

SISTEM PREDIKSI HARGA SAHAM LQ45 DENGAN RANDOM FOREST CLASSIFIER

Skripsi



oleh
KEVIN VALIANT
71150151

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

SISTEM PREDIKSI HARGA SAHAM LQ45 DENGAN RANDOM FOREST CLASSIFIER

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

KEVIN VALIANT
71150151

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

SISTEM PREDIKSI HARGA SAHAM LQ45 DENGAN RANDOM FOREST CLASSIFIER

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 17 Desember 2019



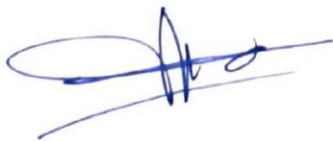
KEVIN VALIANT
71150151

HALAMAN PERSETUJUAN

Judul Skripsi : SISTEM PREDIKSI HARGA SAHAM LQ45
DENGAN RANDOM FOREST CLASSIFIER
Nama Mahasiswa : KEVIN VALIANT
N I M : 71150151
Matakuliah : Skripsi (Tugas Akhir)
Kode : TIW276
Semester : Gasal
Tahun Akademik : 2019/2020

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 17 Desember 2019

Dosen Pembimbing I



Yuan Lukito, S.Kom., M.Cs.

Dosen Pembimbing II



R. Gunawan Santosa, Drs. M.Si.

HALAMAN PENGESAHAN

SISTEM PREDIKSI HARGA SAHAM LQ45 DENGAN RANDOM FOREST CLASSIFIER

Oleh: KEVIN VALIANT / 71150151

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 13 Desember 2019

Yogyakarta, 17 Desember 2019
Mengesahkan,

Dewan Penguji:

1. Yuan Lukito, S.Kom., M.Cs.
2. R. Gunawan Santosa, Drs. M.Si.
3. Widi Hapsari, Dra. M.T.
4. Prihadi Beny Waluyo, SSi., MT.

Dekan

Ketua Program Studi

(Restyandito, S.Kom., MSIS., Ph.D.)

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Penulis banyak menerima bimbingan, bantuan serta dorongan dari berbagai pihak baik yang bersifat moral maupun material. Pada kesempatan ini penulis menyampaikan rasa terima kasih yang kepada:

1. Tuhan Allah yang selalu memberkati penulis dengan semangat dan kekuatan hingga dapat menyelesaikan skripsi ini.
2. Kedua orang tua dan adik penulis, yang selalu mendukung penulis dengan perhatian, kasih sayang, semangat, dan doa yang tiada habisnya hingga penulis dapat menyelesaikan skripsi ini.
3. Yuan Lukito, S.Kom., M.Cs selaku dosen pembimbing pertama dan Drs. R. Gunawan Santosa, M.Si selaku dosen pembimbing kedua yang telah memberikan bimbingan terhadap penulis sehingga skripsi ini dapat diselesaikan.
4. Gloria Virginia, S.Kom., MAI, Ph. D. selaku Kepala Program Studi Informatika UKDW yang telah memberikan bimbingan dan bantuan dalam penyusunan skripsi ini.
5. Sahabat-sahabat, Gabriella Amelia Prasetyo, Albertus Istora Pahala, Antonius Dimas Pradana, Ditrissa Taranadia, Febriyandi, Fransiska Litmantoro Hidayat, Hutrima Gusti Cahyadi, Kevin Giovanni, Yudhistira Audri Pranata, dan Zenas Dwigrasia yang selalu menjadi pengingat, penyemangat, tempat bercerita dan teman bekerja selama pengerjaan skripsi ini.
6. Jessi Febria yang menjadi tempat cerita dan sumber inspirasi serta penghiburan penulis selama pengerjaan skripsi.
7. Semua teman-teman selalu mengingatkan dan menanyakan tentang kemajuan skripsi penulis.
8. Serta seluruh pihak yang tidak dapat penulis sebut satu per satu yang telah memberikan banyak bantuan dalam pengerjaan skripsi ini.

Semoga Tuhan Allah selalu memberikan balasan atas segala kebaikan yang penulis terima. Semoga penelitian ini dapat bermanfaat bagi peneliti dan juga bagi para pembaca.

