

# Studi Literatur Mengenai Prediksi Harga Saham Menggunakan Machine Learning

Kartika Maulida H.<sup>1</sup>, I Gede Susrama Mas Diyasa<sup>2</sup>, Prismahardi Aji R.<sup>3</sup>, Tresna Maulana F<sup>4</sup>

<sup>1,2,3,4</sup> Sains Data, UPN “Veteran” Jawa Timur

<sup>3</sup>[prismahardi.aji.ds@upnjatim.ac.id](mailto:prismahardi.aji.ds@upnjatim.ac.id)

<sup>4</sup>[tresna.maulana.ds@upnjatim.ac.id](mailto:tresna.maulana.ds@upnjatim.ac.id)

<sup>2</sup> Informatika, UPN “Veteran” Jawa Timur

<sup>2</sup>[igsusrama.if@upnjatim.ac.id](mailto:igsusrama.if@upnjatim.ac.id)

\*Corresponding author email: [kartika.maulida.ds@upnjatim.ac.id](mailto:kartika.maulida.ds@upnjatim.ac.id)

**Abstrak**— Prediksi harga saham tidak hanya menarik bagi para ekonom namun juga menarik dibahas untuk para analis. Harga saham yang berfluktuasi dan terpengaruhi oleh banyak faktor menyebabkan prediksi dengan berbagai macam teknik dilakukan. *Machine Learning* adalah salah satu teknik yang sering digunakan untuk memprediksi harga saham. Pada penelitian ini membahas mengenai studi literatur variabel masukan, metode, dan hasil dari literature terdahulu. Harga pembukaan, penutupan, terendah, dan tertinggi saham harian merupakan variabel masukan yang paling banyak digunakan. Metode yang paling banyak digunakan adalah jaringan syaraf tiruan. Metode tersebut juga menghasilkan tingkat kesalahan yang relatif kecil.

**Kata Kunci**— Prediksi Harga Saham, *Machine Learning*, Regresi, Jaringan Syaraf Tiruan, Studi Literatur

penelitian ini akan membahas mengenai variabel-variabel yang digunakan dalam memprediksi harga saham, machine learning yang digunakan, dan hasil dari prediksi saham.

## Tujuan Penelitian

Mengetahui berbagai variabel masukan, metode yang digunakan dan hasil dari penelitian yang telah dilakukan oleh peneliti lain merupakan tujuan dari penelitian ini.

## Manfaat Penelitian

Manfaat penelitian ini adalah memberikan kemudahan kepada peneliti pemula untuk dapat melakukan penelitian pada bidang prediksi harga saham dengan menggunakan *machine learning*.

## II. METODOLOGI

Penelitian ini adalah studi literatur yang merangkum dari beberapa literatur relevan mengenai prediksi harga saham dengan menggunakan machine learning. Metodologi yang digunakan dalam studi literatur ini yaitu identifikasi, seleksi, dan kelayakan. Bagan metodologi dapat dilihat pada gbr 1.

Langkah pertama yang dilakukan yaitu identifikasi literatur melalui pencarian situs IEEE, SpringerLink, dan Google Scholar. Dari ketiga situs tersebut berhasil dikumpulkan sebanyak 33 literatur. Kata kunci yang digunakan dalam pencarian literatur yaitu “*Stock Price*”, “*Stock Market Regression*”, “*Stock Market Price*”, “Prediksi Saham”, “Prediksi Harga Saham”, dan “Pergerakan Harga Saham”.

Setelah literatur terkumpul, penyeleksian literatur dilakukan. Literatur diseleksi melalui judul dan tahun publikasi. Judul literatur dipertimbangkan apakah sesuai dengan tema prediksi saham menggunakan teknik data mining atau tidak. Penyeleksian tahun publikasi literatur dibatasi dari tahun 2000 hingga 2020.

