

## PERAMALAN HARGA SAHAM MENGGUNAKAN *PRINCIPAL COMPONENT ANALYSIS* DAN *HIDDEN MARKOV MODEL*

Erlina Febriani<sup>1</sup>, Drs. Jondri, M.Si,<sup>2</sup> Dr. Deni Saepudin, S.Si.,M.Si<sup>3</sup>

<sup>1,2,3</sup>Prodi Ilmu Komputasi Telkom University

<sup>1</sup>erlinafebriani@gmail.com, <sup>2</sup>jondri@telkomuniversity.ac.id, <sup>3</sup>denisaepudin@telkomuniversity.ac.id

### Abstrak

Prediksi harga saham merupakan salah satu penelitian penting dalam bidang perekonomian. Dalam praktiknya, harga suatu saham dapat diprediksi dengan menggunakan konsep analisis teknikal. Analisis teknikal didasarkan pada prinsip penggunaan data histori harga saham untuk memprediksi pergerakan saham dimasa mendatang. Tujuan dari penelitian ini adalah mengimplementasikan metode *Principal Component Analysis* dan *Hidden Markov Model* di dalam analisis teknikal untuk memprediksi pergerakan harga saham di masa mendatang.

Sistem prediksi menggunakan data saham JKSE.JK, BBNI.JK, dan ANTAM.JK. Hasil dari percobaan dalam tugas akhir ini menunjukkan bahwa PCA dan HMM dapat bekerja dengan baik. Sistem ini memiliki MAPE terbaik sebesar 0,727%.

**Kata kunci :** *prediksi harga saham, time series, Principal Component Analysis, Hidden Markov Model*

### Abstract

*Stock price prediction is one of many important researches in economic field. In practice, stock price may be predicted with the technical analysis approach. Technical analysis based on the principle of using historical data to predict stock price movement in the future. The aim of this project is to implement of Principal Component Analysis and Hidden Markov Model in the Technical Analysis for stock price prediction.*

*The prediction system uses daily stock data of JKSE.JK, BBNI.JK, and ANTAM.JK. The result of the experiment shows that PCA with HMM performed well and gives the Mean Percentage Error (MAPE) of 0,727%.*

**Keywords:** *stock price prediction, time series, Principal Component Analysis, Hidden Markov Model*

### 1. Pendahuluan

#### 1.1 LATAR BELAKANG MASALAH

Saham adalah tanda penyertaan atau pemilikan seseorang atau badan usaha dalam suatu perusahaan atau perseroan terbatas [1]. Menginvestasikan dana dalam bentuk saham merupakan salah satu cara untuk mengendalikan dana agar aset yang dimiliki seorang investor saham adalah tetap atau menurun atau diharapkan memperoleh keuntungan pada masa yang akan datang.

Terdapat dua pendekatan analisis yang sering digunakan oleh para investor saham guna menentukan waktu terbaik untuk membeli dan menjual saham yaitu analisis fundamental dan analisis teknikal. Analisis fundamental adalah usaha untuk memperkirakan kesehatan dan prospek, yaitu kemampuan suatu perusahaan untuk bertumbuh dan

menghasilkan laba di masa depan sedangkan analisa teknikal adalah sebuah metode peramalan gerak harga saham, indeks atau instrumen keuangan dengan menggunakan harga historis yang digambarkan dalam sebuah grafik sebagai alat utama [1][3].

Indikator teknikal merupakan hasil perhitungan matematis yang *inputnya* berdasarkan pada dua hal, yaitu harga dan volume. Sebuah indikator dapat disusun dengan hanya mempertimbangkan harga, atau hanya volume, atau kombinasi dari keduanya.

Menurut Rode (1995) belum ada satupun indikator yang dijadikan sebagai pedoman berinvestasi secara pasti, karena sejauh ini belum ada indikator yang benar-benar sempurna. Hal ini membuat para analis selalu mencari-cari indikator terbaru sebagai petunjuk dalam berinvestasi. Tetapi

tidak semua indikator bisa digunakan. Ada indikator teknikal yang bisa digunakan untuk mengolah informasi dari *stock A* tetapi tidak bisa digunakan untuk mengolah informasi *stock B*. Semua indikator yang dipilih dalam *trading system* yang baik haruslah indikator yang saling mendukung dan saling melengkapi.

