

Prediksi Akurasi Perusahaan Saham Menggunakan SVM dan K-Fold Cross Validation

Ali Akbar Lubis¹, Ng Poi Wong², Frans Mikael Sinaga³

STMIK Mikroskil, Jl. Thamrin No. 112, 124, 140, Telp. (061) 4573767, Fax. (061) 4567789

Program Studi Teknik Informatika, STMIK Mikroskil, Medan

¹ali.akbar@mikroskil.ac.id, ²poiwong@mikroskil.ac.id, ³frans.sinaga@mikroskil.ac.id

Abstrak

Setiap perusahaan akan mengalami pergerakan harga saham. seorang pialang sering sekali kesulitan untuk memutuskan apakah harus membeli dan menjual saham, dikarenakan banyaknya perusahaan saham yang nilai saham nya tidak menentu. Pada penelitian ini, dataset yang digunakan diambil dari perusahaan Blue Chip di Bursa Efek Indonesia sebanyak lima tahun data perusahaan saham. Penelitian ini bertujuan untuk memprediksi akurasi perusahaan saham, dan pada teknik ini dilakukan perbandingan kernel Radial Basis Function dan Polynomial untuk mengatasi data nonlinear. Setelah diperoleh vector terbaik dari setiap data saham, maka akan dilakukan pembagian data saham sesuai dengan fungsi linear pada SVM untuk dihitung akurasi data saham dari setiap data perusahaan saham menggunakan K-Fold Cross Valodation. Hasil pengujian menunjukkan dengan menggunakan K-Fold Cross Validation dapat diketahui bahwa perusahaan BBKA adalah perusahaan yang akurasi data nya lebih baik. Dengan hasil akurasi yang diperoleh adalah RBF = 63.083164% dan Polynomial = 63.083164%.

Kata kunci—Prediksi Akurasi, SVM, K-Fold Cross Validation

Abstract

Every company will experience stock price movements. A brokerage is often difficult to decide whether to buy and sell stock, because there are many stock companies which the stock values are uncertain. In this study, the dataset used was taken from the Blue chip company on the Indonesia Stock Exchange for five years of stock company data. This study aims to predict the accuracy of stock companies, and in this technique a kernel comparison of Radial Basis Function and Polynomial was used to overcome nonlinear data. After obtaining the best vector from each share data, the share data will be distributed according to the linear function in SVM to calculate the accuracy of stock data from each stock company data using K-Fold Cross Valodation. The test results show that by using K-Fold Cross Validation it can be seen that BBKA is a company with better data accuracy. With the results obtained accuracy is RBF = 63.083164% and Polynomial = 63.083164%.

Keywords—Predicition Accuracion, SVM, K-Fold Cross Validation

1. PENDAHULUAN

Saat ini masyarakat semakin memandang saham sebagai salah satu alternatif dalam menginvestasikan dana yang dimilikinya. Hal ini terbukti dengan semakin maraknya kegiatan-kegiatan di pasar saham [1]. Saham merupakan kertas tanda bukti penyertaan kepemilikan modal/dana pada suatu perusahaan yang tercantum dengan jelas nilai nominal, nama perusahaan dan diikuti dengan hak dan kewajiban yang jelas kepada setiap pemegangnya [2]. Pada Bursa Efek Indonesia terdapat banyak perusahaan saham yang terdaftar di dalamnya. Dari perusahaan saham tersebut, pemain saham bahkan pialang saham terkadang kesulitan memprediksi harga saham. Terdapat banyak algoritma untuk memprediksi harga saham, tetapi tidak semua algoritma dapat memprediksi harga saham dengan akurat. Selain itu, setiap saham yang dimiliki oleh perusahaan pemilik saham memiliki akurasi harga saham yang berbeda-beda, sehingga tidak semua saham sesuai dengan algoritma prediksi harga saham yang diterapkan.

