

# Perbandingan Prediksi Harga Saham Menggunakan Metode SVR, RFR, dan DTR

Royan Agil Nugroho<sup>1</sup>, St. Rahmawati<sup>2</sup>, Diyah Kresna Ani A<sup>3</sup>  
Universitas Ahmad Dahlan, Yogyakarta<sup>1,2,3</sup>  
*joko@math.uad.ac.id*

**Abstrak**—Saham adalah surat berharga yang dikeluarkan oleh suatu perusahaan. Dalam hal ini, perlu diprediksi harga saham sekarang yang didasarkan pada harga saham sebelumnya. Data yang digunakan adalah data *time series* saham PT. Kimia Farma, Tbk. dan diprediksi menggunakan metode *Decision Tree Regression (DTR)*, *Random Forest Regression (RFR)*, dan *Support Vector Regression (SVR)*. Penelitian ini bertujuan untuk membandingkan tingkat keakuratan dari metode SVR, RFR, dan DTR. Penelitian ini menunjukkan hasil perbandingan dari ketiga metode tersebut dalam memprediksi harga saham secara akurat. Hasil penelitian menunjukkan bahwa untuk data saham ini model prediksi dengan akurasi terbaik adalah RFR dengan *score MSE-test* 13443.61 dan *R2 score-test* 0.19.

**Kata kunci** : *Prediksi, Harga Saham, Decision Tree Regression, Random Forest Regression, dan Support Vector Regression*

## I. PENDAHULUAN

Pasar Modal merupakan sistem keuangan yang terorganisir, komponen utama dari pasar modal adalah bank komersial, lembaga perantara di bidang keuangan, dan surat-surat berharga yang beredar[1]. Di dalam pasar modal sesuai dengan UU no. 8 tahun 1995 tentang Pasar Modal maka objek yang dijadikan sebagai kegiatan jual beli pasar modal adalah berupa surat-surat berharga yang disebut dengan efek. Efek ini adalah surat berharga berupa surat pengakuan utang, saham, obligasi, surat berharga komersial, tanda bukti utang, dan setiap derivatif dari efek[2].

*International Monetary Fund (IMF)* menyebutkan bahwa memasuki kondisi pandemi saat ini, dunia sedang berada di situasi krisis ekonomi yang bahkan lebih parah dari tahun 2008, tidak hanya berdampak pada pasar uang, namun juga pada pasar modal secara signifikan. Indeks harga saham di dunia menurun drastis, termasuk IHSG di Indonesia. IHSG di Indonesia pernah menyentuh level terendahnya yakni di angka Rp 3.911,71. Saat pandemi Covid 19 ini, terjadi penurunan harga saham yang signifikan pada industri perbankan di Indonesia dan terdapat perbedaan yang signifikan terhadap harga saham ketika sebelum terjadi pandemi Covid 19, saat diumumkannya kasus Covid 19 Pertama di Indonesia, dan 3 bulan pasca diumumkannya kasus Covid 19[3].

Salah satu aset untuk investasi adalah saham. Saham adalah surat berharga yang dikeluarkan oleh suatu perusahaan. Pemegang saham memperoleh pendapatan dari perusahaan yang menerbitkan saham tersebut. Jika penerbit saham (emiten) mampu menghasilkan keuntungan yang besar maka keuntungan yang didapatkan pemegang saham juga semakin besar. Semakin tinggi keuntungan yang ditawarkan, semakin tinggi risiko yang akan dihadapi investor. Oleh karena itu, perlu diprediksi harga saham sekarang yang didasarkan pada harga saham sebelumnya.[4]

Penelitian ini dilakukan untuk memprediksi dan mengidentifikasi data *time series* saham yang diterbitkan secara berkala setiap hari. Identifikasi berguna untuk menemukan pengetahuan dalam data. Berbagai pendekatan untuk prediksi digunakan dalam penelitian ini yaitu teknik *machine learning*[5].