Dalam penelitian ini akan dilakukan peramalan harga saham dengan menggunakan dua metode yaitu *Principal Component Analysis (PCA)* dan *Hidden Markov Model (HMM)*, yang akan digunakan untuk mencari indikator yang paling merepresentasikan pola pergerakan harga saham. Dan kemudian beberapa indikator yang terpilih akan dijadikan variabel observasi pada proses HMM dalam proses peramalan harga harian saham.

**1.2 Tujuan**

Tujuan penulisan tugas akhir ini adalah sebagai berikut:

1. Mencari indikator teknikal yang tepat dengan menggunakan metode *Principal Component Analysis*.
2. Mengetahui prediksi harga saham harian dengan menggunakan metode *Hidden Markov Models* dan indikator teknikal.
3. Menguji dan menganalisis performansi dan keakuratan hasil prediksi harga harian saham dengan *Hidden Markov Models* dan indikator teknikal.

**2 Landasan Teori**

*2.1 Principal Component Analysis*

PCA adalah sebuah transformasi linear untuk menentukan sistem koordinat yang baru dari dataset. PCA berguna pada bidang pengenalan, klasifikasi dan kompresi data citra. Teknik PCA dapat mengurangi

mereduksi dimensi variabel data input menjadi komponen utama yang berdimensi lebih kecil dengan kehilangan informasi minimum, tetapi tanpa menghilangkan informasi maksimum atau informasi penting dari dataset dimana komponen utama yang terbentuk tidak berkorelasi satu dengan yang lainnya.. Proyeksi pada PCA adalah representasi himpunan data X ke dalam bentuk vektor eigen orthonormal dari matriks kovarian data X. Matriks kovarian merupakan korelasi antara variabel-variabel dalam himpunan data X. PCA merupakan proses mendapatkan vektor eigen orthonormal dari matriks kovarian sebagai basis untuk ditransformasi ke ruang data yang baru. Vektor eigen dapat dikatakan sebagai basis asli untuk multi dimensi data X. Nilai eigen terbesar dari matriks kovarian merupakan korelasi terkecil antar variabel dalam ruang data. Selanjutnya PCA akan mencari proyeksi variabel-variabel yang tidak berkorelasi.

**2.1.1 Algoritma PCA**

Misalkan terdapat dataset matriks berukuran  $n \times D$  yang terdiri dari  $n$  observasi  $x_i \in \{1, 2, \dots, D\}$  berdimensi  $D$ .

1. Hitung vektor rata-rata  $\bar{x} \in \{1, 2, \dots, D\}$  dengan :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \tag{2.9}$$

2. Hitung matriks kovariansi  $C$  atau  $cov(x)$  menggunakan persamaan :

$$C = cov(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T}{n-1} \tag{2.10}$$

3. Hitung nilai eigen  $\lambda$  dan vektor eigen  $v$  yang memenuhi persamaan :

$$|C - \lambda I| = 0 \tag{2.11}$$

$$(C - \lambda I)v = 0 \tag{2.12}$$

4. Vektor eigen yang didapatkan merupakan komponen utama untuk membentuk variabel baru, yang merupakan perkalian antara vektor eigen dengan matriks  $X$  yaitu matriks  $X$  yang telah dinormalisasi yang dihitung dengan rumus :

$$x_i = \frac{(x_i - \bar{x})}{\sigma}$$

(2.13)

dimensi dari dataset tanpa tidak menghilangkan informasi penting dari dataset. Tujuan PCA adalah

5. Variansi yang dapat dijelaskan oleh variabel baru ke-i tergantung persentase kontribusi p

dari masing-masing nilai eigen, yang dihitung dengan rumus :

$$f_i = \frac{\lambda_i}{\sum_{j=1}^D \lambda_j} \times 100\% \quad (2.14)$$

6. Penentuan jumlah variabel baru yang digunakan tergantung persentase kontribusi kumulatif dari kumulatif nilai eigen yang telah diurutkan dari nilai yang terbesar. Nilai persentase kontribusi kumulatif sampai komponen ke - r dihitung dengan rumus :

$$F_r = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^D \lambda_i} \times 100\% \quad (2.15)$$

Dengan  $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_D$

### 2.1.2 Teknik Reduksi Dimensi

Teknik reduksi dimensi mengikuti algoritma sebagai berikut :