Harga saham merupakan salah satu faktor untuk mengukur keberhasilan pengelolaan perusahaan, jika harga saham suatu perusahaan selalu mengalami kenaikan, maka investor atau calon investor menilai bahwa perusahaan berhasil dalam mengelola usahanya. Kepercayaan investor atau calon investor sangat bermanfaat bagi perusahaan, karena semakin banyak orang yang percaya terhadap perusahaan tersebut maka keinginan untuk berinvestasi pada perusahaan tersebut semakin kuat [3]. Secara teori, semakin tinggi harga saham suatu perusahaan maka semakin tinggi pula nilai perusahaan yang berbanding lurus dengan tinggi pula keuntungan yang akan diperoleh serta risiko yang akan ditanggung [4].

Ada berbagai jenis perusahaan yang menawarkan saham kepada investor, diantaranya Bank Rakyat Indonesia (BBRI), Astra Internasional (ASII JK), Perusahaan gas Negara (PGAS), Telkom Indonesia (TLKM) dan Bank Central Indonesia (BBCA). Sehingga investor perlu melakukan prediksi perusahaan saham yang paling optimal. Untuk melakukan proses prediksi perusahaan saham ini digunakan metode Support Vector Machine (SVM). Karena SVM merupakan salah satu dari sekian banyak metode yang dapat digunakan untuk menyelesaikan berbagai jenis permasalahan termasuk prediksi [5].

Dengan menerapkan SVM (Support Vector Machine), overfitting cenderung terjadi, dan solusi optimal juga mungkin terjadi secara global. Meskipun SVM yang terpisah memiliki kinerja yang luar biasa, namun kinerja klasifikasi dan kemampuan generalisasi classifier sering dipengaruhi oleh dimensinya atau jumlah variabel fitur. Dengan demikian, para peneliti lebih memilih untuk menggabungkan teknik lain dengan SVM untuk mengembangkan model peramalan yang efisien [6]. SVM menggunakan model linier untuk memisahkan data sampel melalui beberapa pemetaan nonlinear dari vektor input ke dalam ruang fitur dimensi tinggi. Model linier dibangun di ruang baru dapat mewakili nonlinier batas keputusan di ruang asli. SVM bertujuan untuk menemukan jenis model linier khusus, yang disebut memisahkan hyperplanes. Poin pelatihan yang lebih dekat ke hyperplane pemisah yang optimal disebut dukungan vektor, yang menentukan batas keputusan. Dalam kasus umum dimana data tidak dipisahkan secara linear, SVM menggunakan mesin nonlinear untuk menemukan hyperplane yang meminimalkan jumlah kesalahan pada set pelatihan [7].

Prinsip kerja dari metode SVM adalah menentukan model classifier (biasanya berupa hyperplane) yang memisahkan data dari dua kelas yang berbeda. Koefisien-koefisien dari hyperplane classifier dipilih sehingga jarak hyperplane ke titik-titik terdekat dari kedua kelas paling jauh [8]. Vector yang dihasilkan oleh SVM akan diproses dengan menggunakan metode K-Fold Cross Validation (KFCV). KFCV berperan untuk mengetahui tingkat akurasi dari masing-masing data perusahaan saham dari vector terbaik data yang dihasilkan SVM.

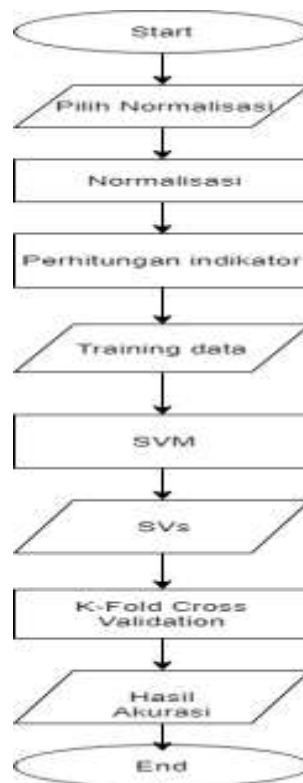
Setiap perusahaan akan mengalami pergerakan harga saham. Investor akan memperhatikan dinamika harga saham setiap saat untuk memperkirakan waktu yang tepat untuk menjual atau membeli saham agar memperoleh keuntungan [9]. Salah satu cara untuk memprediksi harga saham adalah analisis teknikal. Dimana Analisis teknikal merupakan analisis yang menggunakan data historis harga saham suatu perusahaan [9]. Dengan kata lain, analisis teknikal digunakan untuk memprediksi pergerakan saham berdasarkan pola harga saham dengan menggunakan indikator pada jenis saham tertentu. Adapun jenis indikator pada analisis teknikal yaitu XMA (Exponential Moving Average), STI (Stochastic Oscillator), RSI (Relative Strength Index), PROC (Rate of change and momentum), ADO/CLV (Accumulation/Distribution).

