Terakreditasi SINTA Peringkat 4
Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Nomor: 200/M/KPT/2020
Masa berlaku mulai Volume 9 Nomor 1 Tahun 2019 sampai Volume 13 Nomor 2 Tahun 2023
Terbit online pada laman: https://ojs.stmikplk.ac.id/index.php/saintekom/index

SAIN TEKOM
ISSN 2088-1770 (Print)
ISSN 2503-3247 (Online)
Vol. 13 No. 2 (2023) 184-192

JURNAL SAINTEKOM
Sains, Teknologi, Komputer, dan Manajemen

Data Mining Menggunakan Multiple Regression untuk Prediksi Harga Saham Netflix

*Rama Dona Ariyatma[1], Syahrul Fahmi[2]
[1,2)]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Meruya Selatan No. 01, Kembangan, DKI Jakarta
Email: [1]41520120064@student.mercubuana.ac.id, [2]41520110025@student.mercubuana.ac.id

ABSTRACT

Investing in the stock market is an important and fascinating endeavor, especially when we observe significant increases in certain stocks. Currently, Netflix stock is one of the rising stars and sought after by investors. However, along with the potential for high profits, there are certainly risks of losses that need to be anticipated. To mitigate these risks, an investor must make predictions about future stock prices. One method that can be used is data mining, a data processing technique used to discover patterns in data. In this study, data mining was conducted using the multiple regression algorithm to predict the future price of Netflix stock. Python and Jupyter Notebook were used as tools to process the data, which was collected from January 4, 2010, to March 30, 2023, totaling 3334 data points. After data processing, the model yielded a score of 0.99%, indicating a highly reliable model. Additionally, evaluation using RMSE resulted in a value of 3.73, and MAE had a value of 2.80, both derived from 1334 testing data points. With accurate prediction results and the evaluation conducted, an investor can use these findings as a reference when deciding whether to buy or sell Netflix stock.

Keywords : multiple regression; data mining; stock; Netflix

ABSTRAK

Investasi di pasar saham merupakan salah satu hal yang penting dan menarik untuk diikuti, terutama jika kita melihat adanya kenaikan signifikan pada saham-saham tertentu. Saat ini, saham Netflix menjadi salah satu saham yang sedang naik daun dan menjadi incaran para investor. Namun, seiring dengan potensi keuntungan yang tinggi, tentunya ada risiko kerugian yang perlu diwaspadai. Untuk mengurangi risiko kerugian tersebut, seorang investor harus melakukan prediksi harga saham yang akan datang. Salah satu metode yang dapat digunakan adalah data *mining*, yaitu suatu teknik pengolahan data yang digunakan untuk menemukan pola-pola dalam data. Dalam penelitian ini, data *mining* dilakukan dengan menggunakan algoritma *multiple regression* untuk memprediksi harga saham Netflix yang akan datang. Penelitian ini menggunakan Python dan *Jupyter Notebook* sebagai *toolsnya* untuk mengolah data yang diambil dari 4 Januari 2010 hingga 30 Maret 2023 sebanyak 3334 data. Setelah dilakukan pengolahan data, didapatkan hasil skor model sebesar 0.99%, yang menunjukkan bahwa model sangatlah baik. Selain itu, evaluasi menggunakan RMSE juga dilakukan dengan hasil sebesar 3.73 dari 1334 data testing. Dengan hasil prediksi yang akurat dan evaluasi yang dilakukan, seorang investor dapat menggunakan hasil tersebut sebagai acuan dalam memutuskan apakah akan membeli atau menjual saham Netflix.

Kata kunci *: Multiple regression; data mining*; saham; *Netflix*

Diterima Redaksi: 22-5-2023 | Selesai Revisi: 30-8-2023 | Diterbitkan Online: 30-9-2022

1. PENDAHULUAN

Di masa kini, ada banyak sekali jenis investasi yang tersedia untuk dipilih, mulai dari investasi dalam saham, reksa dana, tanah, dan berbagai jenis investasi lainnya. Namun, setiap jenis investasi memiliki kelebihan dan kekurangan masing-masing. Salah satu jenis investasi yang menarik untuk dibahas adalah investasi dalam saham (Priyadi et al., 2019).

Saham adalah jenis investasi yang diperdagangkan melalui bursa saham dengan menggunakan surat berharga sebagai bukti kepemilikan (Pambudi et al., 2023). Contoh perusahaan yang memiliki pergerakan saham yang signifikan di pasar saham adalah *Netflix*, perusahaan yang bergerak di bidang media hiburan. Harga saham Netflix lebih terbuka dan dapat diakses oleh semua orang, sehingga memudahkan bagi seorang *stock trader* dalam berinvestasi di bursa efek.

