Sistem Prediksi Harga Saham LQ45 Dengan Random Forest Classifier

Kevin Valiant[1], Yuan Lukito[2], R. Gunawan Santosa[3]
Program Studi Informatika, Universitas Kristen Duta Wacana
Jl. Dr.. Wahidin Sudirohusodo nomor 5-25, Kota Yogyakarta
[1]kevin.valiant@ti.ukdw.ac.id
[2]yuanlukito@ti.ukdw.ac.id
[3]gunawan@staff.ukdw.ac.id

Abstract*— The purpose of this research is to measure the performance of Random Forest method in predicting change of LQ45 stocks' prices. The dataset is the daily stock summary of company registered in the August 2018-January 2019 version of LQ45 index from July 1st, 2015 to December 31st, 2018. The technical indicators used in this research are On-Balance Volume, Chaikin Oscillator, Moving Average Convergence/Divergence, and Bollinger Bands. Models then was made for each stock ticker, time range, and feature type. The time range are 1-day, 5-day, and 20-day. Feature type are plain which use all the values in daily stock summary and the technical indicator, grouped which the feature are the condition of technical indicator values compared to the day before (up, still, down) and the x-axis (positive, zero, negative), and onehot which the feature are one-hot encoded grouped feature. The formed models then are used to predict the change of stock prices with up, still, and down as possible results. The accuracy is then calculated using confusion matrix. The result of the test on training set shows high accuracy, where plain feature type with time range of 5-day and 20-day even show a perfect accuracy of 100%. The result of the test on test set shows declining performance, but the plain feature type still show the best performance where 3 stock tickers range having accuracy bigger than 60% for 1-day time, 6 stock tickers for 5-day time range, and 14 stock tickers for 20-day time range, 9 of which having accuracy bigger than 70%, and ANTM having accuracy of 80.6%.

Intisari—Data yang digunakan adalah ringkasan saham harian perusahaan yang terdaftar pada indeks LQ45 versi Agustus 2018 – Januari 2019 mulai tanggal 1 Juli 2015 hingga 31 Desember 2018. *Technical indicator* yang digunakan dalam penelitian ini adalah On-Balance Volume, Chaikin Oscillator, Moving Average Convergence/Divergence, dan Bollinger Bands. Data tersebut kemudian dibentuk modelnya untuk setiap kode saham, rentang waktu, dan tipe fitur. Rentang waktu prediksi yang digunakan adalah 1 hari, 5 hari, dan 20 hari. Tipe fitur yang digunakan untuk membentuk model adalah *plain* yang menggunakan seluruh nilai ringkasan saham harian dan *technical indicator*-nya, *grouped* yang fiturnya merupakan kondisi *technical indicator* terhadap hari sebelumnya (naik, tetap, dan turun) dan nilainya terhadap garis nol (positif, nol, negatif), serta *onehot* yang fiturnya merupakan hasil *one-hot encoding* terhadap fitur *grouped*. Model yang dibentuk kemudian digunakan untuk memprediksi perubahan harga saham dengan kemungkinan nilai naik, tetap, atau turun. Nilai akurasi dihitung menggunakan *confusion matrix*. Hasil pengujian terhadap data latih menunjukkan nilai yang sangat baik, dimana tipe fitur plain dengan rentang waktu 5 dan 20 hari mencapai 100%. Hasil pengujian terhadap data uji menunjukkan penurunan dibanding data latih, namun tipe fitur *plain* tetap menunjukkan performa paling baik dimana terdapat tiga kode saham dengan akurasi lebih besar dari 60% untuk rentang waktu satu hari, enam kode saham untuk rentang waktu lima hari, dan empat belas kode saham untuk rentang waktu dua puluh hari, sembilan di antaranya di atas 70%, dengan kode ANTM mencapai akurasi sebesar 80,6%.

***Kata Kunci*— saham, LQ45, technical analysis, random forest**

I. PENDAHULUAN

Salah satu teknologi yang sedang gencar dipelajari adalah *data mining*, dimana aplikasinya sangat luas mulai dari *fraud detection, credit scoring*, hingga *sentiment analysis. Data mining* merupakan proses mengubah data mentah menjadi informasi yang berguna [1]. Proses *data mining* berjalan secara semiotomatis dan digunakan untuk membantu pengambilan keputusan.

Dalam jual beli saham, umumnya pelaku dibagi menjadi dua jenis yaitu *technical analyst* dan *fundamental analyst*. *Fundamental analyst* memperkirakan harga saham berdasarkan nilai nyata dari sebuah perusahaan terdaftar, sedangkan *technical analyst* memprediksi harga saham berdasarkan pola dari ringkasan saham yang menunjukkan sentimen pasar.

Berdasarkan data dari Bursa Efek Indonesia (BEI) pada 22 April 2019, tercatat terdapat 632 perusahaan terdaftar dan saham beredar sebanyak 15.705.859.354 lembar dengan rata-rata frekuensi transaksi per hari sebanyak 389.865 kali. Jumlah data yang sangat banyak tentunya menyulitkan apabila diproses secara manual, sehingga sistem yang memanfaatkan data mining dapat membantu proses analisis dan meningkatkan akurasi prediksi harga saham.

Terdapat banyak faktor yang mempengaruhi perubahan nilai saham, diantaranya kondisi perusahaan, kondisi ekonomi negara atau dunia secara keseluruhan, hingga sentimen pasar. Berbagai variasi metode analisis dapat digunakan untuk memprediksi perubahan nilai saham, namun beberapa penelitian menunjukkan bahwa lebih menguntungkan untuk memprediksi arah perubahan harga saham dibanding nilai saham [1].

Berbagai metode *machine learning* sudah banyak digunakan untuk memprediksi perubahan harga saham, mulai dari *neural network* hingga Support Vector Machine. Penelitian oleh Hegazy [2] menggabungkan Particle Swarm Optimization (PSO) dan Least Square Support Vector

Machine (LS-SVM) untuk memprediksi harga saham harian dan membandingkan hasilnya terhadap single LS-SVM dan Artificial Neural Network (ANN) yang digabungkan dengan algoritma Levenberg-Marquardt (LM). Penelitian tersebut menunjukan bahwa PSO-LS-SVM dapat mengatasi *overfitting* yang terjadi pada ANN-LM, dengan PSO-LS-SVM menunjukkan nilai *error* terendah dan ANN-LM dengan nilai *error* tertinggi. Penelitian lain oleh Madge [3] menggunakan SVM untuk memprediksi arah perubahan harga saham. Penelitian tersebut menemukan bahwa data historis lebih berpengaruh dalam memprediksi perubahan jangka panjang dan perubahan jangka pendek cenderung sulit diprediksi.

Metode yang akan digunakan untuk membangun sistem dalam penelitian ini adalah Random Forest Classifier. Sistem akan menerima data ringkasan saham harian dan menghasilkan label prediksi harga saham yang bernilai naik, turun, atau tetap dan diukur performanya berdasarkan akurasi.