Dalam melakukan analisis teknikal untuk memprediksi akurasi dari masing-masing perusahaan saham digunakan SVM (Support Vector Machine) dengan menggunakan kernel RBF (Radial Basis Function) karena model linear dimana fungsi basis berupa radial basis function, yaitu fungsi yang tergantung pada jarak antara argumennya. SVM digunakan untuk memilih data pelatihan penting. SVM juga akan menghasilkan vector-vektor terbaik dari setiap data perusahaan-perusahaan terkait, data dari masing-masing perusahaan tersebut akan diberikan kelas dari masing-masing data. Data perusahaan yang telah ditraining akan menghasilkan Support Vector dari masing-masing data. Data perusahaan

saham yang akan diketahui akurasi nya adalah data saham yang merupakan Support Vector yang sudah memiliki kelas yang terbagi berdasarkan nilai K pada tahapan K-Fold Cross Validation.

2. METODE PENELITIAN

Pada tahap ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan, masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah dan tantangan yang ada. Proses dalam SVM dan KFCV akan digambarkan dalam bentuk flowchart pada gambar 1.



Gambar 1 Flowchart SVM and K-Fold Cross Validation

1. Normalisasi Data

Pada tahap ini sebelum nilai digunakan untuk proses indikator. Data harga tersebut diproses kedalam formula adaptive combiner linear pada persamaan 1. Proses ini berfungsi memperkecil nilai dengan rentang +1 sampai -1 yang selanjutnya akan digunakan kedalam pengolahan melalui indikator saham yang dipilih. Berikut ini adalah formula adaptive combiner linear untuk mencari nilai normalisasi.

$$y = \frac{2 * x - (\max + \min)}{(\max + \min)} \quad (1)$$

Keterangan:

y : Nilai normalisasi.

X : Harga saham sebenarnya.

Max : Harga saham terbesar dalam periode pengamatan

Min : Harga saham terkecil dalam periode pengamatan.

2. Perhitungan Indikator

Pada proses ini indikator yang digunakan adalah Exponential Moving Average (EMA/XMA). Accumulation/Distribution Oscillator (ADO/CLV). Stochastic indicator (STI). MA (Moving Average). Price Rate of Change (PRO C). dan Chaikin Money Flow (CMF). Indikator yang digunakan pada proses

ini disimbolkan dengan D. Jumlah D yang digunakan saat ini adalah 10. Karena syarat tersebut maka beberapa indikator dibagi lagi menjadi beberapa periode hari yaitu EMA/XMA 10. EMA/XMA 20. EMA/XMA 30. STI 14. STI 19. STI 30. PROC. ADO/CLV. Setiap nomor pendampin indikator menjelaskan beberapa lama hari yang digunakan datanya untuk proses perhitungan. Setiap hasil dari perhitungan disimbolkan dengan x_n Berikut ini adalah contoh proses perhitungan masing-masing indikator dengan menggunakan data saham open, high, low dan close yang sudah dinormalisasi. Sesuai dengan persamaan 2:

$$\%K = \frac{(\text{Today Close} - \text{Lowest low in K period})}{(\text{highest High in K period} - \text{Lowest Low in K period})} \times 100 \quad (2)$$

Ket:

%K : Posisi relatif
 Today close : Harga penutupan pada hari pengamatan
 Lowest low : Harga low terendah pada hari pengamatan
 Highest high : Harga high tertinggi pada periode hari pengamatan.
 (n) : Periode hari pengamatan.