1. Lakukan partisi dimensi himpunan X menjadi l himpunan bagian, sehingga masing-masing menjadi berukuran d dimensi dengan :

$$d = \frac{D}{l} \quad (2.16)$$

2. Lakukan untuk setiap subset data dengan d dimensi sebagai berikut :
  - i. Hitung matriks kovariansi masing-masing subset menurut persamaan 3.
  - ii. Hitung nilai eigen dan vektor eigen masing-masing subset menurut persamaan 4) dan 5).
  - iii. Ambil vektor eigen v dengan nilai eigen terbesar, dimana :
 
$$\lambda^* = \{\lambda_1, \lambda_2, \lambda_l\}$$
  - iv. Lakukan perhitungan observasi baru  $X^*$  berukuran (nx1) menggunakan :

$$X^* = (V^T X_i^T)^T \text{ atau } X^* = (X_i V)^T \quad (2.17)$$

Dengan  $i \in \{1, 2, \dots, l\}$  adalah data normalisasi dan  $v \in \{1, 2, \dots, d\}$  adalah vektor eigen yang mempunyai variansi maksimum.

- i. Gabungkan observasi baru menjadi dataset baru  $X^*$  berukuran (n x l).

## 2.2 Hidden Markov Model

Hidden Markov model merupakan perluasan *Markov Chain*. Setiap stata berkorespondensi dengan

sebuah *observable event* deterministik, *Hidden Markov Model* mencirikan sebuah proses non-deterministik yang menghasilkan keluaran simbol-

simbol observasi pada tiap stata. Observasi menjadi fungsi probabilistik dari stata. Dengan cara ini, hidden markov model dapat dianggap sebagai sebuah proses stokastik *double-embedded* dengan proses

stokastik pokok (urutan stata) yang tidak langsung *observable*.

Hidden Markov model pada dasarnya merupakan sebuah Markov chain dimana keluaran observasi merupakan sebuah variable X acak yang dihasilkan menurut suatu keluaran fungsi probabilistik yang berkaitan dengan tiap stata. Ciri-ciri HMM adalah :

1. Observasi diketahui tetapi urutan keadaan (*state*) tidak diketahui sehingga disebut *hidden*.
2. Observasi adalah fungsi probabilitas keadaan.
3. Perpindahan keadaan adalah dalam bentuk probabilitas.

Misalkan cuaca dalam satu hari dapat dikelompokkan menjadi cerah, hujan, dan berawan.

Jika seseorang dikunci dalam satu ruangan tertutup sehingga dia tidak dapat mengetahui keadaan cuaca diluar, kemudian orang tersebut disuruh menerka

keadaan cuaca, maka pengamatan yang dapat dilakukan hanyalah dengan melihat apakah orang yang masuk ke ruangan terkunci tersebut membawa

payung atau tidak. Masalah seperti ini dapat dimodelkan dalam bentuk *Hidden Markov Model (HMM)*.

2.2.1 Elemen – elemen pada HMM

HMM terdiri dari elemen-elemen dasar sebagai berikut :

1. N, adalah banyaknya *state* pada HMM. HMM terdiri dari N buah *state* dengan ruang *state*  $S=\{s_1,s_2,\dots,s_N\}$  dengan ruang *state* pada waktu t dinyatakan dengan  $Q_t$ .
2. M, adalah banyaknya simbol –simbol observasi yang berbeda pada setiap *state*. Simbol-simbol observasi individual dari setiap *state* yang akan membentuk suatu deretan observasi yang dinyatakan dengan :

$$V = \{v_1, v_2, \dots, v_M\} \quad (2.19)$$

3.  $A = \{a_{ij}\}$ , adalah distribusi peluang-peluang transisi dari *state* i menuju *state* j.

$$a_{ij} = P(s_t = j | s_{t-1} = i) ; \quad 1 \leq i, j \leq N \quad (2.20)$$

4.  $B = [b_{ij}]$  adalah distribusi peluang bersyarat simbol observasi  $v_M$  dalam

suatu *state* j matriks peluang bersyarat

observasi jika proses berada pada *state* j.

$$b_{ij} = P(v_t = v_M | s_t = j) ; \quad 1 \leq i \leq M, 1 \leq j \leq N \quad (2.21)$$

5.  $\pi$  (Distribusi peluang *state* inisial)

$$\pi_i = P(s_1 = i) \quad (2.22)$$

2.2.1 Permasalahan – permasalahan dasar HMM

Ada tiga permasalahan khusus yang dapat diselesaikan oleh metode *Hidden Markov Model*. Diantaranya :

$P(O | \lambda)$  bila diberikan urutan observasi  $O = \{O_1, O_2 \dots O_T\}$  dan sebuah model  $\lambda = (A, B, \pi)$ .