II. Landasan Teori

Berikut merupakan beberapa landasan teori yang digunakan dalam penelitian ini:

a. Saham

Saham merupakan bukti kepemilikan atas sebuah perusahaan. Seorang pemilik saham memiliki hak atas aset dan pendapatan dari perusahaan yang bersangkutan. Dua alasan utama seseorang membeli saham dari suatu perusahaan yaitu untuk mendapatkan bagian keuntungan yang diperoleh perusahaan atau disebut juga dividen dan menjual kembali saham dengan harga yang lebih tinggi.

Saham yang dapat dijual dan dibeli masyarakat umum diperdagangkan di pasar modal. Bursa Efek Indonesia (BEI) merupakan pasar modal Indonesia dimana saham-saham perusahaan yang sudah terdaftar diperdagangkan. Sebuah perusahaan memperdagangkan sahamnya untuk mendapatkan modal tambahan. Agar sahamnya dapat diperdagangkan di pasar saham, sebuah perusahaan terlebih dahulu harus melalui proses Initial Public Offering (IPO).

Beberapa contoh perusahaan yang sudah menawarkan sahamnya di BEI diantaranya PT Bank Central Asia, Tbk dengan kode BBCA, PT Waskita Karya (Persero), Tbk dengan kode WSKT, PT MAP Aktif Adiperkasa, Tbk, dengan kode MAPA, dan PT Indofood Sukses Makmur, Tbk dengan kode INDF. Perusahaan terdaftar juga terbagi ke dalam sektor-sektor tertentu, diantaranya TRADE, PROPERTY dan CONSUMER. Beberapa perusahaan juga terdaftar dalam indeks tertentu, diantaranya LQ45, KOMPAS100, dan IDX30. Indeks merupakan gabungan beberapa saham yang menjadi indikator performa pasar secara umum.

b. Technical Analysis

Technical analysis adalah salah satu metode yang umum digunakan di pasar modal selain *fundamental analysis*. *Technical analysis* menggunakan data historis pasar seperti harga dan volume untuk memprediksi *return* [2]. Pasar modal dianggap bersifat tidak stabil dan tidak efisien, sehingga *technical analyst* mencoba untuk memprediksi harga saham dengan membaca sentimen pasar, berlawanan dengan *fundamental analyst* yang mencoba memprediksi harga saham dengan menghitung nilai intrinsik saham tersebut. Sering dikatakan bahwa perbandingan antara *fundamental analysis* dan *technical analysis* tidaklah berbeda dengan perbandingan antara astronomi dan astrologi [3].

Nilai-nilai yang menjadi dasar indikator dalam *technical analysis* adalah *open* yang menunjukan harga pembukaan, *close* merupakan harga penutupan, *high* menunjukkan harga tertinggi, *low* merupakan harga terendah, dan *volume* menunjukkan jumlah lembar saham yang diperdagangkan pada hari tersebut.

Nilai-nilai tersebut tidak digunakan begitu saja, namun dengan perhitungan tertentu untuk membentuk yang disebut dengan technical indicator. Beberapa contoh *technical indicator* yang sering digunakan adalah Moving Average Convergence/Divergence (MACD) dan On-Balance Volume (OBV).

c. Exponential Moving Average (EMA)

EMA merupakan salah satu *technical indicator* yang banyak digunakan. EMA merupakan pengembangan dari Simple Moving Average (SMA) yang dinilai kurang menggambarkan kondisi terkini karena tidak digunakannya pembobotan. EMA juga menjadi dasar perhitungan untuk beberapa *technical indicator* lainnya, seperti MACD dan Chaikin Oscillator. Rumus untuk mendapatkan EMA adalah:

$$Initial\ \text{SMA} = \frac{A_1+A_2+\cdots+A_{n-1}+A_n}{n} \quad (1)$$

Keterangan:
- A_i : Nilai A pada hari ke-i
- n : jumlah hari

Persamaan (1) digunakan untuk mendapatkan nilai SMA awal. Nilai yang digunakan sebagai dasar dari SMA bermacam-macam, mulai dari harga penutupan, pembukaan, hingga *technical indicator* lainnya.

$$Multiplier = \frac{2}{n+1} \quad (2)$$

Keterangan:
- n : jumlah hari

Persamaan (2) digunakan untuk mendapatkan nilai *multiplier* yang akan digunakan untuk menghitung nilai EMA.

$$EMA_i = (C_i - EMA_{i-1}) * multiplier + EMA_{i-1} \quad (3)$$

Keterangan:
- C_i : Harga penutupan pada hari ke-i
- EMA_{i-1} : Nilai EMA satu hari sebelum hari ke-i

Persamaan (3) digunakan untuk mendapatkan nilai EMA. Nilai awal EMA pada hari ke-n adalah nilai SMA pada hari yang sama.

d. On-Balance Volume (OBV)

OBV adalah indikator *technical analysis* yang nilainya dipengaruhi oleh volume dan harga saham. Nilai OBV yang tinggi mengindikasikan sentimen pasar yang baik, serta OBV

JUTEI Edisi Volume.3 No.2 Oktober 2019
ISSN 2579-3675, e-ISSN 2579-5538
DOI 10.21460/jutei.2019.32.187

dapat digunakan untuk memprediksi *market reversal* [4]. Rumus untuk mendapatkan nilai OBV ditunjukkan pada persamaan 4.

$$OBV_i = \begin{cases} OBV_{i-1} + V_i, C_i > C_{i-1} \\ OBV_{i-1} - V_i, C_i < C_{i-1} \end{cases} \quad (4)$$

Keterangan:
- OBV_i : Nilai OBV pada hari ke-i
- V_i : Volume pada hari ke-i
- C_i : Harga penutupan pada hari ke-i

Pada saat i bernilai 0, nilai OBV diinisialisasi dengan nilai 0.

Gambar 1. Grafik yang menunjukkan harga, volume dan OBV saham Wal-Mart pada November-Desember 2010 (https://school.stockcharts.com/doku.php?id=technical_indicators:on_balance_volume_obv)

Gambar 1 menunjukkan grafik perubahan harga, volume serta nilai OBV dari saham Wal-Mart. Nilai OBV yang naik mengindikasikan sentimen pasar yang baik dan harga saham yang naik, seperti terlihat grafik di awal November. Lalu grafik OBV tampak menurun, menggambarkan harga saham yang melemah seperti yang ditunjukan pada grafik menjelang pertengahan November.

e. Chaikin Oscillator

Chaikin Oscillator adalah indikator *technical analysis* yang nilainya menggambarkan momentum dari Accumulation/Distribution Line (ADL). Nilai dari Chaikin Oscillator berosilasi di atas dan bawah garis nol [5]. Nilai Chaikin Oscillator positif menunjukkan *buying pressure* yang mengindikasikan kenaikan harga saham dan berlaku sebaliknya. Rumus yang digunakan untuk mendapatkan nilai Chaikin Oscillator adalah:

$$MFM_i = \left[\frac{(C_i - L_i) - (H_i - C_i)}{(H_i - L_i)}\right] \quad (5)$$

Keterangan:
- C_i : Harga penutupan pada hari ke-i
- L_i : Harga terendah pada hari ke-i
- H_i : Harga tertinggi pada hari ke-i