Bagi seorang investor, memaksimalkan keuntungan dan meminimalkan potensi kerugian dalam berinvestasi merupakan suatu keharusan (Priyadi et al., 2019).Oleh karena itu, dibutuhkan sebuah strategi *trading* saham Netflix yang tepat, yang mengimplementasikan konsep data *mining*. Data *mining* merupakan metode atau Teknik untuk menemukan informasi yang berasal dari sebuah data. Data *mining* untuk menggali dan menganalisis data dalam jumlah banyak (Hammad et al., 2022).

Strategi *trading* saham Netflix dengan data *mining* menggunakan teknik *multiple regression* untuk melakukan prediksi harga saham di masa depan. Teknik ini dilakukan dengan mengumpulkan data historis harga saham Netflix. Selanjutnya, data-data tersebut diolah dengan menggunakan teknik data *mining* termasuk teknik *multiple regression*, untuk menghasilkan model prediksi harga saham Netflix di masa depan.

Pemilihan strategi *trading* saham Netflix yang tepat, investor dapat meminimalkan risiko dan memaksimalkan potensi keuntungan dalam berinvestasi. Selain itu, teknik *data mining* juga dapat digunakan untuk meningkatkan efektivitas dan efisiensi investasi di berbagai jenis produk investasi lainnya.

Penelitian mengenai prediksi tren pergerakan harga saham sudah banyak dilakukan oleh penelitian-penelitian

DOI: https://doi.org/10.33020/saintekom.v13i2.419

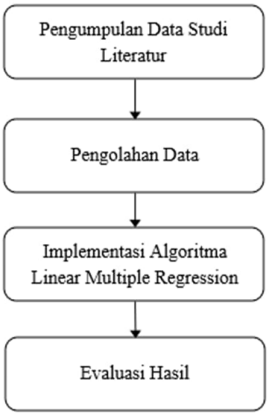
sebelumnya seperti pada penelitian yang dilakukan oleh Eka patriya yang menggunakan algoritma *Support Vector Machine* (SVM) namun yang digunakan adalah saham dari Indeks Harga Saham Gabungan (IHSG) yang menggunakan RMSE untuk hasil evaluasinya didapatkan sebesar 14,334. (Patriya, 2020). Penelitian lain yang dilakukan oleh Arie bayu untoro yang menggunakan Jaringan syaraf tiruan untuk prediksi harga saham yang memiliki tingkat *errors* sebesar 3,38% (Untoro, 2020)

Perkembangan teknologi informasi memungkinkan pengolahan data menjadi lebih cepat atau lebih mudah. Oleh karena itu digunakan sebuah bahasa *pemrograman Python* untuk memprediksi dari sekumpulan data serta *tools jupyter notebook.*

Berdasarkan penelitian-penelitian yang pernah dilakukan untuk prediksi harga saham,maka akan dilakukan penelitian yang menggunakan Teknik *multiple regression* untuk mengetahui apakah nilai *erros* dari RMSE kecil atau besar serta ada tambahan MAE untuk meningkatkan informasi evaluasi yang tidak ada pada penelitian sebelumnya. Dengan menggunakan python sebagai bahasa pemrograman dan *jupyter notebook* sebagai toolsnya.

2. METODE

Kerangka kerja penelitian ini merupakan langkah-langkah yang akan dilakukan dalam melakukan penyelesaian masalah yang akan dibahas. Adapun detail kerangka kerja penelitian tahapan demi tahapannya dapat dilihat seperti Gambar 1 berikut ini.



Gambar 1. Tahapan penelitian

2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini dihimpun melalui berbagai sumber yang ada. Data tersebut mencakup data-data seperti data historis harga emas USD dan harga saham dari berbagai perusahaan. Data-data tersebut diperoleh melalui sebuah website *investing.com* yang merupakan website penyedia informasi seputar

DOI: https://doi.org/10.33020/saintekom.v13i2.419

indeks/saham, komoditas, valuta asing, dan harga obligasi.

2.2. *Linear Regression*

Algoritma *linear regression* adalah jenis aturan *classification and regression* pada data *mining* selain *Linear Regression* yang termasuk pada golongan ini adalah *Support Vector Machine, Logistic Regression* dan lain-lain. Analisis *linear regression* adalah teknik *data mining* untuk menentukan bahwa terdapat hubungan antara variabel yang ingin diramalkan dengan variabel lain.

Algoritma *linear regression* menggunakan model Persamaan 1-3 berikut.

$$a = \frac{(\sum y)(\sum x^2)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2} \quad (1)$$

$$b = \frac{n(\sum xy)-(\sum x)(\sum y)}{n(\sum x^2)-(\sum x)^2} \quad (2)$$

$$y = a + b.x \quad (3)$$

x merupakan *variable dependent* atau variabel respon, a dan b merupakan konstanta (Indarwati et al., 2018).