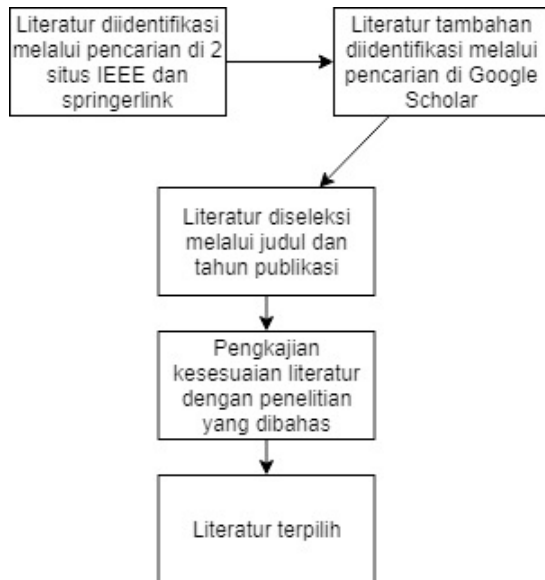
Setelah penyeleksian literatur melalui judul dan tahun publikasi, literatur yang tersisa dikaji. Pengkajian literatur meliputi kesesuaiannya dengan penelitian yang dibahas dan ketersediaan artikel secara *full text*. Tahap terakhir adalah literatur terpilih akan dibahas pada penelitian ini. Literatur terpilih yang akan dibahas sebanyak 15 artikel.

## I. PENDAHULUAN

Prediksi harga saham merupakan topik yang menarik dalam bidang machine learning. Hal ini dikarenakan faktor-faktor yang mempengaruhi harga saham tidak dapat diprediksi. Beberapa penelitian terdahulu mengacu pada analisis fundamental perusahaan dan sentiment analisis baik dari berita maupun dari sosial media. Kinerja perusahaan, tingkat kepercayaan investor, kondisi negara, peraturan pemerintah, tingkat inflasi, dll merupakan factor-faktor yang biasanya mempengaruhi keputusan investor [1].

Penanaman modal di saham merupakan salah satu bentuk investasi oleh investor. Asal investor tidak hanya berasal dari dalam negeri tetapi juga dapat dari luar negeri. Hal ini mengakibatkan banyak faktor yang mempengaruhi pergerakan harga saham. Ketika melakukan investasi saham, investor dan pelaku bisnis memerlukan informasi yang lengkap, relevan, akurat, dan tepat waktu. Informasi tersebut dapat mendorong investor dalam mengambil keputusan investasi [2].

Informasi-informasi yang tersedia yang dapat mendukung keputusan investor sudah banyak tersebar. Perusahaan yang terdaftar pada bursa saham harus melaporkan laporan keuangannya. Selain itu, informasi dari berita, kebijakan perusahaan maupun pemerintahan, dan opini investor terhadap suatu saham dapat terekam di media sosial. Oleh karena itu, telah banyak penelitian yang menggunakan machine learning untuk memprediksi pergerakan harga saham. Sehingga pada



Gbr 1. Metodologi Penelitian

### III. HASIL DAN PEMBAHASAN

Daftar singkatan yang akan digunakan terdapat pada tabel 1.

TABEL I  
DAFTAR SINGKATAN

No	Singkatan	Kepanjangan dari
	IHS	Indeks Harga Saham
1	SVR	<i>Support Vector Regression</i>
2	JST	Jaringan Syaraf Tiruan
	MLR	<i>Multiple Linear Regression</i>
3	MA	<i>Moving Average</i>
4	WMA	<i>Weighted Moving Average</i>
5	RMSE	<i>Root Mean Square Error</i>
	MAE	<i>Mean Absolute Error</i>
6	MAPE	<i>Mean Absolute Percentage Error</i>
7	RSI	<i>Relative Strength Index</i>
8	MACD	<i>Moving average Convergence Divergence</i>
	CCI	<i>Commodity Channel Index</i>
9	ATR	<i>Average True Range</i>
10	NTA	<i>Net Tangible Asset</i>
11	LA	<i>Liquid Asset</i>
12	DE	<i>Debt to Equity</i>
13	ZS	<i>Altman Z-Score</i>
14	AT	<i>Asset Turnover</i>

Daftar literatur yang digunakan terdapat pada tabel2.