Persamaan (5) digunakan untuk mendapatkan nilai Money Flow Multiplier (MFM).

$$MFV = MFM * VP \quad (6)$$

Keterangan:
- VP : Volume selama periode tersebut

Persamaan (6) digunakan untuk mendapatkan nilai Money Flow Volume (MFV).

$$ADL = Previous\ ADL + Current\ Period's\ MFV \quad (7)$$

Persamaan (7) digunakan untuk mendapatkan nilai ADL, yang kemudian dihitung momentumnya untuk mendapatkan nilai Chaikin Oscillator.

$$CO = (N1\_EMA\ of\ ADL - N2\_EMA\ of ADL) \quad (8)$$

Keterangan:
- N1_EMA of ADL = Nilai EMA jangka pendek dari ADL
- N2_EMA of ADL = Nilai EMA jangka panjang dari ADL

Persamaan (8) menunjukkan rumus yang digunakan untuk mendapatkan nilai Chaikin Oscillator, yang merupakan selisih EMA jangka panjang dan jangka pendek ADL.

Gambar 2. Grafik yang menunjukkan harga, volume dan Chaikin Oscillator saham Microsoft pada September-Desember 2010. (https://school.stockcharts.com/doku.php?id=technical_indicators:chaikin_oscillator)

Gambar 2 menunjukkan harga, volume, dan Chaikin Oscillator saham Microsoft. Terlihat bahwa saat nilai Chaikin Oscillator positif di bulan Oktober, harga saham naik. Sebaliknya, nilai Chaikin Oscillator yang negatif menunjukkan sentimen pasar yang buruk sehingga harga

saham turun seperti yang terlihat pada grafik di bulan November.

f. Moving Average Convergence/Divergence (MACD)

MACD merupakan salah satu *technical indicator* yang menunjukkan tren. MACD terdiri dari dua buah garis yang disebut sebagai garis MACD dan garis sinyal, serta sebuah histogram. Untuk mendapatkan nilai garis MACD dan sinyal digunakan perhitungan menggunakan EMA. Adapun rumus dari nilai garis MACD, sinyal dan histogram adalah:

$$MACD\ line\ =\ EMA_{12} - EMA_{26} \qquad (9)$$

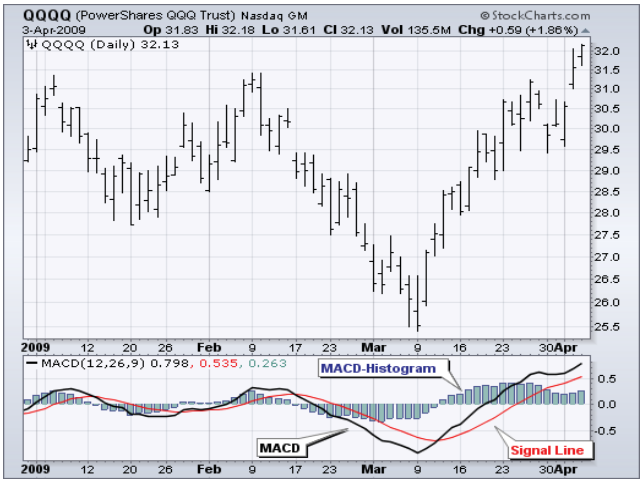
Persamaan (9) digunakan untuk mendapatkan nilai garis MACD, dengan menggunakan EMA dari harga penutupan. Jangka waktu pendek dan panjang dapat diubah, namun 12 dan 26 hari umumnya lebih sering digunakan.

$$Signal\ line\ =\ EMA_9\ of\ MACD\ line \qquad (10)$$

Persamaan (10) digunakan untuk mendapatkan nilai garis sinyal. Jangka waktu EMA dapat diubah, namun umumnya 9 hari lebih sering digunakan.

$$Histogram\ line\ =\ MACD\ line - Signal\ line \qquad (11)$$

Persamaan (11) menunjukkan rumus yang digunakan untuk mendapatkan nilai garis histogram.



Gambar 3. Grafik yang menunjukkan harga, garis MACD, garis sinyal dan histogram Powershares QQQ Trust pada Februari-April 2009. (https://school.stockcharts.com/doku.php?id=technical_indicators:moving_average_convergence_divergence_macd)

Gambar 3 menunjukkan menunjukkan harga, garis MACD, garis sinyal dan histogram Powershares QQQ Trust. Terlihat bahwa pada tanggal 17 Februari, garis MACD memotong garis sinyal ke bawah, yang menunjukkan perubahan pasar menjadi *bearish*. Pada tanggal 11 Maret terlihat garis MACD memotong garis sinyal ke atas, yang menunjukkan perubahan pasar menjadi *bullish*.

g. Bollinger Bands

Bollinger bands merupakan indikator teknikal yang menunjukan *volatility* dari saham. *Bollinger bands* terdiri dari tiga buah garis, yaitu *upper band*, *lower band*, dan *middle band*. Rumus perhitungan nilai *upper band*, *lower band*, dan *middle band* adalah:

$$Upper\ band\ =\ SMA_n + (\sigma\ of\ Close_n * m) \qquad (12)$$

Keterangan:

- SMA_n : Nilai SMA selama n-hari
- σ of $Close_n$: Standar deviasi dari harga penutupan selama n-hari
- m : angka pengali

Persamaan (12) digunakan untuk mendapatkan *upper band*.

$$Lower\ band\ =\ SMA_n - (\sigma\ of\ Close_n * m) \qquad (13)$$

Persamaan (13) digunakan untuk mendapatkan nilai *lower band*.

$$Middle\ band\ =\ SMA_n \qquad (14)$$

Persamaan (14) menunjukkan rumus yang digunakan untuk mendapatkan nilai *middle band*. Nilai n yang umum digunakan adalah 20 hari dan m sebesar 2. Kondisi *overbought* terjadi saat nilai penutupan berada di atas *upper band*. *Overbought* merupakan kondisi di mana harga sudah di atas ambang perkiraan harga sebenarnya. Sebaliknya, kondisi *oversold* terjadi saat nilai penutupan berada di bawah *lower band*. *Oversold* merupakan kondisi di mana harga sudah di bawah ambang perkiraan harga sebenarnya. Kondisi *oversold* dan *overbought* tidak selalu menunjukkan *market reversal* dan perlu digabungkan dengan *technical indicator* lainnya untuk mendapatkan perkiraan pasar yang lebih baik.

Gambar 4. Grafik yang menunjukkan harga, Bollinger Bands serta standar deviasi dari SPDR S&P 500 ETF Trust pada April-Juni 2009. (https://school.stockcharts.com/doku.php?id=technical_indicators:bollinger_bands)

Gambar 4 menunjukkan harga, Bollinger Bands serta standar deviasi dari SPDR S&P 500 ETF Trust. Terlihat bahwa pada awal Mei dan awal Juni 2009, harga melewati upper band, yang menunjukkan kondisi overbought.