2. Masalah kedua (*Decoding*)

Memilih sekuen *state*  $\{Q = q_1 q_2 \dots q_T\}$  atau urutan keadaan yang paling optimal bila diberikan urutan observasi  $O = \{O_1 O_2 \dots O_T\}$  paling optimal dan sebuah model  $\lambda = (A, B, \pi)$ .

3. Masalah ketiga (*Learning*)

Memilih model  $\lambda$  yang sesuai agar  $P(O | \lambda)$  optimal. Untuk memperoleh model yang paling sesuai, digunakan observasi training yang akan melakukan adaptasi terhadap parameter-parameter HMM sehingga akan dihasilkan model yang terbaik.

2.3 Akurasi Sistem

Akurasi sistem yang akan digunakan yaitu *Mean Absolute Percentage Error* (MAPE), yaitu [11]:

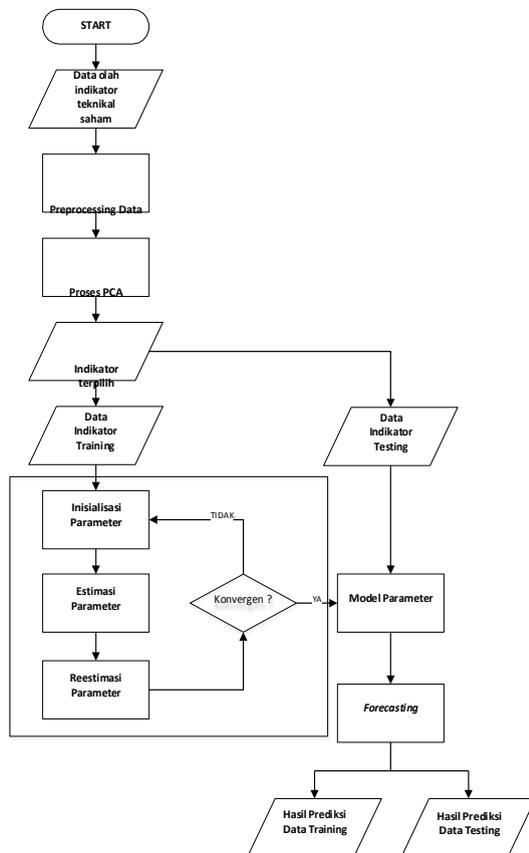
$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|a_i - p_i|}{|a_i|} \times 100 \% \quad (2.40)$$

dimana,  $a_i$  adalah harga saham aktual,  $p_i$  adalah prediksi harga saham dihari ke  $i$  dan  $n$  adalah jumlah hari yang di uji.

(2.21)

1. Masalah pertama (*Evaluation*)  
Masalah pertama dalam HMM yang harus dipecahkan adalah bagaimana caranya menghitung

### 3. Perancangan Sistem



Gambar 3-1 Flowchart proses pengolahan indikator saham

Secara umum gambar diatas dapat dijelaskan bahwa data hasil perhitungan indikator saham masuk ke dalam tahap *preprocessing* data dan kemudian masuk ke dalam tahap proses PCA untuk dilakukan reduksi data, hasil dari reduksi data adalah indikator terpilih yang kemudian dibagi menjadi dua, yaitu digunakan untuk data indikator *training* dan data indikator *testing*. Kemudian data indikator training sebagai barisan observasi di inisialisasi untuk menghasilkan parameter-parameter HMM taksiran awal, kemudian di estimasi untuk mendapatkan parameter yang sesuai. Dengan mereestimasi parameter akan diperoleh model parameter yang sesuai. Model parameter tersebut digunakan untuk proses *forecasting* (peramalan) yang nantinya akan menghasilkan hasil prediksi data *training* dan hasil prediksi data *testing*.