Dalam *machine learning*, salah satu komponen penting dalam membangun algoritma adalah *data mining*. *Data mining* merupakan suatu proses penemuan pengetahuan atau deteksi pola dalam kumpulan data yang besar[6]. Pengolahan data ini bertujuan untuk mengekstrak informasi dan menampilkan visualisasi pada data. Fungsi utama data mining ialah untuk mendeskripsikan dan memprediksi data. Proses pengumpulan dan ekstraksi informasi dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan matematika, statistika, atau teknologi *artificial intelligence*(AI) [7]. Metode *data mining* yang digunakan dalam penelitian ini antara lain yaitu *Support Vector Regression*, *Random Forest Regression*, dan *Decision Tree Regression*.

*Support Vector Regression* merupakan metode regresi dari *Support Vector Machine* yang biasa digunakan untuk mengatasi *overfitting* dan memiliki performa yang baik untuk kasus regresi. Akan tetapi, untuk mendapatkan performa yang baik dibutuhkan pula parameter yang tepat. Selain membutuhkan parameter yang tepat, penentuan fitur adalah salah satu hal yang penting untuk melatih *Support Vector Regression*. Maka dari itu dibutuhkan suatu metode optimasi untuk mencapai performa model *Support Vector Regression* yang akurat [8]. Konsep *Support Vector Regression* didasarkan pada *risk minimization*, yaitu untuk mengestimasi suatu fungsi dengan cara meminimalkan batas atas dari *generalization error*, sehingga *Support Vector Regression* mampu mengatasi *overfitting* [4].

*Random Forest* merupakan salah satu algoritma yang melakukan pengklasifikasian yang terdiri dari beberapa pohon keputusan. Setiap pohon keputusan dibangun dengan menggunakan vektor acak. Pendekatan umum yang digunakan untuk menyisipkan vektor acak dalam pembentukan pohon [9]. *Random Forest* juga merupakan sebuah model *ensemble* yang istimewa sebab memiliki *bagging* di dalamnya. Metode *bagging* ini membuat model RF menjadi lebih tahan terhadap kemungkinan *overfit* dan juga lebih stabil [10]. Meskipun dalam praktiknya metode ini sangat sulit untuk didefinisikan namun memiliki tingkat optimasi yang baik dari pada metode-metode lain yang sudah dilakukan penelitian.

*Decision Tree* merupakan salah satu Algoritma Klasifikasi dan Prediksi. Metode ini bekerja dengan cara mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan yang dapat dengan mudah dipahami dengan bahasa alami [11]. *Decision Tree* juga berfungsi untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara sejumlah variabel input dengan variabel target lainnya.

Ada beberapa penelitian tentang prediksi dengan menggunakan beberapa model algoritma yang mendukung penelitian ini. Antara lain [12] telah meneliti tingkat akurasi prediksi tertinggi harga saham dengan membandingkan dua algoritma yaitu *Naive Bayes* dan *Decision Tree*. Hasil penelitian menunjukkan algoritma *Decision Tree* lebih baik dibanding *Naive Bayes*. Pada dasarnya prediksi harga saham sudah dilakukan dengan menggunakan beberapa metode, namun dengan menggunakan ketiga metode ini diharapkan dapat terlihat metode mana yang menghasilkan prediksi yang baik dan optimal karena karakteristik dari data saham merupakan data *time series* yang bergerak kontinu terhadap waktu [13].

## II. METODE PENELITIAN

Analisis regresi adalah alat statistik untuk menyelidiki hubungan antar variabel. Biasanya, penyidik berusaha untuk memastikan efek kausal dari satu variabel terhadap variabel lain efek kenaikan harga terhadap permintaan, misalnya, atau efek perubahan jumlah uang beredar terhadap tingkat inflasi[14].

Metode penelitian merupakan suatu tata cara atau langkah-langkah yang digunakan untuk kepentingan penelitian. Data yang digunakan adalah data *time series* saham PT. Kimia Farma, Tbk. yang diperoleh dari *finance.yahoo.com* dengan menggunakan metode *Support Vector Regression*, *Random Forest Regression* dan *Decision Tree Regression* untuk memprediksi akurasi terbaik.