JUTEI Edisi Volume.3 No.2 Oktober 2019
ISSN 2579-3675, e-ISSN 2579-5538
DOI 10.21460/jutei.2019.32.187

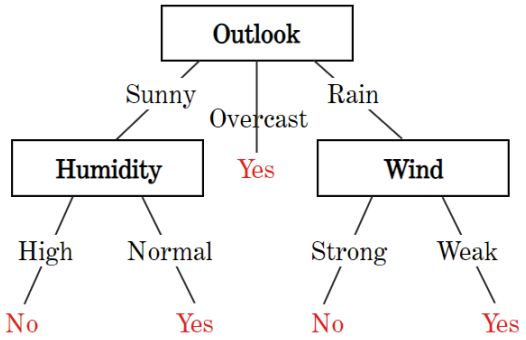
h. Data Mining

Data mining merupakan proses menambang atau mengekstrak *knowledge* dari data yang sangat banyak [6]. Aplikasi *data mining* umumnya berjalan secara semiotomatis untuk kebutuhan pemilihan keputusan [7]. Data sendiri memiliki berbagai macam bentuk, mulai dari teks, gambar, hingga suara.

i. Decision Tree

Decision tree merupakan salah satu metode klasifikasi yang menghasilkan *rule* untuk memprediksi data. Sebuah *decision tree* berbentuk seperti *flowchart* yang terdiri dari *internal node* yang merepresentasikan keputusan terhadap sebuah atribut, *leaf node* yang menunjukkan kelas/label, dan *branch* yang menggambarkan kemungkinan hasil dari tes [8]. Tahap pembentukan *decision tree* adalah:

1. Untuk masing-masing nilai dari setiap atribut, pilih kelas yang paling sering muncul dan hitung *error*-nya.
2. Hitung *error* masing-masing atribut dengan menjumlahkan *error* dari setiap nilainya (langkah 1).
3. Pilih atribut dengan *error* terkecil sebagai *internal node*, dengan nilai atribut tersebut sebagai *branch*-nya.
4. Apabila *error* dari nilai atribut bernilai 0, maka *branch*-nya dilanjutkan dengan *leaf node* yang berisikan kelas yang sesuai.
5. Selain itu ulangi dari langkah pertama dengan mengabaikan atribut yang sudah menjadi *internal node* di jalur yang sama. Hal ini terus dilakukan hingga seluruh jalur berujung di *leaf node*.



Gambar 5. Contoh *decision tree.*
(https://www.thelearningmachine.ai/tree-id3)

Gambar 5 menunjukkan decision tree yang terbentuk. Sebagai contoh penggunaannya, terdapat kondisi outlook sunny, temperature hot, humidity high, dan windy true. Internal node pertama adalah outlook dan branch yang dipilih adalah sunny, sehingga internal node kedua adalah humidity. Branch berikutnya adalah high dan menghasilkan leaf node bernilai N sehingga hasil prediksi dari data uji tersebut adalah N.

j. Random Forest

Random Forest merupakan sebuah *ensemble method* yang melakukan prediksi dengan mencari nilai rata-rata dari hasil prediksi beberapa model independen [9]. Metode ini merupakan pengembangan dari *decision tree* yang cenderung rawan terhadap *overfit* sehingga sulit digunakan apabila data uji berisi data yang belum pernah ditemui. Sebuah Random Forest adalah *classifier* yang terdiri dari sekumpulan *classifier* berbentuk pohon {h(x,Θk),k=1, …} dimana {Θk} yang merupakan vektor-vektor acak yang terdistribusi secara independen dan setiap pohon memberikan *vote* untuk memilih kelas yang paling populer berdasarkan input x. Performa yang baik dari Random Forest berhubungan dengan kualitas yang baik dari masing-masing pohon dan korelasi antar pohon, dimana korelasi antar pohon didefinisikan sebagai korelasi umum dari prediksi terhadap *out-of-bag* (OOB) *samples*. OOB *samples* merupakan sekumpulan observasi yang tidak digunakan untuk membangun pohon yang sedang dibentuk, digunakan untuk memperkirakan error dari prediksi dan mengevaluasi pentingnya variabel [10]. Tahap pembentukan Random Forest adalah:

1. Pilih jumlah pohon (n) yang akan dibentuk (semakin besar umumnya semakin baik, namun memperlambat proses prediksi dan nilai akurasi mungkin tidak berubah banyak).
2. Bentuk n buah *decision tree* berdasarkan dengan pembagian data (umumnya 2/3 untuk data latih dan 1/3 untuk data uji), dimana data dipilih secara acak.
3. Masing-masing *decision tree* memberikan *vote* untuk memilih kelas yang paling sesuai berdasarkan aturannya masing-masing.

k. Akurasi

Akurasi merupakan salah satu indikator performa klasifikasi yang menunjukkan ketepatan hasil prediksi dibandingkan seluruh data uji. Untuk menunjukkan jumlah hasil prediksi dan aktual, umumnya digunakan confusion matrix berukuran n x n, dimana n adalah jumlah kelas/label [11].

TABEL 1
CONTOH CONFUSION MATRIX DENGAN N=2

	Predicted Negative	**Predicted Positive**
Actual Negative	a	b
Actual Positive	c	d

Tabel 1 merupakan contoh *confusion matrix* berukuran 2 x 2, dimana a menunjukkan nilai *true negative*, b menunjukkan nilai *false positive*, c menunjukkan nilai *false negative*, dan d menunjukkan nilai *true positive*. Nilai akurasi didapatkan dari nilai *true negative* + *true positive* dibagi total seluruhnya.

III. METODOLOGI PENELITIAN

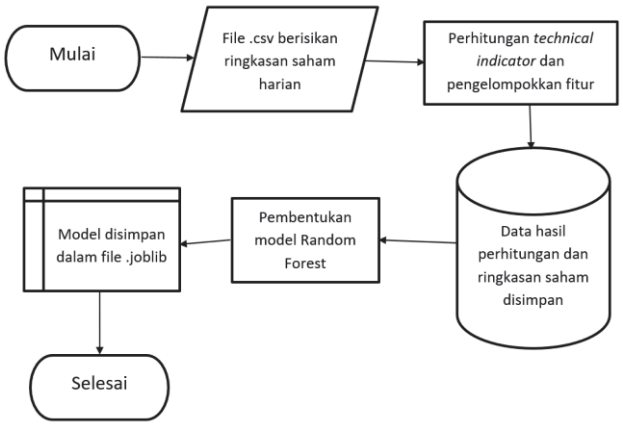
a. Sumber Data

Data yang berupa ringkasan saham harian perusahaan-perusahaan yang terdaftar pada indeks LQ45 versi Agustus 2018-Januari 2019 diambil dari data Phillip Sekuritas Indonesia. Emiten Waskita Beton Precast Tbk. (WSBP), Barito Pacific Tbk. (BRPT), Indika Energy Tbk. (INDY), dan Media Nusantara Citra Tbk. (MNCN) tidak termasuk dikarenakan data yang kurang atau juga pernah *suspended*. Data yang digunakan adalah data sejak tanggal 1 Juli 2015 hingga 30 Juni 2018 untuk data latih dan 1 Juli 2018 hingga 31 Desember 2018.