3. Support Vector Machine

Sebelum masuk ke proses SVM, data yang akan digunakan sebagai data latih untuk SVM yang sudah dinormalisasi, dan sudah dilakukan proses perhitungan indikator yaitu: Open, High, Low, Close, XMA10, XMA20, XMA30, STI14, STI19 untuk memperoleh support vector dari setiap data saham. Sesuai dengan persamaan 3:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^h \quad (3)$$

4. K-Fold Cross Validation

Data yang telah melewati proses SVM akan menghasilkan support vector dari setiap data, vector terbaik dari setiap data akan diberikan label/kelas pada tahapan SVM, data yang telah diberikan kelas akan diuji dengan menggunakan K-Fold Cross Validation.



Gambar 2 K-Fold Cross Validation

K-Fold Cross Validation digunakan untuk mengetahui tingkat akurasi dari masing-masing perusahaan yang terlibat. Ada beberapa langkah cara kerja dari K-Fold Cross Validation untuk menentukan akurasi dari masing-masing data perusahaan saham, yaitu menentukan banyak nilai K, merandom data set, membagi partisi pada data training dan data testing, kemudian menentukan akurasi dari nilai K. Tahapan untuk mengetahui tingkat akurasi dengan K-Fold Cross Validation dapat dilihat pada gambar 2 di atas.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Data yang digunakan pada penelitian ini untuk mendapatkan gambaran analisis proses prediksi data perusahaan saham TLKM, BBRI, PGAS, ASII, BBRI yang diambil dari Bursa Efek Indonesia. Proses pelatihan data digunakan untuk menghasilkan model yang akan digunakan pada proses testing. Proses yang dilakukan pada tahap training adalah sebagai berikut:

1. Input Data

Jumlah data saham yang digunakan pada proses training terdiri dari 1437 data dari masing-masing perusahaan saham selama 5 tahun.

2. Konversi Data

Data saham dikonversi dengan tahapan normalisasi agar dapat disesuaikan dengan libsvm.dll.

3. Training Data

Proses training merupakan proses yang bertujuan untuk melatih SVM dalam mengenali dataset dan membentuk sebuah model berdasarkan pelatihan tersebut. Sebelum melakukan proses training, buat sebuah file csv excel yang berisi data saham dari perusahaan.

3.2 Pengujian

Proses pengujian (testing) terhadap model yang sudah dibangun pada tahap training. Data yang digunakan pada penelitian ini menggunakan 10 indikator saham berdasarkan analisis teknikal yang telah

melewati tahapan normalisasi data. Kemudian dihitung berdasarkan persamaan tiap-tiap indikator saham yang sudah ditetapkan. Indikator saham yang telah melewati tahapan normalisasi dan perhitungan indikator saham berdasarkan persamaan indikator saham yang ditetapkan akan dilanjutkan pada proses tahapan SVM. Pada proses SVM akan ditentukan nilai kelas pada setiap fitur. Hasilnya dapat dilihat pada tabel 1 berikut:

Tabel 1 Penentuan Kelas dari Data Support Vector

No	Open	High	Kelas (Open)	Kelas (High)	Total	Kelas
1	-0.008000	-0.014151	-1	-1	-2	-1
2	-0.004800	-0.003145	-1	-1	-2	-1
3	-0.008000	-0.018868	-1	-1	-2	-1
4	-0.008000	-0.014151	-1	-1	-2	-1
5	-0.006400	-0.001572	-1	-1	-2	-1
6	0.014151	0.022913	1	1	2	1
7	0.040000	0.064465	1	1	2	1
8	0.024000	0.022013	1	1	2	1
9	0.014400	0.003145	1	1	2	1
10	-0.011200	-0.017296	-1	-1	-2	-1
11	-0.012800	0.007862	-1	1	0	1
12	0.022400	0.022013	1	1	2	1
13	-0.008000	0.011006	-1	1	0	1
14	0.003200	-0.007862	1	-1	0	1
15	-0.083200	-0.094340	-1	-1	-2	-1
16	-0.073600	-0.070755	-1	-1	-2	-1
17	-0.052800	-0.066038	-1	-1	-2	-1
18	-0.052800	-0.037736	-1	-1	-2	-1
19	-0.035200	-0.034591	-1	-1	-2	-1
20	-0.035200	-0.051887	-1	-1	-2	-1
21	-0.056000	-0.050314	-1	-1	-2	-1
22	-0.033600	-0.048742	-1	-1	-2	-1
23	-0.072000	-0.070755	-1	-1	-2	-1
24	-0.062400	-0.070755	-1	-1	-2	-1
25	-0.012800	-0.029874	-1	-1	-2	-1
26	-0.008000	-0.017296	-1	-1	-2	-1
27	-0.014400	-0.028302	-1	-1	-2	-1
28	-0.020800	-0.020440	-1	-1	-2	-1
29	-0.016000	-0.011006	-1	-1	-2	-1

Tabel 1 di atas digunakan untuk menentukan posisi letak kelas pada setiap data.