GAMBAR 1. DIAGRAM ALIR PENELITIAN

### Support Vector Regression

*Support Vector Regression* merupakan penerapan *Support Vector machine (SVM)* untuk kasus regresi, *output* yang dihasilkann berupa bilangan riil atau kontinu. *Support Vector Regression* merupakan metode yang dapat mengatasi *overfitting*, sehingga akan menghasilkan performa yang bagus[15]. *Support Vector Regression* digunakan untuk menemukan suatu fungsi  $f(x)$  yang mempunyai deviasi  $\varepsilon$  paling besar dari nilai aktual  $y$ , untuk semua data training. Jika nilai  $\varepsilon = 0$ , maka kita dapatkan suatu regresi yang sempurna [16].

Misalkan  $\lambda$  set data *training*,  $(x_j, y_j)$  dimana  $j = 1, 2, \dots, \lambda$  dengan *input*  $x = \{x_1 x_2 x_3\} \dots \subseteq \mathcal{R}^N$  dan *output* yang bersangkutan  $y = \{y_1, \dots, y_\lambda\} \subseteq \mathcal{R}$ , sehingga dengan menggunakan *Support Vector Regression* akan ditemukan suatu fungsi  $f(x)$  yang mempunyai *deviasi* paling besar  $\varepsilon$  dari target aktual  $y_j$  untuk semua data training[17]. Maka  $\varepsilon$  sama dengan 0 akan didapat regresi yang sempurna. Dengan metode SVR diperoleh fungsi regresi dibawah ini:

$$f(x) = w^T \varphi(x) + b \quad (1)$$

Dimana

$w$  : vektor pembobot berdimensi L

$\varphi(x)$  : fungsi yang memetakan x pada ruang dengan l dimensi

$b$  : bias

Agar memperoleh generalisasi yang baik untuk fungsi regresi  $f(x)$ , dapat dilakukan dengan meminimalkan norm dari  $w$ . Oleh karena itu perlu menyelesaikan problem optimasi berikut:

$$\min \frac{1}{2} \|w\|^2 \quad (2)$$

Dengan syarat:

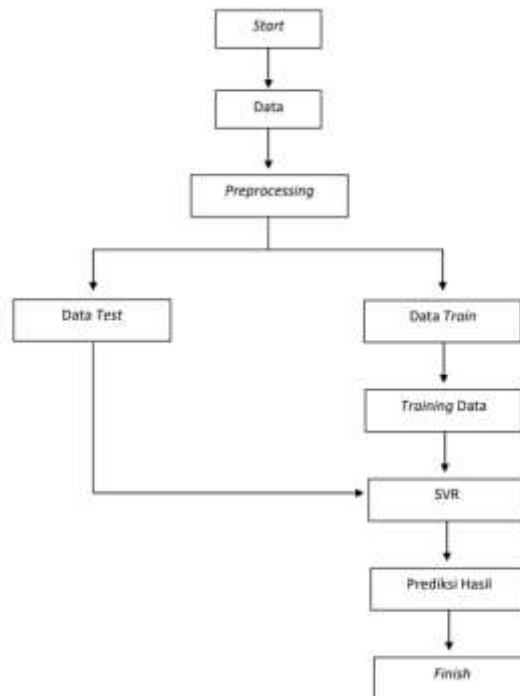
$$y_i - w\varphi(x_i) - b \leq \varepsilon \quad (3)$$

$$w\varphi(x_i) - y_i + b \leq \varepsilon, i = 1, 2, \dots, \lambda \quad (4)$$

$$\sum e^2 = \sum (y - \hat{y})^2 \quad (5)$$

$$= \sum(y - (b_0 + b_1x))^2 \tag{6}$$

Asumsikan bahwa ada suatu fungsi  $f$  yang dapat mengaproksimasi semua titik  $(x_j, y_j)$  dengan presisi  $\epsilon$ . asumsikan bahwa semua titik ada dalam rentang  $f \pm \epsilon$  (*feasible*). Dalam hal ini kemungkinan ada beberapa titik yang mungkin keluar dari rentang, sehingga dapat di tambahkan variabel slack  $\xi, \xi^*$  untuk mengatasi masalah pembatas yang tidak layak dalam problem optimasi [19].