b. Alur Kerja Sistem

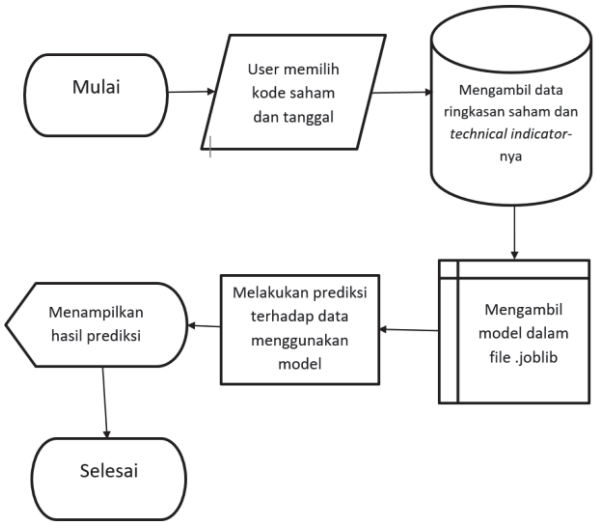
Secara garis besar, alur kerja sistem dibagi menjadi dua bagian yaitu proses latih dan proses prediksi. Gambar 6 menunjukkan proses latih. Data dibaca dari file .csv yang berisikan ringkasan saham harian, kemudian dihitung nilai-nilai *technical indicator*-nya kemudian dikelompokkan fitur-fiturnya. Data tersebut kemudian disimpan ke dalam basis data. Data yang sama kemudian dibagi menjadi data latih dan data uji dan digunakan untuk membentuk model Random Forest untuk setiap kode saham, rentang prediksi dan tipe fitur. Model yang terbentuk kemudian disimpan dalam bentuk file .joblib.

Tiga tipe fitur yang dibentuk adalah plain, grouped dan onehot. Tipe fitur plain didapatkan dari nilai open, high, low, close, volume, OBV, Chaikin Oscillator, MACD, signal, histogram, upper band dan lower band. Sebagai contoh, kode saham EXCL pada tanggal 6 Juni 2016 memiliki fitur dengan nilai 3.480 (open), 3.680 (high), 3.470 (low), 3.620 (close), 184.506 (volume), 1.722.517 (OBV), -64210.9814991219 (Chaikin Oscillator), -14.8601285160294 (MACD), -63.4153032939874 (signal), 48.5551747779581 (histogram), 3646.13114269311 (upper band) dan 3123.86885730689 (lower band). Tipe fitur grouped merupakan nilai hasil perubahan dari nilai technical indicator yang digunakan pada tipe fitur plain menjadi ordinal. Fitur-fitur yang digunakan adalah OBV_COMPARISON, CO_COMPARISON, MACD_COMPARISON, SIGNAL_COMPARISON, HISTOGRAM_COMPARISON yang merupakan perbandingan nilai fitur tersebut terhadap hari sebelumnya yang bernilai 1 (naik), 0 (tetap) dan -1 (turun) serta OBV_POSITION, CO_POSITION, MACD_POSITION, SIGNAL_POSITION, HISTOGRAM_POSITION yang merupakan perbandingan nilai fitur tersebut terhadap titik nol yang bernilai 1 (positif), 0 (nol), dan -1 (negatif) dan juga BB_CONDITION yang merupakan perbandingan nilai close terhadap upper band dan lower band yang bernilai 1 (overbought), 0 (normal), dan -1 (oversold). Sebagai contoh kode saham MEDC pada tanggal 20 September 2017 bernilai 1 (OBV_COMPARISON), 1 (OBV_POSITION), 1 (CO_COMPARISON), 1 (CO_POSITION), 1 (MACD_COMPARISON), 1 (MACD_POSITION), -1 (SIGNAL_COMPARISON), 1 (SIGNAL_POSITION), 1 (HISTOGRAM_COMPARISON), -1 (HISTOGRAM_POSITION), 0 (BB_CONDITION). Tipe fitur onehot merupakan hasil one-hot encoding untuk tiap-tiap fitur grouped. Tiap fitur pada grouped dipecah menjadi tiga fitur yaitu naik, turun dan tetap untuk fitur dengan akhiran COMPARISON lalu positif, negatif dan nol untuk fitur dengan akhiran POSITION serta overbought, oversold, dan normal untuk BB_CONDITION. Nilai dari fitur-fitur onehot adalah Boolean dengan 1 yang berarti True dan 0 yang berarti False.



Gambar 6. Alur proses latih

Gambar 7 menunjukkan alur proses prediksi. *User* memberikan input berupa kode saham dan tanggal, lalu sistem mengambil ringkasan saham beserta *technical indicator* dari basis data. Sistem juga mengambil model yang sebelumnya disimpan dalam bentuk file .joblib, kemudian model tersebut digunakan untuk memprediksi ringkasan saham dari kode saham dan tanggal pilihan *user*. Hasil prediksi kemudian ditampilkan pada antar muka *web*.



Gambar 7. Alur proses prediksi

c. Rancangan Basis Data

Data yang digunakan dalam proses latih dan prediksi disimpan ke dalam beberapa tabel. Tabel prediksi_stocks menyimpan informasi ringkasan harga saham, kondisi perubahan yang sebenarnya, dan hasil prediksi berdasarkan tiap rentang waktu dan tipe fitur untuk setiap kode saham pada setiap harinya.

TABEL 2
DESKRIPSI TABEL PREDIKSI_STOCKS

No.	Nama Kolom	Tipe Data	Keterangan
1	ID	Integer	Auto Increment
2	KODE_SAHAM	Varchar	Unique
3	TANGGAL	Datetime	Unique
4	OPEN	Integer	Not Null
5	HIGH	Integer	Not Null
6	LOW	Integer	Not Null
7	CLOSE	Integer	Not Null
8	VOLUME	Integer	Not Null

JUTEI Edisi Volume.3 No.2 Oktober 2019
ISSN 2579-3675, e-ISSN 2579-5538
DOI 10.21460/jutei.2019.32.187

9	AFTER1	Integer	
10	AFTER5	Integer	
11	AFTER20	Integer	
12	PLAIN1	Integer	
13	PLAIN5	Integer	
14	PLAIN20	Integer	
15	GROUPED1	Integer	
16	GROUPED5	Integer	
17	GROUPED20	Integer	
18	ONEHOT1	Integer	
19	ONEHOT5	Integer	
20	ONEHOT20	Integer	