TABEL III  
DAFTAR LITERATUR

Auth or	Variabel Independent	Variabel Dependent	Metode	Hasil
[1]	Harga tertinggi, harga terendah, harga	Prediksi harga saham periode berikutnya	<i>Backpropagation</i> JST	Margin error masih terlalu besar sehingga diperlukan

Auth or	Variabel Independent	Variabel Dependent	Metode	Hasil
	penutup suatu saham pada 1 hari			data yang lebih banyak untuk mengetahui pola prediksi
[3]	IHS penutupan harian, jumlah saham yang diperdagangkan, laba perusahaan, dan kondisi perusahaan	Prediksi harga saham	Jaringan Syaraf Tiruan	Nilai RMSE kecil yaitu 0.00244
[4]	Harga pembuka, penutup, tertinggi dan terendah saham selama 3 tahun	Prediksi IHSG ke depan	Gabungan SVR dan JST	Tingkat MAPE (175,41) dan RMSE (3,16%) relatif kecil
[5]	Harga pembuka, penutup, tertinggi dan terendah saham selama 1 tahun	Prediksi harga penutupan saham	<i>Backpropagation Neural Network</i>	Nilai RMSE (0,0626) dan MAE (0,0456) kecil
[6]	Harga pembuka, penutup, tertinggi, terendah, dan jumlah saham yang diperdagangkan harian selama 2 tahun	Prediksi harga saham	MLR dengan <i>K-Means</i> dan <i>MA</i> .	Nilai MSE (15087.465), RMSE (122.831), dan MAPE (3.255) cukup kecil
[7]	Histori harga saham selama 3 tahun, indikator analisa teknikal (MA10, WMA10, Momentum, RSI, MACD, BIAS10, PSY10)	Prediksi pergerakan harga saham	SVR	Nilai RMSE cukup baik (892.01)
[8]	Harga penutupan saham dan indikator (Simple MA,	Prediksi harga saham	SVR dengan <i>feature selection</i> menggunakan	92% akurasi.

Author	Variabel Independent	Variabel Dependent	Metode	Hasil	Author	Variabel Independent	Variabel Dependent	Metode	Hasil
	Exponential MA, momentum, rate of change, MACD, CCI, RSI, and stochastic % K and stochastic % D)		an PSO			dan AT.			menunjukkan kinerja yang paling baik. Nilai MAE (0.5462) dan RMSE (0.8164) kecil
[9]	Harga pembuka, penutup, tertinggi, terendah, dan jumlah saham yang diperdagangkan harian selama 2 tahun	Prediksi harga saham	Multiple Regression	Akurasi Prediksi tinggi pada ketiga dataset yaitu 95%, 89% dan 97%	[14]	Tick-by-tick data (waktu, harga, perubahan harga, volume, jumlah, sinyal jual atau beli)	Prediksi tren harga saham	Outlier miming algorithm dan K-Means Clustering	Tingkat kesuksesan tepat dan memuaskan
[10]	Harga penutupan saham dan berita di web selama 4 tahun	Prediksi harga saham	Recurrent Neural Network	MAPE 2.03-2.17. Korelasi tinggi antara berita dengan pergerakan harga saham	[15]	Tanggal, harga penutupan, volume, total harga transaksi, dan data sosial media	Prediksi harga saham	SVM	SVM dengan sentiment index memiliki akurasi yang lebih tinggi daripada metode SVM saja
[11]	Harga pembuka, penutup, tertinggi, terendah saham selama 6 tahun, dan indicator (Exponential MA, MACD, ATR, dan RSI)	Prediksi pergerakan harga saham	Logistic Regression dan SVM	Memberikan kinerja yang baik dibandingkan metode Perceptron. Error measurement 10.01-16.31	[16]	Harga saham, sosial media, dan berita	Prediksi saham	Neural Network dengan pengaplikasian Deep Learning	Akurasi yang lebih tinggi 83,22%
[12]	Harga pembukaan, penutupan saham dan status di twitter	Prediksi harga saham	Random Forest dan Naïve Bayes	Prediksi dengan sentimen analisis memberikan hasil yang cukup akurat. Akurasi paling tinggi dengan data social media adalah 67.37%					
[13]	NTA, LA, DE, ZS	Prediksi tren saham	Regression	SMO Regression					