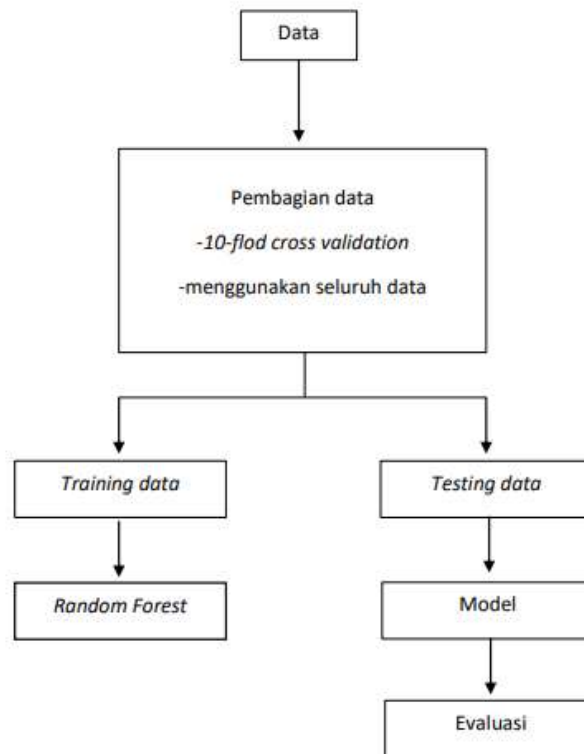
GAMBAR 2. DIAGRAM ALIR ALGORITMA SUPPORT VEKTOR REGRESSION

### Random Forest Regressor

*Random Forest Regressor* merupakan pengembangan dari *Decision Tree Regressor*, yaitu prediktor pohon. *Random Forest Regressor* yaitu konsep untuk membuat sejumlah besar pohon keputusan berkolerasi dimana semua pohon keputusan bertindak sebagai model ansambel[19]. Setiap pohon keputusan meletakkan prediksi kelas dan keputusan akhir didasarkan pada hasil maksimum [20].

*Random Forest Regressor* adalah biasa digunakan untuk menyelesaikan masalah yang berhubungan dengan klasifikasi, regresi dan sebagainya. Algoritma ini berupa kombinasi dari beberapa *tree prediction* atau *Decision Tree Regressor* di mana hasil prediksi dari *Random Forest Regressor* didapatkan melalui hasil terbanyak dari setiap individual *Decision tree Regressor*. Untuk *Random Forest Regressor* yang terdiri dari N trees dirumuskan sebagai:

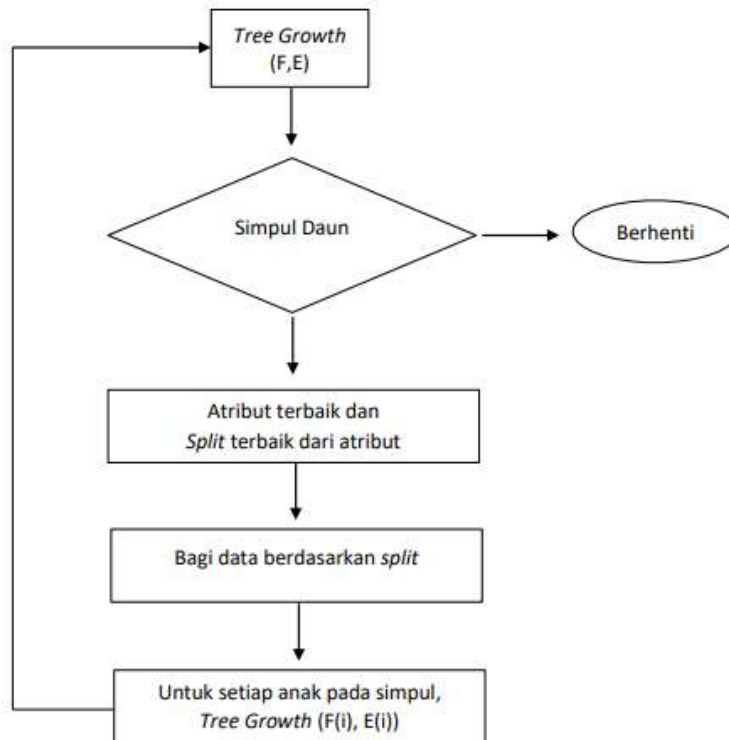
$$l(y) = \operatorname{argmax}_c (\sum_{n=1}^N I_{h_n(y)=c}) \tag{7}$$