Tabel 2 menunjukkan tabel prediksi_stocks yang menyimpan data ringkasan saham dan *technical indicators*. Kolom KODE_SAHAM dan TANGGAL bersifat unik dimana pada satu tanggal tidak ada lebih dari satu baris dengan kode saham yang sama dan sebaliknya. Kolom OPEN, HIGH, LOW, CLOSE, dan VOLUME menyimpan nilai-nilai dari ringkasan saham untuk setiap saham pada tanggal tertentu. Nilai AFTER1, AFTER5, dan AFTER20 menunjukkan kondisi perubahan harga saham yang sebenarnya setelah 1, 5 dan 20 hari kerja. Nilai PLAIN1, PLAIN5, dan PLAIN20 menunjukkan nilai prediksi menggunakan angka nilai ringkasan saham dan nilai *technical indicator*-nya. Nilai GROUPED1, GROUPED5, dan GROUPED20 menunjukkan nilai prediksi menggunakan fitur yang dibentuk dari kondisi perubahan nilai *technical indicators*. Nilai ONEHOT1, ONEHOT5, ONEHOT20 menunjukkan nilai prediksi menggunakan fitur hasil *one-hot encoding* dari fitur *grouped*. Nilai kolom AFTER, PLAIN, GROUPED, dan ONEHOT bernilai -1 yang berarti turun, 0 yang berarti tetap, dan 1 yang berarti naik.

TABEL 3
DESKRIPSI TABEL PREDIKSI_PLAINSTOCKS

No.	Nama Kolom	Tipe Data	Keterangan
1	ID	Integer	Auto Increment
2	KODE_SAHAM	Varchar	Unique
3	TANGGAL	Datetime	Unique
4	OPEN	Integer	Not Null
5	HIGH	Integer	Not Null
6	LOW	Integer	Not Null
7	CLOSE	Integer	Not Null
8	VOLUME	Integer	Not Null
9	OBV	Float	Not Null
10	CO	Float	Not Null
11	MACD	Float	Not Null
12	SIGNAL	Float	Not Null
13	HISTOGRAM	Float	Not Null
14	BOLLINGER_HIGH	Float	Not Null
15	BOLLINGER_LOW	Float	Not Null

Tabel 3 menunjukkan tabel prediksi_plainstocks yang menyimpan data ringkasan saham dan *technical indicators* untuk setiap kode saham pada setiap harinya.

TABEL 4
DESKRIPSI TABEL PREDIKSI_GROUPEDSTOCKS

No.	Nama Kolom	Tipe Data	Keterangan
1	ID	Integer	Auto Increment
2	KODE_SAHAM	Varchar	Unique
3	TANGGAL	Datetime	Unique
4	OBV_COMPARISON	Integer	
5	OBV_POSITION	Integer	
6	CO_COMPARISON	Integer	
7	CO_POSITION	Integer	
8	MACD_COMPARISON	Integer	
9	MACD_POSITION	Integer	
10	SIGNAL_COMPARISON	Integer	
11	SIGNAL_POSITION	Integer	
12	HISTOGRAM_ COMPARISON	Integer	
13	HISTOGRAM_POSITION	Integer	
14	BB_CONDITION	Integer	

Tabel 4 menunjukkan tabel prediksi_groupedstocks yang berisi fitur fitur yang telah dibentuk dari kondisi nilai ringkasan dan *technical indicators*. Nilai Comparison menunjukkan kondisi perbandingan nilai *technical indicators* terhadap hari sebelumnya, yang bernilai -1 (turun), 0 (tetap), dan 1 (naik). Nilai Condition kecuali BB_CONDITION menunjukkan kondisi nilai *technical indicator* terhadap garis nol, yang bernilai -1 (negatif), 0, dan 1 (positif). BB_CONDITION bernilai -1 (*oversold*), 0 (normal), dan 1 (*overbought*).

TABEL 5
DESKRIPSI TABEL PREDIKSI_ONEHOTSTOCKS

No.	Nama Kolom	Tipe Data	Keterangan
1	ID	Integer	Auto Increment
2	KODE_SAHAM	Varchar	Unique
3	TANGGAL	Datetime	Unique
4	OBV_COMPARISON_ NAIK	Integer	
5	OBV_COMPARISON_ TETAP	Integer	
6	OBV_COMPARISON_ TURUN	Integer	
7	OBV_POSITION_POSITIF	Integer	
8	OBV_POSITION_NOL	Integer	
9	OBV_POSITION_NEGATIF	Integer	
10	CO_COMPARISON_NAIK	Integer	
11	CO_COMPARISON_TETAP	Integer	
12	CO_COMPARISON_ TURUN	Integer	
13	CO_POSITION_POSITIF	Integer	
14	CO_POSITION_NOL	Integer	
15	CO_POSITION_NEGATIF	Integer	
16	MACD_COMPARISON_ NAIK	Integer	
17	MACD_COMPARISON_ TETAP	Integer	
18	MACD_COMPARISON_ TURUN	Integer	
19	MACD_POSITION_ POSITIF	Integer	
20	MACD_POSITION_NOL	Integer	
21	MACD_POSITION_NEGAT IF	Integer	
22	SIGNAL_COMPARISON_ NAIK	Integer	
23	SIGNAL_COMPARISON_ TETAP	Integer	
24	SIGNAL_COMPARISON_ TURUN	Integer	
25	SIGNAL_POSITION_ POSITIF	Integer	
26	SIGNAL_POSITION_NOL	Integer	

27	SIGNAL_POSITION_ NEGATIF	Integer	
28	HISTOGRAM_ COMPARISON_NAIK	Integer	
29	HISTOGRAM_ COMPARISON_ TETAP	Integer	
30	HISTOGRAM_ COMPARISON_TURUN	Integer	
31	HISTOGRAM_POSITION_ POSITIF	Integer	
32	HISTOGRAM_POSITION_ NOL	Integer	
33	HISTOGRAM_POSITION_ NEGATIF	Integer	
34	BB_CONDITION_ OVERBOUGHT	Integer	
35	BB_CONDITION_ OVERSOLD	Integer	
36	BB_CONDITION_ NORMAL	Integer	

Tabel 5 menunjukkan tabel prediksi_onehotstocks yang berisikan nilai hasil *one-hot encoding* terhadap fitur-fitur *grouped stocks*. Nilai 1 menunjukkan *True* dan 0 menunjukkan *False*.

d. Rancangan Pengujian

Model yang sudah dihasilkan kemudian akan diujikan pada data terbaru. Data yang diujikan diambil merupakan data ringkasan saham tanggal 1 Juli 2018 hingga 31 Desember 2018. Setelah model yang terbentuk diaplikasikan terhadap data uji, didapatkan hasil prediksi yang kemudian akan dihitung untuk mendapatkan nilai akurasi. Nilai akurasi kemudian akan digunakan untuk melakukan evaluasi dan analisis terhadap performa algoritma Random Forest dalam memprediksi perubahan harga saham LQ45.