### 4. Analisis Hasil Pengujian

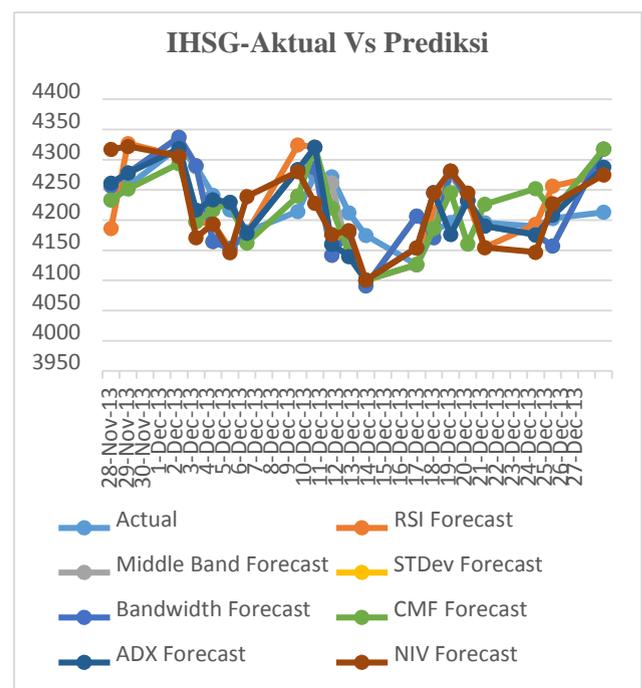
#### 4.1 IHSG (JKSE.JK)

Tabel 4. 1 Nilai Eigen, Persentase Variansi, dan Persentase Kumulatif Variansi IHSG

PC ke-	Nilai Eigen	% Variansi	% Kumulatif Variansi
1	11,269	42,490	42,490
2	7,171	27,036	69,525
3	2,646	9,975	79,500
4	2,056	7,752	87,252
5	1,247	4,703	91,955
6	1,077	4,062	96,017
7	1,056	3,983	100,000

Pada tabel terlihat bahwa dengan jumlah 7 PC yang memiliki nilai eigen lebih besar dari 1 dengan persentase kumulatif variansi yaitu 100%, artinya 31 PC tersebut dapat mewakili dari total keseluruhan dimensi yaitu 31 dimensi serta dapat menjelaskan penyebaran data sebesar 100%. Persentase kumulatif variansi yang menunjukkan 100% artinya proses reduksi dengan metode PCA tidak menghilangkan informasi di dalamnya.

Berdasarkan analisis PCA tersebut maka indikator RSI, *Middle Band 20-Day SMA*, *20-Day Standar Deviation*, *Bandwidth*, *Chaikin Money Flow*, *ADX* dan *NIV* yang dipakai dalam proses pengolahan selanjutnya, indikator-indikator tersebut digunakan untuk membuat *observation sequences* pada proses HMM.



Gambar 4. 1 Grafik Data Testing IHSG-Aktual VS Prediksi

Tabel 4. 2 Mean Absolute Percentage Error (MAPE) IHSB

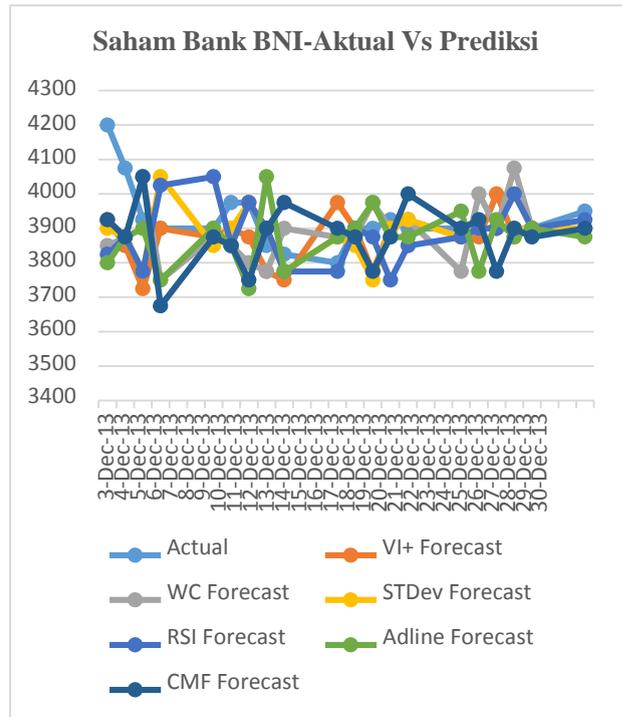
	Training	Testing
RSI	<b>0,727</b>	<b>0,772</b>
Middle Band 20-Day SMA	1,096	1,135
20-Day Standar Deviation	1,374	1,437
Bandwidth	1,567	1,666
Chaikin Money Flow	1,779	1,900
ADX	1,997	2,064
NIV	2,167	2,242

Pada Tabel 4.2 diberikan nilai MAPE untuk semua data IHSB, baik data *training* dan data *testing*. Indikator RSI memiliki nilai mape terbaik pada data *training* sebesar 0,727% dan pada data *testing* sebesar 0,772%.