2.3. *Multiple Linear Regression*

Multiple Linear Regression (MLR) merupakan teknik statistik yang menggunakan beberapa variabel penjelas untuk memprediksi hasil dari variabel respon. Tujuan dari *Multiple Linear Regression* (MLR) adalah untuk memodelkan hubungan linier antara variabel penjelas (*independen*) dan variabel respon (*dependen*).

Dalam metode ini untuk mengekspresikan antar variable yang digunakan sebagai kombinasi *linear* (Triyanto et al, 2019), nantinya akan digunakan dalam pembuatan model.

2.4. RMSE (*Root Mean Square Error*)

RMSE digunakan sebagai evaluasi hasil dari pemodelan yang akan menentukan nilai *errors* kecil atau besar pada sebuah pemodelan. RMSE adalah metrik yang digunakan untuk mengukur seberapa besar kesalahan prediksi dalam suatu model. RMSE menghitung selisih antara nilai prediksi yang dihasilkan oleh model dengan nilai aktual yang ada dalam data, kemudian menghitung nilai rata-rata kuadrat dari selisih tersebut, dan akhirnya mengambil akar kuadrat dari nilai tersebut. (Rahmatullah et al, 2023)

2.5. MAE (*Mean Absolute Error*)

Digunakan *Mean Absolute Error* (MAE) sebagai salah satu untuk evaluasi hasil. MAE adalah metode pengukuran kesalahan (*error*) pada model peramalan

DOI: https://doi.org/10.33020/saintekom.v13i2.419

Tabel 1. Data Saham Netflix

	Date	Price	Open	High	Low	Vol.	Change %
0	03/30/2023	338.43	340.27	343.29	335.30	7.12M	1.93%
1	03/29/2023	332.03	326.29	332.85	325.73	6.29M	2.63%
2	03/28/2023	323.52	326.06	333.32	321.28	6.48M	-1.26%
3	03/27/2023	327.66	327.55	336.44	324.41	8.62M	-0.22%
4	03/24/2023	328.39	320.63	331.83	320.63	13.00M	2.50%
…	…	…	…	…	…	…	
3333	01/04/2010	7.64	7.93	7.96	7.57	17.24M	-2.92%

Tabel 2. Hasil Seleksi Data Saham Netflix

	Date	Price	Open	High	Low
0	01/04/2010	7,64	7.93	7.96	7.57
1	01/05/2010	7,36	7.65	7.66	7.26
2	01/06/2010	7,62	7.36	7.67	7.20
3	01/07/2010	7,49	7.73	7.76	7.46
4	01/08/2010	7,61	7.50	7.74	7.47
…	…	…	…	…	…
3333	03/30/2023	338.43	340.27	343.29	335.30

atau prediksi. Metode ini menghitung rata-rata kesalahan absolut antara nilai prediksi dengan nilai aktual (*riil*) pada setiap observasi atau sampel data yang ada (Irawan et al,2021).

MAE mengukur besarnya kesalahan prediksi dalam satuan yang sama dengan data yang diamati. Oleh karena itu, MAE umumnya digunakan ketika unit data yang diamati memiliki arti penting yang sama, seperti dalam peramalan harga saham, suhu udara, atau penjualan produk. Evaluasi model menggunakan *Mean Absolute Error* (MAE) yang mempresentasikan nilai *error* dari model asosiasi. Jika nilai MAE semakin mendekati nol, maka model tersebut semakin baik.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan Jupyter notebook sebagai *tools* dan Python sebagai bahasa pemogramannya.

3.1. Pengumpulan data

Untuk mengolah data digunakan Pandas dan Numpy .Data yang di ambil atau yang digunakan dimulai dari tanggal 4 Januari 2010 – 30 Maret 2023 yang berjumlah sebesar 3334 *rows* berikut bagian dari data seperti Tabel 1 . Sebelum menggunakan data pada Tabel 1 Peneliti akan mengubah atau menghilangkan data yang tidak perlu serta mengubah index 0 menggunakan data yang paling lama atau paling awal yaitu tanggal 4 Januari 2010 berikut Tabel 2.

DOI: https://doi.org/10.33020/saintekom.v13i2.419

3.2. Data cleaning

Peneliti melakukan *cleaning data* untuk memastikan data yang digunakan tidak memiliki nilai kosong format dalam rekomendasi. *Data cleaning* tersebut meliputi *remove duplicate,drop wrong data, missing value.* Hasil dari proses *cleaning* data yang dilakukan tidak menghasilkan perubahan apapun. Hal ini disebabkan karena memang tidak ada data yang salah atau duplikasi oleh karena itu data sudah siap digunakan.