IV. Hasil dan Analisis

Hasil rata-rata akurasi pada pengujian terhadap data latih untuk tiap tipe fitur dan rentang waktu dapat dilihat pada Tabel 6. Terlihat bahwa prediksi menggunakan tipe fitur *plain* memiliki akurasi yang lebih tinggi dan mendekati sempurna daripada tipe fitur *grouped* dan *onehot*, yang rata-rata akurasinya menunjukkan nilai yang hampir identik. Terlihat juga dengan semakin besarnya rentang waktu, akurasi semakin meningkat juga.

TABEL 6
RATA-RATA HASIL PENGUJIAN TERHADAP DATA LATIH

Tipe Fitur dan Rentang Waktu	Akurasi
Plain1	0.999
Plain5	1
Plain20	1
Grouped1	0.65
Grouped5	0.694
Grouped20	0.72
OneHot1	0.65
OneHot5	0.694
OneHot20	0.72

Hasil rata-rata akurasi pada pengujian terhadap data uji untuk tiap tipe fitur dan rentang waktu dapat dilihat pada Tabel 7. Kembali terlihat bahwa prediksi menggunakan tipe fitur *plain* memiliki akurasi yang lebih tinggi dan mendekati sempurna daripada tipe fitur *grouped* dan *onehot*, yang rata-rata akurasinya juga menunjukkan nilai yang hampir identik. Kembali terlihat juga dengan semakin besarnya rentang waktu, akurasi semakin meningkat. Penurunan nilai akurasi terjadi untuk seluruh tipe fitur dan rentang waktu. Hal ini menunjukkan bahwa secara umum model yang dibentuk bersifat *overfitting*, dimana model dapat memprediksi data-data historis dengan baik, namun belum dapat memprediksi data-data baru dengan baik.

TABEL 7
RATA-RATA HASIL PENGUJIAN TERHADAP DATA UJI

Tipe Fitur dan Rentang Waktu	Akurasi
Plain1	0.523
Plain5	0.549
Plain20	0.555
Grouped1	0.483
Grouped5	0.507
Grouped20	0.527
OneHot1	0.484
OneHot5	0.507
OneHot20	0.526

Tabel 8 menunjukkan detil hasil pengukuran akurasi untuk tiap kode saham dan rentang waktu untuk tipe fitur *plain*. Hasil akurasi pada pengujian terhadap data uji untuk fitur *plain* menunjukkan hasil yang paling baik di antara ketiga tipe fitur. Terdapat tiga kode saham dengan akurasi lebih besar dari 60% untuk rentang waktu satu hari, enam kode saham untuk rentang waktu lima hari, dan empat belas kode saham untuk rentang waktu dua puluh hari, sembilan di antaranya di atas 70%. Akurasi pengujian terhadap kode saham ANTM dengan rentang dua puluh hari bahkan bernilai 80,6%.

TABEL 8
HASIL PENGUJIAN TERHADAP DATA UJI UNTUK TIPE FITUR PLAIN

Kode Saham	Akurasi		
	Plain1	Plain5	Plain20
ADRO	0.581	0.484	0.5
ADHI	0.411	0.508	0.347
AKRA	0.581	0.516	0.653
ANTM	0.524	0.589	0.806
ASII	0.524	0.573	0.556
BBCA	0.492	0.484	0.347
BBNI	0.476	0.581	0.702
BBRI	0.508	0.435	0.516
BBTN	0.46	0.492	0.532
BJBR	0.46	0.516	0.726
BKSL	0.548	0.548	0.427
BMRI	0.581	0.54	0.677
BSDE	0.492	0.524	0.492
ELSA	0.452	0.387	0.427
EXCL	0.508	0.508	0.694
GGRM	0.5	0.46	0.315
HMSP	0.589	0.629	0.726
ICBP	0.468	0.669	0.54
INCO	0.524	0.548	0.46
INDF	0.637	0.597	0.484
INKP	0.581	0.5	0.339
INTP	0.581	0.685	0.718
ITMG	0.565	0.629	0.734
JSMR	0.556	0.581	0.645
KLBF	0.516	0.573	0.556
LPKR	0.484	0.476	0.798

JUTEI Edisi Volume.3 No.2 Oktober 2019
ISSN 2579-3675, e-ISSN 2579-5538
DOI 10.21460/jutei.2019.32.187

LPPF	0.5	0.581	0.339
MEDC	0.621	0.573	0.661
PGAS	0.508	0.5	0.444
PTBA	0.484	0.565	0.516
PPTP	0.403	0.605	0.573
SCMA	0.516	0.484	0.492
SMGR	0.54	0.548	0.742
SRIL	0.452	0.548	0.5
SSMS	0.492	0.685	0.702
TLKM	0.556	0.573	0.524
UNTR	0.621	0.573	0.5
UNVR	0.573	0.589	0.492
WIKA	0.508	0.589	0.435
WSKT	0.548	0.524	0.565

Tabel 9 menunjukkan detil hasil pengukuran akurasi untuk tiap kode saham dan rentang waktu untuk tipe fitur *grouped*. Hasil pengujian terhadap tipe fitur *grouped* menunjukkan hasil yang lebih buruk dari tipe fitur *plain*, dimana tidak ada kode saham dengan akurasi lebih besar dari 60% untuk rentang waktu satu hari, satu kode saham untuk rentang waktu lima hari, enam kode saham untuk rentang waktu dua puluh hari, dua di antaranya memiliki akurasi di atas 70%.

TABEL 9
HASIL PENGUJIAN TERHADAP DATA UJI UNTUK TIPE FITUR GROUPED

Kode Saham	Akurasi		
	Grouped1	Grouped5	Grouped20
ADRO	0.548	0.427	0.363
ADHI	0.492	0.508	0.589
AKRA	0.508	0.46	0.597
ANTM	0.427	0.46	0.532
ASII	0.427	0.492	0.556
BBCA	0.508	0.516	0.702
BBNI	0.484	0.589	0.484
BBRI	0.476	0.5	0.548
BBTN	0.484	0.46	0.661
BJBR	0.444	0.556	0.403
BKSL	0.508	0.548	0.5
BMRI	0.565	0.573	0.532
BSDE	0.46	0.395	0.435
ELSA	0.476	0.5	0.411
EXCL	0.5	0.54	0.556
GGRM	0.508	0.605	0.75
HMSP	0.5	0.484	0.403
ICBP	0.484	0.573	0.581
INCO	0.508	0.427	0.54
INDF	0.508	0.492	0.411
INKP	0.347	0.508	0.411
INTP	0.581	0.597	0.605
ITMG	0.427	0.589	0.516
JSMR	0.581	0.516	0.524
KLBF	0.46	0.532	0.54
LPKR	0.427	0.556	0.613
LPPF	0.468	0.476	0.516
MEDC	0.444	0.556	0.548
PGAS	0.5	0.379	0.435
PTBA	0.476	0.532	0.556
PPTP	0.54	0.54	0.589
SCMA	0.46	0.565	0.605
SMGR	0.492	0.516	0.581
SRIL	0.468	0.444	0.516
SSMS	0.395	0.516	0.516
TLKM	0.516	0.508	0.46
UNTR	0.419	0.5	0.444
UNVR	0.556	0.419	0.508
WIKA	0.516	0.516	0.548
WSKT	0.452	0.419	0.548