Hasil kelas Open diperoleh dari:

1. Jika data Open kurang dari nol maka akan masuk ke kelas -1.
2. Jika data Open lebih sama dengan dari nol maka akan masuk ke kelas +1.

Hasil kelas High diperoleh dari:

1. Jika data High kurang dari nol maka akan masuk ke kelas -1.
2. Jika data High lebih sama dengan dari nol maka akan masuk ke kelas +1.

Total diperoleh dari hasil penjumlahan kelas open dan high.

Penentuan Kelas. diperoleh dari hasil total.

1. Jika total kurang dari nol maka akan masuk ke kelas -1.
2. Jika total lebih sama dengan dari nol maka akan masuk ke kelas +1.

Dalam melakukan pengujian dengan *K-Fold Cross Validation*, langkah-langkah yang dilakukan adalah:

- Menentukan banyak nilai K
- Merandom dataset.
- Membagi partisi pada data training dan data testing
- Menentukan Akurasi dari nilai K

Hasil pengujiannya dapat dilihat pada tabel 2 berikut ini:

Tabel 2 Pengujian K = 5

K	Data Acak	Open	High
K = 1	15	-0.083200	-0.094340
	7	0.040000	0.064465
	16	-0.073600	-0.070755
	6	0.014151	0.022913
	20	-0.035200	-0.051887
	25	-0.012800	-0.029874
K = 2	12	0.022400	0.022013
	4	-0.008000	-0.014151
	24	-0.062400	-0.070755
	28	-0.020800	-0.020440
	13	-0.008000	0.011006
	11	-0.012800	0.007862
K = 3	5	-0.006400	-0.001572
	27	-0.014400	-0.028302
	23	-0.072000	-0.070755
	8	0.024000	0.022013
	17	-0.052800	-0.066038
	3	-0.008000	-0.018868
K = 4	2	-0.004800	-0.003145
	21	-0.056000	-0.050314
	18	-0.052800	-0.037736
	1	-0.008000	-0.014151
	9	0.014400	0.003145
	29	-0.016000	-0.011006
K = 5	10	-0.011200	-0.017296
	14	0.003200	-0.007862
	26	-0.008000	-0.017296
	19	-0.035200	-0.034591
	22	-0.033600	-0.048742

Kemudian pengujian dilakukan dengan kernel Radial Basis Function dan Polynomial pada tahapan SVM, untuk melihat apakah perbedaan kernel yang digunakan dapat mempengaruhi hasil akurasi dari setiap perusahaan saham yang digunakan. Hasil pengujiannya dapat dilihat pada tabel 3 berikut:

Tabel 3 Pengujian RBF dan Polynomial

Perusahaan	Banyak Data (Tahun)	Open	High
ASII	1	60.728745%	60.728745%
	2	61.866126%	61.866126%
	3	60.786974%	60.786974%
	4	59.123344%	59.123344%

	5	60.522023%	60.522023%
BBCA	1	62.348178%	62.348178%
	2	63.083164%	63.083164%
	3	57.937585%	57.937585%
	4	57.288481%	57.288481%
	5	59.706362%	59.706362%
BBRI	1	57.489879%	57.489879%
	2	55.172414%	55.172414%
	3	55.902307%	55.902307%
	4	54.230377%	54.230377%
	5	53.915171%	53.915171%
PGAS	1	55.465587%	55.465587%
	2	52.738337%	52.738337%
	3	55.766621%	55.766621%
	4	55.963303%	55.963303%
	5	58.156607%	58.156607%
TLKM	1	61.943320%	61.943320%
	2	61.866126%	61.866126%
	3	58.344640%	58.344640%
	4	61.060143%	61.060143%
	5	61.011419%	61.011419%