GAMBAR 3. DIAGRAM ALIR ALGORITMA *RANDOM FOREST REGRESSION*

### Decision Tree Regressor

Decision Tree Regressor merupakan algoritma machine learning yang banyak digunakan dalam masalah klasifikasi dan regresi, bahkan masalah multi output. Decision Tree Regressor merupakan algoritma yang sangat kuat, mampu menyesuaikan kumpulan data yang kompleks menjadi simple. Decision Tree Regressor juga dapat menemukan interaksi tersembunyi antara sejumlah calon variabel input menggunakan sebutan variabel target dan juga dapat memadukan antara ekspolarasi data dan pemodelan sebagai akibatnya sangat baik menjadi langkah awal dalam proses

pemodelan. Elemen Decision Tree Regressor ada tiga yaitu simpul root, simpul ranting dan simpul leaf



[21].

GAMBAR 4. DIAGRAM ALIR ALGORITMA DECISION TREE REGRESSION

Memilih nilai *gain* tertinggi sebagai akar, didasarkan pada hasil perhitungan nilai *gain*. Untuk menghitung *gain* digunakan rumus seperti yang tertera dalam persamaan berikut:

$$Entropy(S) = \sum_{i=1}^n -\pi * \log_2 * \pi \quad (8)$$

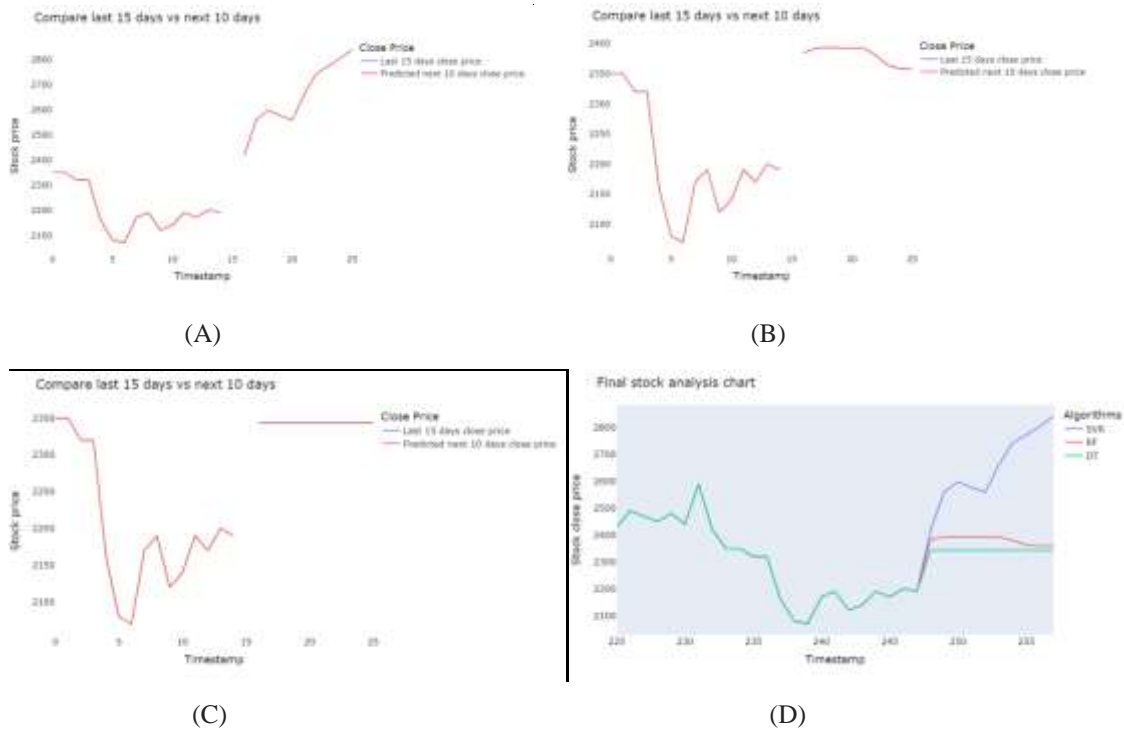
$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (9)$$