©UKDW

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
UCAPAN TERIMA KASIH.....	vi
INTISARI.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	1
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Metodologi Penelitian.....	2
1.6. Sistematika Penulisan.....	3
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	5
2.1. Tinjauan Pustaka.....	5
2.2. Landasan Teori.....	5
2.2.1. Saham.....	5
2.2.2. Technical Analysis.....	6
2.2.3. Exponential Moving Average (EMA).....	7
2.2.4. On-Balance Volume (OBV).....	8
2.2.5. Chaikin Oscillator.....	10
2.2.6. Moving Average Convergence/Divergence.....	12
2.2.7. Bollinger Bands.....	14
2.2.8. Data Mining.....	16
2.2.9. Decision Tree.....	16
2.2.10. Random Forest Classifier.....	18
2.2.11. Akurasi.....	19

BAB 3 METODOLOGI PENELITIAN.....	21
3.1. Gambaran Umum Sistem	21
3.2. Analisis Kebutuhan Sistem	21
3.2.1. Analisis Data	21
3.2.2. Analisis Kebutuhan Perangkat Lunak.....	22
3.2.3. Analisis Kebutuhan Perangkat Keras.....	22
3.3. Alur Kerja Sistem	22
3.4. Rancangan Basis Data Sistem	25
3.5. Rancangan Antar Muka	29
3.5.1. Halaman Home.....	29
3.5.2. Halaman Update.....	29
3.5.3. Halaman Data.....	30
3.5.4. Halaman Chart	30
3.5.5. Halaman Tree	31
3.5.6. Halaman Result.....	31
3.5.7. Halaman Summary.....	32
3.6. Perancangan Pengujian Sistem.....	32
BAB 4 IMPLEMENTASI DAN ANALISIS SISTEM.....	33
4.1. Implementasi Sistem	33
4.1.1. Implementasi Antarmuka	33
4.1.2. Hasil Sistem	37
4.1.3. Analisis Sistem.....	41
BAB 5 KESIMPULAN DAN SARAN	43
5.1. Kesimpulan.....	43
5.2. Saran	43
DAFTAR PUSTAKA	44
LAMPIRAN 1. Source Code	46
LAMPIRAN 2. Pindaian Kartu Konsultasi.....	51
LAMPIRAN 3. Grafik OHLCV Saham LQ45 Periode Juli 2015-Desember 2018	53

DAFTAR TABEL

Tabel 2. 1 Contoh data harga penutupan saham perusahaan x selama 14 hari.	7
Tabel 2. 2 Contoh data ringkasan saham perusahaan x selama 5 hari.	10
Tabel 2. 3 Contoh data EMA-12 dan EMA-26 selama 14 hari dari perusahaan x.	12
Tabel 2. 4 Contoh data harga penutupan saham perusahaan x selama 14 hari.	14
Tabel 2. 5 Contoh data untuk memprediksi apakah seseorang akan bermain golf	16
Tabel 2. 6 Contoh confusion matrix dengan $n=2$	20
Tabel 3. 1 Deskripsi tabel prediksi_stocks.....	25
Tabel 3. 2 Deskripsi tabel prediksi_plainstocks.....	26
Tabel 3. 3 Deskripsi tabel prediksi_groupedstocks.....	27
Tabel 3. 4 Deskripsi tabel prediksi_onehotstocks.....	28
Tabel 4. 1 Akurasi rata-rata pengujian terhadap data latih	37
Tabel 4. 2 Akurasi rata-rata pengujian terhadap data uji.	37
Tabel 4. 3 Akurasi pengujian terhadap data uji untuk tipe fitur plain.....	38
Tabel 4. 4 Akurasi pengujian terhadap data uji untuk tipe fitur grouped	39
Tabel 4. 5 Akurasi pengujian terhadap data uji untuk tipe fitur onehot.....	40