4. KESIMPULAN

Berdasarkan pengujian yang sudah dilakukan dapat disimpulkan bahwa

1. Banyaknya data perusahaan saham dapat mempengaruhi akurasi dari data saham tersebut. Dapat dilihat dari pemanfaatan teknik K-Fold Cross Validation dengan membagi partisi data saham $K = 5$, akurasi terbaik yang diperoleh dari perusahaan ASII = 63.083164% sebanyak 2 tahun data (2011 – 2012), perusahaan BBCA = 63.083164% sebanyak 2 tahun data (2011 – 2012), perusahaan BBRI = 57.489879% sebanyak 1 tahun data (2011), perusahaan PGAS = 58.156607% sebanyak 5 tahun data (2011 – 2015) dan perusahaan TLKM = 61.943320% sebanyak 1 tahun data (2011).
2. Dengan melakukan teknik perbandingan kernel RBF dan Polynomial pada tahapan SVM, ternyata tidak mempengaruhi hasil akurasi data yang diperoleh dari K-Fold Cross Validation.
3. Dengan teknik pengujian K-Fold Cross validation data perusahaan BBCA adalah perusahaan yang menghasilkan akurasi terbaik sebesar 63.083164%.
4. Dengan melakukan pengujian data terbesar yaitu 5 tahun data dari setiap perusahaan saham, perusahaan TLKM adalah perusahaan yang akurasinya terbaik sebesar 61.011419%.

5. SARAN

Untuk pengembangan penelitian ini lebih lanjut, maka saran yang dapat disampaikan sebagai berikut:

1. Perlu dikombinasikan dengan analisis fundamental untuk mendukung hasil yang telah diperoleh pada tahapan pengujian dengan menggunakan K-Fold Cross Validation.
2. Perlu ditambahkan teknik lain dalam penentuan nilai K terbaik pada tahapan K-Fold Cross Validation dalam pengujian data perusahaan saham yang telah ditentukan.

DAFTAR PUSTAKA

- [1] Haryuningputri, M, Widyarti, E, T. (2012), Pengaruh Rasio Profitabilitas Dan Eva Terhadap Harga Saham Pada Sektor Industri Manufaktur di Bei Tahun 2007-2010, Diponegoro Journal of Management, Vol 1 No 4, pages 67-69

- [2] Fahmi, Irham (2012), Analisis Laporan Keuangan cetakan ke-2, Alfabeta, Bandung
- [3] Zuliarni, S. (2012), Pengaruh Kinerja Keuangan Terhadap Harga Saham Pada Perusahaan Mining and Mining Service Di Bursa Efek Indonesia (BEI), Jurnal Aplikasi Bisnis, Vol 3 No. 1, page 36-48
- [4] Zamzamy, F.R, Setiawan. E, Azizah, E. N, (2018) Reaksi Sinyal Keuangan Terhadap Harga Saham Sektor Pertanian di Indonesia, Jurnal Bisnis dan Manajemen, 8, pages 133-140
- [5] Naufal, M. F. (2017). Peramalan Jumlah Wisatawan Mancanegara yang Datang ke Indonesia Berdasarkan Pintu Masuk Menggunakan Metode Support Vector Machine (SVM)
- [6] Yingjun Chen, Yontao Hao. (2017). A Feature Weighted Support Vector Machine And K-Nearest Neighbor Algorithm for Stock Market Indices Prediction. Expert Syst. Appl. 80, Pages 340-355.
- [7] Yongsheng Ding, Xinping Song, Yueming Zen. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Syst. Appl. 34, Vol 4, pages 3081-3089.
- [8] Rahmansyah, Nugraha. (2016). Analisa Algoritma Support Vector Machine (SVM) Dalam Memprediksi Nasabah Yang Berpeluang Kredit Macet, Jurnal KomTekInfo, Vol. 3, No. 1, pages 67-77.
- [9] Firdaniza, Jondri (2014). Prediksi Trend Pergerakan Harga Saham dengan Hidden Markov Model (HMM) dan Support Vector Machine (SVM), Jurnal Matematika Integratif, 10, pages 19-24.