#### A. Variabel

Variabel input untuk memprediksi harga saham dapat berbentuk bermacam-macam. Variabel yang paling umum digunakan adalah data harga saham dari harga pembukaan, penutupan, terendah, tertinggi, volume, dan total harga transaksi harian. Analisis mendasar menggunakan laporan keuangan perusahaan juga dapat dilakukan dengan memperhatikan variabel NTA, LA, DE, ZS, dan AT. Indikator yang dibentuk dari harga penutupan saham juga dapat dipertimbangkan sebagai variabel penentu. Indikator yang paling sering digunakan adalah Simple MA, Exponential MA, momentum, MACD, dan RSI. Selain data kuantitatif yang telah disebutkan, variabel penentu harga saham juga dapat berupa data kualitatif. Data tersebut didapatkan dari sosial media dan berita-berita mengenai saham. Durasi atau rentang waktu data yang diambil dapat bermacam-macam dari 2 minggu hingga 6 tahun. Semakin banyak data masukan, diharapkan semakin dapat membentuk pola sebaran untuk prediksi harga saham.

Formula Simple MA, dimana  $C_t$  adalah harga penutupan saham dapat dilihat pada (1)

$$SMA = (1/N) \sum_{k=1}^N C_t \quad (1)$$

Formula Exponential MA dapat dilihat pada (2)

$$EMA = (C_t \times \alpha) + (EMA_{t-1} \times (1 - \alpha)), c > n(2)$$

Formula Momentum dapat dilihat pada (3)

$$MOM = C_t - C_{t-x} \quad (3)$$

Formula MACD dapat dilihat pada (4)

$$MACD = EMA_{12} - EMA_{26} \quad (4)$$

Formula RSI dapat dilihat pada (5)

$$RSI = 100 - (100 / (1 + (\sum_{i=1}^{n-1} Up_{t-1} / n) / (\sum_{i=1}^{n-1} Dw_{t-1} / n))) \quad (5)$$

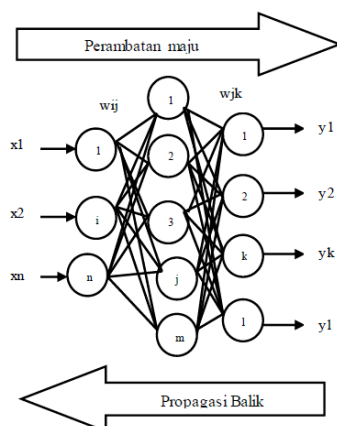
#### B. Metode

Perhitungan yang digunakan untuk mengetahui kepopuleran suatu metode pada penelitian ini dapat dilihat pada tabel III.

TABEL IIIII  
DAFTAR METODE

No	Metode yang Digunakan	Jumlah
1.	JST / ANN	4
2.	SVR	3
3.	Multiple Linear Regression	2
4.	K-Means	2
5.	Moving Average	1
6.	Recurrent Neural Network	1
7.	Logistic Regression	1
8.	SVM	2
9.	Random Forest	1
10.	Naïve Bayes	1
11.	SMO Regression	1
12.	Outlier Mining Algorithm	1
13.	Deep Learning	1

Metode yang paling populer pada studi literatur ini adalah JST / ANN. JST merupakan metode pembelajaran yang dibimbing (*supervised learning*). Hal yang unik dari JST adalah terdiri dari banyak layer seperti *input layer*, *hidden layer*, dan *output layer*. Pembaruan bobot pada layer-layer dari *output layer* sampai kembali ke *input layer* dinamakan *BackPropagation*.



Gbr 2. Arsitektur ANN BackPropagation [5]

SVM dan SVR juga populer pada studi literatur ini. SVM dan SVR memiliki konsep yang sama. Hal yang membedakan adalah keluaran hasilnya. Support Vector membentuk *hyperplane* yang mengklasifikasikan data dengan memisahkan secara linier dalam dimensi yang lebih tinggi [11].

Selain itu, Multiple Linear Regression dan K-Means juga termasuk salah satu metode yang populer. Multiple Linear Regression terdiri dari lebih dari satu variabel  $X$  dan variabel  $Y$  [17]. Keduanya merupakan kuantitatif data.  $\beta$  adalah koefisien, maka MLR dinotasikan pada (6)

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{in}\beta_n \quad (6)$$

#### IV. KESIMPULAN

Variabel yang digunakan untuk menjadi data masukan prediksi harga saham dapat bermacam-macam. Variabel yang paling banyak digunakan yaitu harga pembuka, penutup, terendah, dan tertinggi saham pada satu hari. Metode yang paling populer yaitu metode jaringan syaraf tiruan / ANN. Pengaplikasian *deep learning* pada *machine learning* juga digunakan untuk prediksi harga saham. Dari studi literatur yang dilakukan, metode yang menghasilkan nilai RMSE, MAE, dan MAPE paling sedikit adalah JST. Penelitian ke depannya dapat melakukan perbaikan pada metode-metode lain agar dapat meminimalkan nilai error dan meningkatkan akurasi.

