i>Google trends</i> "Bank Mandiri: (Worldwide)	-1.6260 (0.4696)	-8.5860 (0.0000)

Pertama, proses uji stasioner menggunakan tes *Augmented dickey-fuller* (ADF), mendapatkan hasil seperti di Tabel 1. Harga saham bank BCA, bank BRI, dan bank Mandiri pada data *original* tidak bernilai stasioner, dan juga semua kata kunci *Google trends* pada data *original* tidak bernilai stasioner melainkan non-stasioner dari tingkat signifikan 5%, maka dari itu dilakukan proses differensing untuk mengubah skala pada data agar menjadi stasioner. Berdasarkan tes ADF, data hasil differensing pertama menunjukkan bahwa data stasioner pada tingkat signifikan 5%. Jadi harga saham bank BCA, bank BRI, dan bank Mandiri memenuhi kondisi hubungan *cointegration* dan *Granger*, dan semua kata kunci *Google trends* memenuhi kondisi hubungan *cointegration* dan *Granger*.

Hasil pengujian *Cointegration test*

Table 2 Hasil pengujian *Cointegration test* dalam t-statistik (p-value)

<i>Google trends</i>	<i>First-order differencing</i>
Kata kunci <i>Google trends</i> "BBCA: (Indonesia)	-6.3128 (3.3332)
Kata kunci <i>Google trends</i> "Bank BCA: (Indonesia)	-6.3137 (3.3162)
Kata kunci <i>Google trends</i> "BBCA: (Worldwide)	-6.2889 (3.7689)
Kata kunci <i>Google trends</i> "Bank BCA: (Worldwide)	-6.3067 (3.4397)
Kata kunci <i>Google trends</i> "BBRI: (Indonesia)	-25.2377 (0.0000)
Kata kunci <i>Google trends</i> "Bank BRI: (Indonesia)	-25.2435 (0.0000)
Kata kunci <i>Google trends</i> "BBRI: (Worldwide)	-25.1651 (0.0000)
Kata kunci <i>Google trends</i> "Bank BRI: (Worldwide)	-25.2394 (0.0000)
Kata kunci <i>Google trends</i> "BMRI: (Indonesia)	-27.8606 (0.0000)
Kata kunci <i>Google trends</i> "Bank Mandiri: (Indonesia)	-27.8474 (0.0000)
Kata kunci <i>Google trends</i> "BMRI: (Worldwide)	-27.8134 (0.0000)
Kata kunci <i>Google trends</i> "Bank Mandiri: (Worldwide)	-27.8851 (0.0000)

Kedua, *Cointegration test* digunakan untuk menguji hubungan *cointegration* antara harga saham dan *Google trends*, seperti pada Table 2. Hasil menunjukkan, pada kata kunci bank BCA tidak bernilai stasioner terhadap harga saham bank BCA, maka pada kata kunci *Google trends* "BBCA: (Indonesia)", "Bank BCA: (Indonesia)", "BBCA: (Worldwide)", dan "Bank BCA: (Worldwide)" tidak memiliki hubungan *cointegration* terhadap harga saham bank BCA. Selain itu kata kunci bank BRI dan bank Mandiri, mereka berada ditingkat signifikan 5%, maka harga saham

bank BRI dan bank Mandiri memiliki hubungan *cointegration* terhadap kata kunci *Google trends* yang berkaitan dari masing-masing bank.

Hasil pengujian *Granger causality analysis*

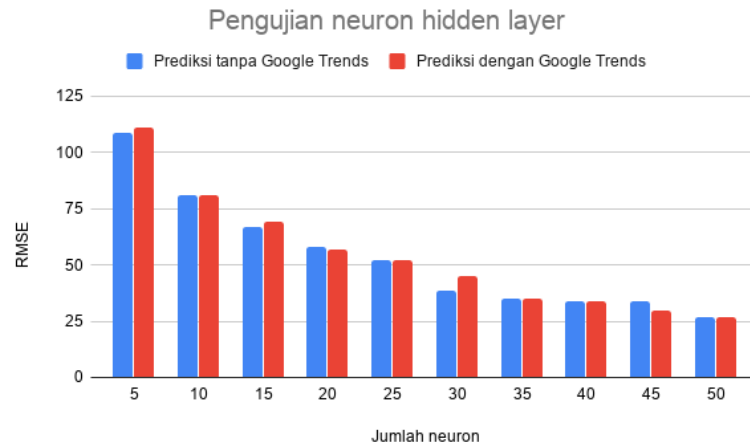
Table 3 Hasil pengujian *Granger causality analysis*