keterangan:

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah Partisi atribut A
- Pi : Proporsi dari Si terhadap S
- |Si| :Jumlah kasus pada partisi ke i
- |S| : Jumlah kasus dalam S

### III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data *time series* PT. Kimia Farma, Tbk. yang diunduh dari *finance.yahoo.com* dan dengan menggunakan metode *Decision Tree Regression*, *Random Forest Regression* dan *Support Vector Regression*. Penelitian ini bertujuan untuk membandingkan hasil prediksi dan menganalisis metode yang paling paling baik tingkat keakuratannya. Hasil yang diperoleh adalah sebagai berikut:



GAMBAR 5. HASIL KESIMPULAN DENGAN PREDIKSI (A) SVR, (B) RFR, (C) DTR, DAN PERBANDINGAN AKURASI TIGA METODE

PERBANDINGAN TIGA METODE REGRESI				
Algoritma		Support Vektor Regressor	Random Forest Regressor	Decision Tree Regressor
RMSE	Train	122.09	44.11	3.74
	Test	143.95	115.95	142.51
MSE	Train	14905.37	1945.71	14.0
	Test	20722.31	13443.61	20309.42
MAE	Train	103.43	28.43	2.11
	Test	129.15	75.01	104.86
Variance Regressor Score	Train	0.90	0.98	0.99
	Test	0.60	0.28	(-0.21)
R2 score	Train	0.89	0.98	0.99
	Test	(-0.24)	0.19	(-0.21)
MGD	Train	0	0	1.99
	Test	0	0	0
MPD	Train	5.45	0.67	0
	Test	8.37	5.63	8.38

GAMBAR 6. TABEL HASIL AKURASI DARI TIGA METODE

Metode Regresi dikatakan paling baik apabila score error mendekati nol (0). Pada tabel metode *Support Vector Regression*, *Random Forest Regression* dan *Decision Tree Regression* hasil yang diperoleh pada setiap metode memiliki jarak yang cukup jauh dan metode yang menghasilkan nilai yang mendekati 0 adalah metode *Random Forest Regressor*. Dapat disimpulkan bahwa metode yang paling baik untuk digunakan dalam prediksi harga saham di Kimia Farma adalah metode *Random Forest Regressor* dengan score *MSE-test* 13443.61 dan *R2 score-test* 0.19.

## IV. SIMPULAN DAN SARAN

Penelitian ini menggunakan tiga metode regresi yaitu SVR, RFR, DTR karena metode-metode ini cocok untuk memprediksi harga saham. Selanjutnya diterapkan metode tersebut pada dataset saham Kimia Farma Tbk untuk dianalisis. Dataset saham Kimia Farma Tbk. Berisi tentang naik turunnya harga saham perusahaan kimia farma Tbk. pada setiap harinya. Tujuan akhir dari penelitian ini adalah untuk membandingkan tingkat keakuratan dari metode SVR, RFR dan DTR. Hasil penelitian menunjukkan bahwa untuk data saham ini model prediksi dengan akurasi terbaik adalah RFR dengan *score MSE-test* 13443.61 dan *R2 score-test* 0.19.