DAFTAR GAMBAR

Gambar 2. 1. Grafik yang menunjukkan harga, volume dan OBV saham Wal-Mart pada November-Desember 2010	9
Gambar 2. 2. Grafik yang menunjukkan harga, volume dan Chaikin Oscillator saham Microsoft pada September-Desember 2010	11
Gambar 2. 3. Grafik yang menunjukkan harga, garis MACD, garis sinyal dan histogram saham Powershares QQQ Trust pada Februari-April 2009	13
Gambar 2. 4. Grafik yang menunjukkan harga, Bollinger Bands serta standar deviasi dari SPDR S&P 500 ETF Trust pada April-Juni 2009	15
Gambar 2. 5. Decision tree yang terbentuk dari contoh data	18
Gambar 3. 1. Contoh format data yang disimpan dalam bentuk file .csv.	22
Gambar 3. 2. Alur proses latih	23
Gambar 3. 3. Alur proses prediksi	25
Gambar 3. 4. Antar muka halaman home	29
Gambar 3. 5. Antar muka halaman update	29
Gambar 3. 6. Antar muka halaman data	30
Gambar 3. 7. Antar muka halaman chart	30
Gambar 3. 8. Antar muka halaman tree	31
Gambar 3. 9. Antar muka halaman update	31
Gambar 3. 10. Antar muka halaman summary	32
Gambar 4. 1. Antar muka halaman home	33
Gambar 4. 2. Antar muka halaman update	34
Gambar 4. 3. Antar muka halaman data	34
Gambar 4. 4. Antar muka halaman chart	35
Gambar 4. 5. Antar muka halaman tree	35
Gambar 4. 6. Antar muka halaman result	36
Gambar 4. 7. Antar muka halaman summary	36

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

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 (Sharma et al., 2014). 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.

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.

1.2. Rumusan Masalah

Permasalahan yang akan dibahas pada penelitian ini adalah:

1. Bagaimana cara mengimplementasikan metode Random Forest Classifier untuk memprediksi harga saham?
2. Bagaimana performa metode Random Forest Classifier dalam memprediksi harga saham berdasarkan akurasi?

1.3. Batasan Masalah

Pada penelitian ini, penulis memberikan batasan masalah, antara lain:

1. Data yang digunakan dibatasi menjadi ringkasan saham perusahaan yang terdaftar pada LQ45 versi Agustus 2018-Januari 2019 selain Waskita Beton Precast Tbk. (WSBP), Barito Pacific Tbk. (BRPT), Indika Energy Tbk. (INDY), dan Media Nusantara Citra Tbk. (MNCN).
2. Data yang digunakan diambil dari data Phillip Sekuritas Indonesia terhitung sejak 1 Juli 2015-30 Juni 2018 (data latih) dan 1 Juli 2018-31 Desember 2018 (data uji).
3. Rentang waktu perubahan harga saham yang digunakan adalah 1 hari, 5 hari, dan 20 hari.
4. Hasil prediksi adalah prediksi harga dengan label bernilai naik, turun, atau tetap.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah mengimplementasikan metode Random Forest Classifier untuk memprediksi harga saham dan mengukur performanya berdasarkan akurasi.

1.5. Metodologi Penelitian

Dalam penyelesaian tugas akhir ini, penulis menggunakan beberapa metode, yaitu:

1. Pengumpulan Data

Pada tahap ini, data yang dibutuhkan dikumpulkan untuk digunakan pada tahap selanjutnya.

2. Perancangan dan Pembuatan Sistem

Tahap ini terdiri dari beberapa sub-tahapan, yaitu:

2.1. Pelabelan Data

Dataset yang ada dilabeli sesuai dengan kondisi harga saham pada hari tersebut dibanding 1 hari, 5 hari, dan 20 hari berikutnya. Label yang digunakan dalam penelitian ini adalah naik, turun, dan tetap.

2.2. Penghitungan *Technical Indicators*

Seluruh dataset kemudian dihitung nilai *technical indicators*-nya, yang meliputi On-Balance Volume, Chaikin Oscillator, Moving Average Convergence/Divergence, dan Bollinger Bands.

2.3. Pembentukan Fitur

Dataset yang sudah dihitung nilai *technical indicator*-nya kemudian dibentuk menjadi beberapa dataset, yaitu *plain* yang fiturnya adalah data ringkasan saham harian serta *technical indicator*-nya, *grouped* yang fiturnya adalah 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*.

2.4. Klasifikasi

Pada tahap ini dibentuk model Random Forest untuk setiap kode saham, tipe fitur, dan rentang waktu.