Tabel 10 menunjukkan detil hasil pengukuran akurasi untuk tiap kode saham dan rentang waktu untuk tipe fitur *onehot*. Hasil pengujian terhadap tipe fitur *onehot* tidak jauh berbeda di mana lima kode saham dengan akurasi lebih besar dari 60% untuk rentang waktu dua puluh hari, dua di antaranya memiliki akurasi di atas 70%. Selisih akurasi antara akurasi pengujian terhadap tipe fitur *grouped* dan *onehot* sangat kecil, dimana hanya kode saham TLKM untuk rentang waktu satu hari yang memiliki selisih lebih besar dari 5%. Hal ini menunjukkan bahwa tidak terdapat perubahan yang signifikan dari *one-hot encoding* yang dilakukan, bahkan hasilnya terkadang lebih buruk daripada nilai akurasi tipe fitur *grouped*.

TABEL 10
HASIL PENGUJIAN TERHADAP DATA UJI UNTUK TIPE FITUR ONEHOT

Kode Saham	Akurasi		
	OneHot1	OneHot5	OneHot20
ADRO	0.524	0.427	0.371
ADHI	0.508	0.508	0.589
AKRA	0.508	0.468	0.597
ANTM	0.435	0.46	0.532
ASII	0.419	0.5	0.556
BBCA	0.508	0.508	0.702
BBNI	0.508	0.589	0.484
BBRI	0.484	0.484	0.524
BBTN	0.476	0.444	0.661
BJBR	0.46	0.556	0.427
BKSL	0.524	0.524	0.508
BMRI	0.597	0.573	0.524
BSDE	0.46	0.395	0.427
ELSA	0.476	0.508	0.395
EXCL	0.5	0.556	0.573
GGRM	0.508	0.597	0.742
HMSP	0.484	0.476	0.403
ICBP	0.484	0.573	0.565
INCO	0.524	0.435	0.524
INDF	0.5	0.5	0.411
INKP	0.355	0.524	0.403
INTP	0.573	0.581	0.605
ITMG	0.46	0.581	0.5
JSMR	0.565	0.524	0.524
KLBF	0.46	0.532	0.532
LPKR	0.427	0.565	0.605
LPPF	0.46	0.484	0.532
MEDC	0.435	0.556	0.556
PGAS	0.516	0.379	0.452
PTBA	0.476	0.524	0.589
PPTP	0.548	0.54	0.573
SCMA	0.452	0.556	0.613
SMGR	0.5	0.516	0.565
SRIL	0.46	0.452	0.524
SSMS	0.403	0.516	0.5
TLKM	0.46	0.524	0.452
UNTR	0.444	0.508	0.435
UNVR	0.548	0.419	0.508
WIKA	0.508	0.516	0.54
WSKT	0.508	0.516	0.5

V. KESIMPULAN

Kesimpulan yang didapat dari penelitian adalah:

1. Hasil pengujian menunjukkan bahwa tipe fitur *plain* menghasilkan akurasi yang lebih tinggi daripada tipe fitur *grouped* dan *onehot*. Penerapan *one-hot encoding* terhadap fitur-fitur *grouped* juga tidak berpengaruh signifikan terhadap nilai akurasi.
2. Semakin besar rentang waktu, nilai akurasi umumnya akan bernilai semakin besar, hal ini berlaku untuk seluruh tipe fitur.
3. Secara umum, tingkat akurasi tidak berkisar jauh dari 50%. Namun dalam beberapa kasus nilai akurasi melebihi 70%, bahkan nilai akurasi kode saham ANTM untuk tipe fitur *plain* dan rentang waktu dua puluh hari mencapai 80,6%.
4. *Chart* dari kode saham ANTM pada periode April 2016 hingga Desember 2018 cenderung bersifat *sideways* dengan nilai yang hampir selalu berada di antara 600 dan 1.000. Hal ini mungkin menjadi salah satu faktor tingginya nilai akurasi kode saham ANTM untuk tipe fitur *plain*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa sehingga penulis dapat menyelesaikan penelitian ini. Penulis juga mengucapkan terima kasih kepada Bapak Yuan Lukito dan Bapak R. Gunawan Santosa yang telah membimbing dan mengarahkan penelitian ini sehingga penelitian ini dapat selesai.

DAFTAR PUSTAKA

[11] A.-S. Chen, M. T. Leung, and H. Daouk, "Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index," *Computers & Operations Research*, vol. 30, no. 6, pp. 901–923, 2003.

[12] O. Hegazy, O. S. Soliman, and M. A. Salam, "A machine learning model for stock market prediction," *arXiv preprint arXiv*:*1402.7351*, 2014.

[13] S. Madge and S. Bhatt. "Predicting Stock Price Direction using Support Vector Machines," *Independent Work Report Spring*, 2015.

[14] A. Sharma et al, "Application of Data Mining–A Survey Paper," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, pp. 2023-2025, 2014.

[15] R. Yamamoto, "Intraday technical analysis of individual stocks on the Tokyo Stock Exchange," *Journal of Banking & Finance*, vol. 36, no. 11, pp. 3033-3047, 2012. Available: 10.1016/j.jbankfin.2012.07.006.

[16] A. Lo, H. Mamaysky, and J. Wang, "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," 2000.

[17] W. W. H. Tsang and T. T. L. Chong, T. T. L, "Profitability of the on-balance volume indicator," *Economics Bulletin*, vol. 29, no. 3, pp. 2424-2431, 2009.

[18] K. Utthammajai and P. Leesutthipornchai, "Association Mining on Stock Index Indicators," *International Journal of Computer and Communication Engineering*, vol. 4, no. 1, pp. 46–49, 2015.

[19] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Amsterdam: Elsevier, 2012.

[20] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine*, vol. 17, no. 3, pp. 37-37, 1996.

[21] M. Kamber, L. Winstone, W. Gong, S. Cheng, and J. Han, "Generalization and decision tree induction: efficient classification in data mining," *Proceedings Seventh International Workshop on Research Issues in Data Engineering. High Performance Database Management for Large-Scale Applications*, pp. 111–120, Apr. 1997.

[22] M. Denil, D. Matheson and N. De Freitas, "Narrowing the gap: Random forests in theory and in practice," *International conference on machine learning,* pp. 665-673, Jan. 2014.

[23] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.

[24] S. Visa et al, "Confusion Matrix-based Feature Selection," *MAICS,* pp. 120-127, Apr. 2011.

JUTEI Edisi Volume.3 No.2 Oktober 2019
ISSN 2579-3675, e-ISSN 2579-5538
DOI 10.21460/jutei.2019.32.187