## DAFTAR PUSTAKA

- [1]. Anoraga, P., & Pakarti, P. (2001). Pengantar Pasar Modal (Edisi Revisi). *PT. Asdi Mahasatya. Jakarta.*
- [2]. Sudarmanto, E., Khairad, F., Damanik, D., Purba, E., Peranginangin, A. M., Arfandi, S. N., ... & Astuti, A. (2021). *Pasar Uang dan Pasar Modal.* Yayasan Kita Menulis.
- [3]. Putri, H. T. (2020). Covid 19 dan Harga Saham Perbankan di Indonesia. *Eksis: Jurnal Ilmiah Ekonomi Dan Bisnis, 11(1), 6-9*
- [4]. Yasin, H., Prahutama, A., & Utami, T. W. (2014). Prediksi harga saham menggunakan support vector regression dengan algoritma grid search. *Media Statistika, 7(1), 29-35.*
- [5]. Oktavianti, I., Ermatita, E., & Rini, D. P. (2019). Analisis Pola Prediksi Data Time Series menggunakan Support Vector Regression, Multilayer Perceptron, dan Regresi Linear Sederhana. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 3(2), 282-287.*
- [6]. Awad, M., & Khanna, R. (2015). Support vector regression. In *Efficient learning machines* (pp. 67-80). Apress, Berkeley, CA.
- [7]. Geron, Aurelien.(2019).Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. United States of America.
- [8]. Alfredo, A., Jondri, J., & Rismala, R. (2015). Prediksi Harga Saham menggunakan Support Vector Regression dan Firefly Algorithm. *eProceedings of Engineering, 2(2).*
- [9]. Sandag, G. A. (2020). Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *CogITO Smart Journal, 6(2), 167-178.*
- [10]. Azhar, Y., Mahesa, G. A., & Mustaqim, M. C. (2021). Prediksi pembatalan pemesanan hotel menggunakan optimalisasi hiperparameter pada algoritme Random Forest. *Jurnal Teknologi dan Sistem Komputer, 9(1), 15-21.*
- [11]. Orpa, E. P. K., Ripanti, E. F., & Tursina, T. (2019). Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4. 5. *JUSTIN (Jurnal Sistem dan Teknologi Informasi), 7(4), 272-278.*
- [12]. Sitorus, H., & Tarihoran, Y. (2018). Analisis Harga Saham PT Astra Internasional Tbk Menggunakan Data Dari Bursa Efek Indonesia dalam Jangka Waktu Pendek Menggunakan Metode Naïve Bayes dan Decision Tree-J48. *TeIKA, 8(1), 21-33.*
- [13]. Rahmadayanti, C., Rabbani, H., & Rohmawati, A. A. (2018). Model GARCH dengan Pendekatan Conditional Maximum Likelihood untuk Prediksi Harga Saham. *Indonesia Journal on Computing (Indo-JC), 3(2), 21-28.*
- [14]. Sykes, A. O. (1993). An introduction to regression analysis.
- [15]. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support Vector Regression Machines. *Advances in Neural Information Processing Systems (Vol. 9).*
- [16]. Santosa, B. (2007). Feature Selection with Support Vector Machines Applied on Tornado Detection. *IPTEK The Journal for Technology and Science, 18(1).*
- [17]. Alfredo, A., Jondri, J., & Rismala, R. (2015). Prediksi Harga Saham menggunakan Support Vector Regression dan Firefly Algorithm. *eProceedings of Engineering, 2(2).*
- [18]. Astuti, Y. A. (2011). Analisis perbandingan teknik Support Vector Regression (SVR) dan decision tree C4. 5 dalam data mining.
- [19]. Agustyaningrum, C. I., Gata, W., Nurfalalah, R., Radiyah, U., & Maulidah, M. (2020). KOMPARASI ALGORITMA NAIVE BAYES, RANDOM FOREST DAN SVM UNTUK MEMREDIKSI NIAT PEMBELANJA ONLINE. *Jurnal Informatika, 20(2), 164-173.*



- [20]. Kabir, M. R., Ashraf, F. B., & Ajwad, R. (2019, December). Analysis of different predicting model for online shoppers' purchase intention from empirical data. In *2019 22nd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-6). IEEE.
- [21]. Azmi, Z., & Dahria, M. (2013). Decision tree berbasis algoritma untuk pengambilan keputusan. *Jurnal Ilmiah SAINTIKOM*, 12(3).