	Lags					
	1	2	3	4	5	6
Panel BRI A	H0: Kata kunci <i>Google trends</i> “BBRI: (Indonesia)” tidak menyebabkan <i>Granger</i> Harga saham bank BRI					
F-stat	0.0222	0.4130	0.6670	0.5252	0.4950	0.4124
p-value	0.8815	0.6619	0.5726	0.7173	0.7801	0.8709
Panel BRI B	H0: Kata kunci <i>Google trends</i> “Bank BRI: (Indonesia)” tidak menyebabkan <i>Granger</i> Harga saham bank BRI					
F-stat	0.7221	0.3965	0.2518	0.2610	0.5244	0.4505
p-value	0.3958	0.6728	0.8601	0.9029	0.7579	0.8447
Panel BRI C	H0: Kata kunci <i>Google trends</i> “BBRI: (Worldwide)” tidak menyebabkan <i>Granger</i> Harga saham bank BRI					
F-stat	1.2574	0.6898	0.6466	0.9359	1.0006	0.9295
p-value	0.2627	0.5021	0.5854	0.4427	0.4167	0.4733
Panel BRI D	H0: Kata kunci <i>Google trends</i> “Bank BRI: (Worldwide)” tidak menyebabkan <i>Granger</i> Harga saham bank BRI					
F-stat	0.2077	0.1013	0.1137	0.0835	0.2426	0.2041
p-value	0.6487	0.9037	0.9521	0.9875	0.9434	0.9755
Panel Mandiri A	H0: Kata kunci <i>Google trends</i> “BMRI: (Indonesia)” tidak menyebabkan <i>Granger</i> Harga saham bank Mandiri					
F-stat	0.0246	0.1767	0.1030	0.1217	0.2703	0.2529
p-value	0.8755	0.8381	0.9583	0.9747	0.9293	0.9581
Panel Mandiri B	H0: Kata kunci <i>Google trends</i> “Bank Mandiri: (Indonesia)” tidak menyebabkan <i>Granger</i> Harga saham bank Mandiri					
F-stat	0.0489	0.7103	0.8917	2.2959	1.8670	1.5835
p-value	0.8251	0.4920	0.4452	0.0582	0.0985	0.1498
Panel Mandiri C	H0: Kata kunci <i>Google trends</i> “BMRI: (Worldwide)” tidak menyebabkan <i>Granger</i> Harga saham bank Mandiri					
F-stat	1.5634	2.0433	1.5657	1.4218	1.1664	1.5639
p-value	0.1497	0.1307	0.1968	0.2255	0.3246	0.1556
Panel Mandiri D	H0: Kata kunci <i>Google trends</i> “Bank Mandiri: (Worldwide)” tidak menyebabkan <i>Granger</i> Harga saham bank Mandiri					
F-stat	0.0563	0.5339	0.5928	1.5073	1.8465	1.6884
p-value	0.8125	0.5866	0.6199	0.1987	0.1022	0.1218

Ketiga, *Granger causality analysis* dilakukan untuk mengeksplorasi secara statistik kata kunci *Google trends* dapat membantu memprediksi harga saham bank, dengan jumlah lag dari 1 sampai 6, seperti yang ditunjukkan pada Tabel 3. Pada Tabel 3 kata kunci *Google trends* tidak ada yang berada pada tingkat signifikan 5% tetapi ada beberapa yang berada pada tingkat signifikan 10%, yaitu pada Panel Mandiri B lag 4 dan 5 kata kunci “Bank Mandiri: (Indonesia)”, maka kata kunci *Google trends* tersebut memiliki hubungan *Granger* yang lebih dominan untuk diprediksi. Namun untuk kata kunci *Google trends* yang berkaitan dengan bank BRI tidak menunjukkan hubungan *Granger* dikarenakan tidak signifikan pada 5% maupun 10%.

Hasil dari pengujian ketiga dapat disimpulkan, hanya beberapa kata kunci *Google trends* yang dapat digunakan untuk prediksi harga saham yaitu pada harga saham bank Mandiri dengan kata kunci *Google trends* “Bank Mandiri: (Indonesia)” pada lag 4 dan 5.