#### REFERENSI

- [1] R. B. Purnama, "Perancangan Prediksi Untuk Menentukan Indeks Harga Saham Menggunakan Jaringan Syaraf Tiruan," *Kinetik*, vol. 2, no. 2, p. 125, 2017.
- [2] D. Arista and Astohar, "ANALISIS FAKTOR – FAKTOR YANG MEMPENGARUHI RETURN SAHAM (Kasus pada Perusahaan Manufaktur yang Go Public di BEI periode tahun 2005 - 2009)," *J. Ilmu Manaj. dan Akunt. Terap.*, vol. 14, no. 1, p. 29, 2013.
- [3] E. Mala, S. Rochman, A. Djunaidy, T. Informatika, and F. T. Informasi, "Prediksi harga saham yang mempertimbangkan faktor eksternal menggunakan jaringan saraf tiruan," vol. 1, no. 2, pp. 5–11, 2014.
- [4] L. E. Siahaan, R. F. Umbara, S. Si, and Y. Sibaroni, "Prediksi Indeks Harga Saham dengan Metode Gabungan Support Vector Regression dan Jaringan," vol. 2, pp. 21–30, 2017.
- [5] A. Novita, "Prediksi Pergerakan Harga Saham Pada Bank Terbesar Di Indonesia Dengan Metode Backpropagation Neural Network," pp. 965–972, 2013.
- [6] A. Izzah and R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier," vol. 2, no. 3, pp. 141–149, 2017.
- [7] L. Y. Kurniawati, H. Tjandrasa, and I. Ariesanti, "Prediksi pergerakan harga saham menggunakan support vector regression," *J. SimanteC*, vol. VIII, no. 1, pp. 11–21, 2013.
- [8] Z. Rustam and P. Kintandani, "Application of Support Vector Regression in Indonesian Stock Price Prediction with Feature Selection Using Particle Swarm Optimisation," *Model. Simul. Eng.*, vol. 2019, 2019.
- [9] M. Z. Asghar, F. Rahman, F. M. Kundi, and S. Ahmad, *Development of stock market trend prediction system using multiple regression*, vol. 25, no. 3. Springer US, 2019.
- [10] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," *Proc. - 5th IEEE Int. Conf. Big Data Serv. Appl. BigDataService 2019, Work. Big Data Water Resour. Environ. Hydraul. Eng. Work. Medical, Heal. Using Big Data Technol.*, pp. 205–208, 2019.
- [11] I. R. Parray, S. S. Khurana, M. Kumar, and A. A. Altalbe, "Time series data analysis of stock price movement using machine

- learning techniques,” *Soft Comput.*, vol. 0123456789, 2020.
- [12] T. Cakra, “Stock Price Prediction using Linear Regression based on Sentiment Analysis,” *Int. J. Sci. Eng. Res.*, vol. 6, no. 3, pp. 1655–1659, 2015.
- [13] H. L. Siew and M. J. Nordin, “Regression techniques for the prediction of stock price trend,” *ICSSBE 2012 - Proceedings, 2012 Int. Conf. Stat. Sci. Bus. Eng. "Empowering Decis. Mak. with Stat. Sci.*, pp. 99–103, 2012.
- [14] L. Zhao and L. Wang, “Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm,” *Proc. - 2015 IEEE 5th Int. Conf. Big Data Cloud Comput. BDCloud 2015*, pp. 93–98, 2015.
- [15] Y. Wang and Y. Wang, “Using social media mining technology to assist in price prediction of stock market,” *Proc. 2016 IEEE Int. Conf. Big Data Anal. ICBDA 2016*, 2016.
- [16] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, “Stock market prediction using machine learning classifiers and social media, news,” *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, 2020.
- [17] D. J. Olive, *Linear regression*. 2017.