2.5. Evaluasi

Pada tahap ini performa model yang dibentuk dievaluasi dengan mengukur akurasi menggunakan *confusion matrix*.

3. Analisis

Analisis kemudian dilakukan terhadap nilai akurasi yang didapatkan dari hasil evaluasi terhadap sistem.

1.6. Sistematika Penulisan

Penulisan laporan tugas akhir ini dibagi menjadi 5 bagian, yaitu:

Bab 1 Pendahuluan yang berisi tentang gambaran umum penelitian. Bab ini terdiri dari latar belakang penelitian, rumusan masalah, batasan-batasan masalah, hipotesis, metodologi penelitian, serta sistematika penulisan dari penelitian.

Bab 2 Tinjauan Pustaka dan Landasan Teori yang berisi bahasan penelitian dari berbagai referensi mengenai topik penelitian.

Bab 3 Metodologi Penelitian yang berisi rancangan pembangunan sistem, arsitektur sistem, algoritma yang digunakan dalam pembuatan sistem, alur kerja

sistem, serta pemenuhan kebutuhan akan *hardware* maupun *software* untuk mendukung penelitian yang akan dilakukan oleh penulis.

Bab 4 Hasil dan Pembahasan yang berisi uraian detil implementasi sistem dan analisis dari implementasi sistem.

Bab 5 Kesimpulan dan Saran yang berisi kesimpulan dari hasil penelitian serta saran-saran untuk penelitian selanjutnya yang sejenis.

©UKPDW

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Metode klasifikasi menggunakan Random Forest telah berhasil diimplementasikan untuk memprediksi harga saham LQ45, walaupun hasil yang didapatkan masih jauh dari sempurna. Berdasarkan analisis terhadap hasil pengujian yang telah penulis lakukan, dapat diambil kesimpulan sebagai berikut:

1. Hasil pengujian menunjukkan bahwa tipe fitur *plain* menghasilkan akurasi yang lebih tinggi daripada tipe fitur *grouped* dan *one-hot*. 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*.

5.2. Saran

Saran yang dapat penulis berikan terhadap penelitian lanjutan dan pengembangan sistem selanjutnya adalah sebagai berikut:

1. Pengujian dilakukan dengan menggunakan *technical indicator* yang jumlahnya sangat banyak, sedangkan dalam penelitian ini hanya digunakan empat *technical indicator*.
2. Penggunaan *database management system* selain SQLite, karena SQLite hanya mengizinkan satu proses tulis dalam satu waktu, sehingga memakan waktu saat proses input data.

DAFTAR PUSTAKA

- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, 30(6), 901-923.
- Denil, M., Matheson, D., & De Freitas, N. (2014, January). Narrowing the gap: Random forests in theory and in practice. In *International conference on machine learning* (pp. 665-673).
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. *arXiv preprint arXiv:1402.7351*.
- Kamber, M., Winstone, L., Gong, W., Cheng, S., & Han, J. (1997, April). Generalization and decision tree induction: efficient classification in data mining. In *Proceedings Seventh International Workshop on Research Issues in Data Engineering*. High Performance Database Management for Large-Scale Applications (pp. 111-120). IEEE.
- Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance*, 55(4), 1705-1765.
- Madge, S., & Bhatt, S. (2015). Predicting Stock Price Direction using Support Vector Machines. *Independent Work Report Spring*.

- Sharma, A., Sharma, R., Sharma, V. K., & Shrivatava, V. (2014). Application of Data Mining—A Survey Paper. *International Journal of Computer Science and Information Technologies*, 5(2), 2023-2025.
- Tsang, W. W. H., & Chong, T. T. L. (2009). Profitability of the on-balance volume indicator. *Economics Bulletin*, 29(3), 2424-2431.
- Utthammajai, K., & Leesutthipornchai, P. (2015). Association mining on stock index indicators. *International Journal of Computer and Communication Engineering*, 4(1), 46.
- Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011, April). Confusion Matrix-based Feature Selection. In *MAICS* (pp. 120-127).
- Yamamoto, R. (2012). Intraday technical analysis of individual stocks on the Tokyo Stock Exchange. *Journal of Banking & Finance*, 36(11), 3033-3047.