Hasil pengujian neuron *hidden layer*



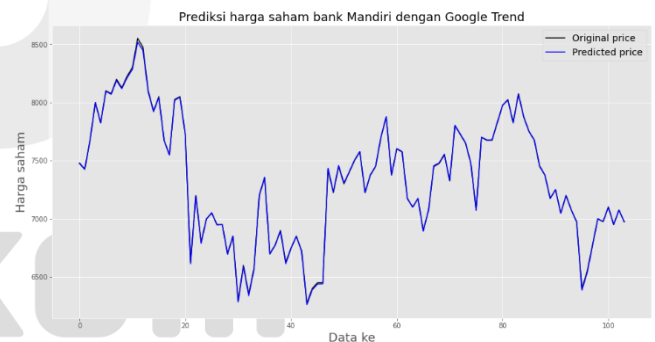
Gambar 4. 1 Pengujian neuron *hidden layer*

Pada Gambar 4.1 pengujian terhadap jumlah neuron pada *hidden layer* dalam model ANN, dengan jumlah neuron dari 5 sampai 50. Pengujian ini membandingkan hasil prediksi tanpa *Google trends* dan prediksi dengan *Google trends*. Hasil prediksi menunjukkan bahwa *hidden layer* dengan neuron 50 memiliki nilai yang optimal berdasarkan prediksi tanpa *Google trends* dan prediksi dengan *Google trends*, maka pada penelitian ini menggunakan *hidden layer* dengan neuron 50. Pada pengujian ini prediksi tanpa *Google trends* menggunakan data input harga saham pada lag 1 dan data harga saham aktual. Sedangkan prediksi dengan *Google trends* menggunakan data input harga saham pada lag 1, data harga saham aktual, dan data *Google trends* pada lag 1.

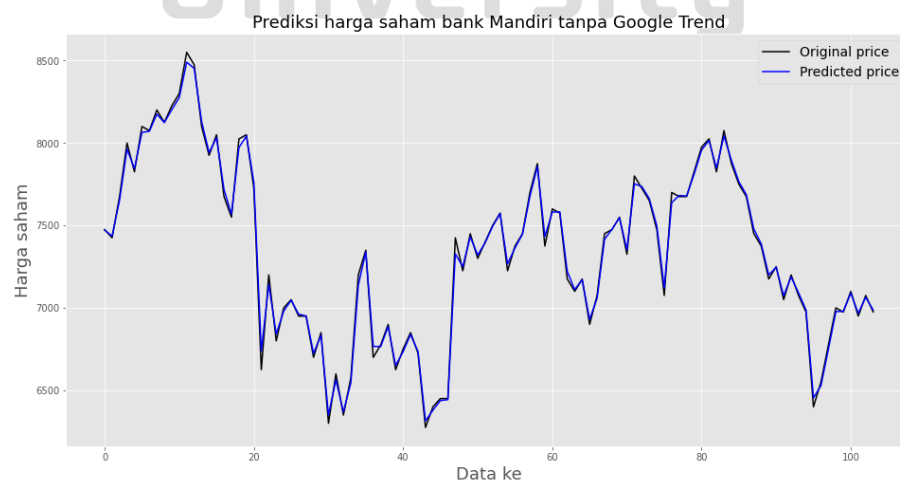
Hasil pengujian prediksi harga saham



Gambar 4. 2 Prediksi dengan *Google trends* lag 4



Gambar 4. 3 Prediksi dengan *Google trends* lag 5



Gambar 4. 4 Prediksi tanpa *Google trends*

Hasil prediksi dengan *Google trends* berdasarkan dari proses *Granger* yang menunjukkan bahwa pada bank Mandiri dan kata kunci *Google trends* “Bank Mandiri: (Indonesia)” memiliki hubungan *Granger* dengan tingkat signifikan 10%. Hasil prediksi ini menghitung nilai prediksi satu minggu kedepan berdasarkan harga *close* dan jumlah data, untuk memprediksi data yang ditentukan pada penelitian ini menggunakan data sebelumnya se pada garis berwarna biru adalah hasil prediksi dan garis berwarna hitam adalah data *original*. Pada Gambar 4.2 memiliki nilai RMSE: 19.0244, dan Gambar 4.3 memiliki nilai RMSE: 19.7397 hasil menunjukkan nilai yang sama untuk kedua data *Google trends* yang di uji. Pada Gambar 4.4 adalah hasil prediksi tanpa *Google trends* memiliki nilai RMSE: 27.0604.