**4.2 Saham BNI (BBNI.JK)**

Tabel 4. 3 Nilai Eigen, Persentase Variansi, dan Persentase Kumulatif Variansi Saham BNI

PC ke-	Nilai Eigen	% Variansi	% Kumulatif Variansi
1	9,7734	40,869	40,869
2	5,5830	23,347	64,216
3	3,1321	13,097	77,314
4	2,5155	10,519	87,833
5	1,7015	7,115	94,95
6	1,2079	5,051	100,00



Gambar 4. 2 Grafik Saham Data Testing Bank BNI-Aktual VS Prediksi

Tabel 4. 4 Mean Absolute Percentage Error (MAPE) Saham BNI

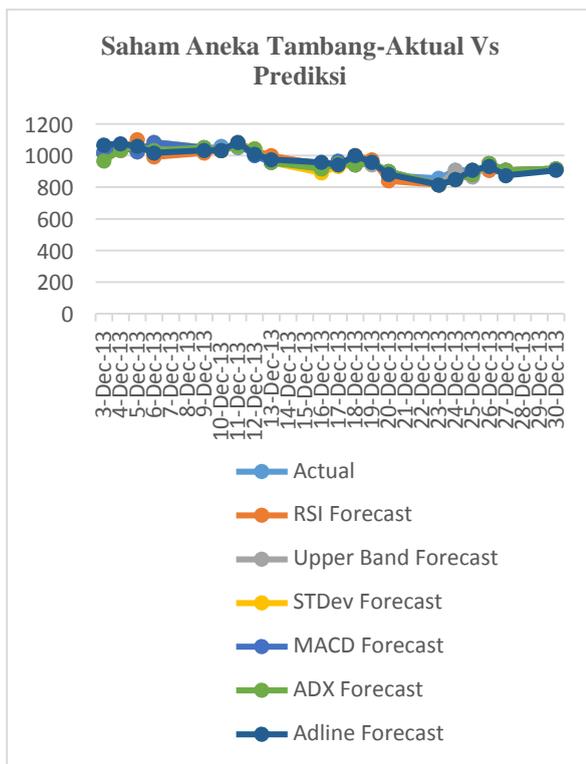
	Training	Testing
<i>Vortex Indicator+</i>	0.742	1.367
<i>Weighted Close</i>	1.175	2.236
<i>20-Day Standar Deviation</i>	1.381	2.646
<i>RSI</i>	1.628	3.368
<i>Adline</i>	1.792	3.978
<i>Chaikin Money Flow</i>	1.968	4.293

Pada Tabel 4.4 diberikan nilai MAPE untuk semua data saham BNI, baik data *training* dan data *testing*. Indikator *Vortex Indicator+* memiliki nilai MAPE terbaik pada data *training* sebesar 0,742% dan pada data *testing* sebesar 1,367%.

4.3 Saham Aneka Tambang (ANTAM.JK)

Tabel 4. 5 Nilai Eigen, Persentase Variansi, dan Persentase Kumulatif Variansi Saham Antam

PC ke-	Nilai Eigen	% Variansi	% Kumulatif Variansi
1	10.894	42,424	42,424
2	5,717	22,261	64,684
3	4,091	15,931	80,615
4	2,186	8,514	89,129
5	1,759	6,850	95,980
6	1,033	4,021	100,00



Gambar 4. 3 Grafik Saham Data Testing Aneka Tambang -Aktual VS Prediksi

Tabel 4. 6 Mean Absolute Percentage Error (MAPE) Saham Aneka Tambang

	Training	Testing
RSI	2,765	2,188
Upper Band 20-Day SMA	3,405	3,235
20-Day Standar Deviation	4,090	4,324
MACD	4,644	5,015
ADX	5,187	5,772
Adline	6,381	7,412

Pada Tabel 4.6 diberikan nilai MAPE untuk semua data saham Aneka Tambang, baik data *training* dan data *testing*. Indikator RSI memiliki nilai MAPE terbaik pada data *training* sebesar 2,765% dan data *testing* sebesar 2,188%.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil analisis dan pengujian pada bab sebelumnya, dapat disimpulkan:

1. Hidden Markov Model dapat diimplementasikan sebagai prediksi harga saham dengan hasil untuk IHSG dan Saham Aneka Tambang memiliki ketepatan prediksi yang lebih baik pada indikator yang sama, yaitu RSI. Hal ini ditunjukkan pada hasil MAPE data Training dan Testing, yaitu Indikator RSI pada IHSG memiliki MAPE untuk data training = 0,727% dan MAPE untuk data testing = 0,772%, dan Indikator RSI pada saham Aneka Tambang memiliki MAPE untuk data training = 2,765% dan MAPE untuk data testing = 2,188%.
2. Hidden Markov model pada Saham Bank BNI memiliki ketepatan prediksi harga saham lebih baik pada indikator ke-1, yaitu *vortex indicator+*, MAPE untuk data training = 0,742% dan MAPE untuk data testing = 1,367 %.
3. MAPE yang dihasilkan merupakan *Excellent Forecasting* karena lebih kecil dari 10%.

5.2 Saran

Berdasarkan percobaan yang dilakukan pada bab sebelumnya, Saran untuk penelitian selanjutnya adalah sebaiknya dilakukan penelitian dengan jumlah PC yang berbeda, jumlah saham dan indikator diperbanyak, dan prediksi harga saham bisa dilakukan secara *real time*.

**Daftar Pustaka :**

- [1] Badge, Jyoti. 2012. *Forecasting of Indian Stock Market by Effective Macro-Economic Factors and Stochastic Model*. India : Scienpress Ltd.
- [2] Suhartono & Fadhillah Qudsi. 2009. *Portofolio & Bursa Efek*. Yogyakarta : Sekolah Tinggi Ilmu Manajemen YKPN.
- [3] Wira, Desmond. 2012. *Analisis Teknikal untuk Profit Maksimal*. Exceed.
- [4] MA, Xueying. 2011. *PCA-Fuzzy-SVR Stock Price Prediction*. Amsterdam : University of Amsterdam.
- [5] Berlinti, Rosalina. 2006. *Peramalan Time Series Indeks Harga Saham Gabungan Menggunakan Elman Recurrent Neural Network dengan Algoritma Backpropagation Through Time*. Bandung: Tugas Akhir Jurusan Teknik Informatika Sekolah Tinggi Teknologi Telkom.
- [6] Sadeq, Ahmad.2008. *Analisis Prediksi Indeks Harga Saham Gabungan Dengan Metode Arima*. Tesis. Program Magister Manajemen Pascasarjana Universitas Dipenogoro.
- [7] Mustafidah, Anifatul. 2009. *Peramalan Harga Saham Menggunakan Jaringan Saraf Tiruan Metode Probabilistik*. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Maulana Malik Ibrahim.
- [8] Lendasse, Amaury & dkk. 2001. *Dimension reduction of technical indicators for the prediction of financial time series- Application to the BEL20 Market Index*. *European Journal of Economic and Social Systems*.
- [9] Firdaniza, dkk. 2006. *Hidden Markov Model*. Dosen Jurusan Matematika FMIPA Universitas Padjadjaran. Bandung : Unpad.
- [10] Rabiner. L.R.1989. *A Tutorial On Hidden Markov Models And Selected Applications In Speech Recognition*. *Proceedings of the IEEE*.
- [11] Gupta, Aditya dan Dhingra, Bhuwan. 2012. *Stock Market Prediction Using Hidden Markov Model*. *Proceedings of the IEEE*.
- [12] Handayani, Vivi Putri. 2013. *Klasifikasi Penyakit Kanker Prostat menggunakan Principal Component Analysis (PCA) dan Least Squares Support Vector Machine (LS-SVM)*. Bandung: Institut Teknologi Telkom.
- [13] Solihati, Tifani Intan. 2014. *Peramalan Harga Saham Menggunakan Hidden Markov Models*. Bandung: Universitas Telkom.
- [14] Yendriani, Devy. 2015. *Prediksi Harga Saham Menggunakan Hidden Markov Model (HMM) dan Fuzzy Model*. Bandung: Universitas Telkom.
- [15] Zhang, Yingjian. 2001. *Prediction Of Financial Time Series With Hidden Markov Models*. China : Shandong University.