Table 4 Hasil prediksi harga saham satu minggu kedepan berdasarkan nilai RMSE

	BCA	BRI	Mandiri
Prediksi tanpa <i>Google trends</i>	197.8571	176.2167	27.0604
Prediksi dengan <i>Google trends</i>	-	-	19.0244

Pada Tabel 4 menunjukkan hasil prediksi tanpa *Google trends* dan prediksi dengan *Google trends*. Hasil prediksi dengan *Google trends* terdapat nilai kosong atau tidak diprediksi karena pada bank tersebut tidak memiliki dominan terhadap data *Google trends*, maka untuk prediksi dengan *Google trends* hanya bank Mandiri yang memiliki dominan terhadap data *Google trends* dan hasil menunjukkan bahwa prediksi dengan *Google trends* memiliki nilai RMSE lebih kecil dibandingkan prediksi tanpa *Google trends*.

5. Kesimpulan

1. Berdasarkan hasil *Cointegration test* dan *Granger causality analysis* dengan tingkat signifikan 10%, hanya kata kunci “Bank Mandiri: (Indonesia)” yang berpotensi mempengaruhi harga saham bank Mandiri.
2. Prediksi harga saham Bank Mandiri dengan tambahan prediktor data *Google trends* memberikan hasil yang sedikit lebih baik dibandingkan prediksi tanpa *Google trends*.

Daftar Pustaka

- [1] M. Qiu and Y. Song, “Predicting the direction of stock market index movement using an optimized artificial neural network model,” *PLoS One*, vol. 11, no. 5, pp. 1–11, 2016, doi: 10.1371/journal.pone.0155133.
- [2] J. Zahedi and M. M. Rounaghi, “Application of Artificial Neural Network Models and Principal Component Analysis Method in Predicting Stock Prices on Tehran Stock Exchange,” *Physica A*, 2015, doi: 10.1016/j.physa.2015.06.033.
- [3] Y. Z. Yu Lean, Zhao Yaqing, Tang Ling, “Oil Consumption Prediction with Google Trends.pdf.”
- [4] H. Hu, L. Tang, S. Zhang, and H. Wang, “Predicting the direction of stock markets using optimized neural networks with Google Trends,” *Neurocomputing*, vol. 285, pp. 188–195, 2018, doi: 10.1016/j.neucom.2018.01.038.
- [5] Y. Kara, M. Acar, and Ö. Kaan, “Expert Systems with Applications Predicting direction of stock price index movement using artificial neural networks and support vector machines : The sample of the Istanbul Stock Exchange,” *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5311–5319, 2011, doi: 10.1016/j.eswa.2010.10.027.
- [6] T. Preis, H. S. Moat, and H. Eugene Stanley, “Quantifying trading behavior in financial markets using google trends,” *Sci. Rep.*, vol. 3, pp. 1–6, 2013, doi: 10.1038/srep01684.
- [7] J. Guo and Q. Ji, “How does market concern derived from the Internet affect oil prices ?,” *Appl. Energy*, vol. 112, pp. 1536–1543, 2013, doi: 10.1016/j.apenergy.2013.03.027.
- [8] M. Kuhn and K. Johnson, *Applied Predictive Modeling*.
- [9] Juanda & Junaidi, “Ekonomika Deret Waktu Teori dan Aplikasi,” 2014.
- [10] J. L. Dods and D. E. A. Giles, “Alternative strategies for ‘augmenting’ the dickey-fuller test: Size-robustness in the face of pre-testing,” *J. Stat. Comput. Simul.*, vol. 53, no. 3–4, pp. 243–258, 1995, doi: 10.1080/00949659508811709.
- [11] C. Hiemstra and J. D. Jones, “Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation,” vol. XLIX, no. 5, 1994.
- [12] N. D. Lewis, *Neural networks for time series forecasting with R*. 2017.
- [13] P. J. P. P.G. Howlett, *Advances in Industrial Control*. 2006.
- [14] A. Goksu, “Forecasting Euro and Turkish Lira Exchange Rates with Artificial Neural Networks (ANN),” vol. 4, no. 4, pp. 307–316, 2014, doi: 10.6007/IJARAFMS/v4-i4/1361.
- [15] F. S. Hillier and S. Editor, *FORECASTING WITH ARTIFICIAL NEURAL*.

- [16] Q. Shao and R. Lund, "Computation and characterization of autocorrelations and partial autocorrelations in periodic arma models," *J. Time Ser. Anal.*, vol. 25, no. 3, pp. 359–372, 2004, doi: 10.1111/j.1467-9892.2004.00356.x.