3.3. Modeling

Model pembelajaran mesin digunakan untuk menganalisis pola data dan mendapatkan pemahaman. Hal ini dapat menghasilkan wawasan seperti visualisasi pola atau prediksi nilai di masa depan.

Tahapan ini merupakan bagian yang paling menarik dalam proyek ilmu data karena melibatkan penggunaan pembelajaran mesin sebagai komponen penting.

Dalam analisis ini, sebuah model *multiple regression* diimplementasikan dengan memanfaatkan tiga *variable independen* yaitu '*Open','High'*, dan '*Low*'. Dalam kerangka kerja ini '*Price*' diidentifikasi sebagai *variable dependen.* Yang prediksinya didasarkan pada ketiga *variable independen* tersebut. Selanjutnya pemisahan data dilakukan dengan menggunakan fungsi *train_test_split,* dimana 60% dari data diperuntukkan sebagai data *training* dan sisanya, yang mencakup 40% diperuntukkan sebagai data *testing.*

Pemodelan dilakukan dengan menggunakan pustaka *LinearRegression* dari *sklearn.linear_model*. Hasil dari pemodelan ini, telah di dapat nilai *intercept* -0,0274. Untuk *coefficients* tiap variabel, telah dijabarkan dalam tabel 3.

Tabel 3. *coefficients model*

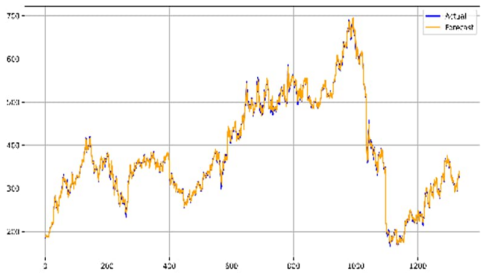
variable	*Coefficients*
Open	- 0.5329
High	0.79843
Low	0.73471

Jadi telah di dapatkan rumus pada model *multiple regression* pada persamaan 4.

$$price = -0.027 - 0.53(open) + 0.79(high) + 0.73(low) \quad (4)$$

Berikut visualisasi data dimana data *actual* dan data *prediction* diperlihatkan digunakan library *matplotlib* untuk prosesnya. Data *actual* berwarna biru, sedangkan prediksi berwarna orange

DOI: https://doi.org/10.33020/saintekom.v13i2.419



Gambar 2. Visualisasi *data actual* dengan data prediksi

Terlihat pada Gambar 2 hampir mirip dengan data aslinya. Pada contoh Tabel 4 kita bisa lihat *score hasil train data* dan *test* data sebagai berikut :

Tabel 4. *Score Acurracy data*

Test data	Train data
0.99900	0.99983

Selanjutnya pemodelan akan dilakukan evaluasi di tahapan evaluasi

3.4. Evaluasi

Pada tahap evaluasi, dilakukan pengecekan untuk memastikan bahwa pengetahuan baru yang ditemukan memiliki nilai dan kebaruan yang cukup. Model-model yang lolos evaluasi akan dipertahankan, sementara model-model yang tidak lolos akan ditinjau kembali untuk mencari solusi alternatif yang dapat meningkatkan hasil yang terbaik.

Evaluasi tersebut menggunakan RMSE dan MAE pada hasil prediksi test data. *Root mean squeared Error* 3,73 error evaluasi RMSE sebesar 3,73 dan MAE sebesar 2,80 dari 1334 test data dan 2000 train data.

Adapun saran bagi penelitian selanjutnya Data mining mempunyai banyak sekali algoritma lainnya selain *multiple regression.* Misalnya *decision tree, KNN, K-means, Random Forest* dan lain sebagainya yang bisa digunakan sebagai penelitian selanjutnya sebagai perbandingan apakah prediksi lebih efektif atau tidaknya.

sedangkan *Mean absolute error* sebesar 2,80. Terlihat cukup lumayan kecil terhadap *error* hasil prediksi tersebut. Berikut tabel hasil sebagian data asli dan data prediksi dapat dilihat pada Tabel 5.

Tabel 5. Nilai perbandingan data asli dengan data prediksi

Data asli	Prediksi	Perbedaan
186.22	187.69	1,47
185.73	186.62	0,89
187.86	187.67	0,19
189.56	191.14	1,58
190.12	189.90	0,22
190.42	189.88	0,54
187.02	187.03	0,01
188.82	186.92	1,9
188.62	189.29	0,67
189.94	189.31	0,63

Pada Tabel 5 bahwa rata-rata nilai prediksi dengan data asli sangatlah tipis bahkan ada perbedaanya hanya 0.01 saja akan tetapi masih ada prediksi perbedaanya 1,9 atau bisa dibilang sangatlah tinggi karena kita bisa melihat

DOI: https://doi.org/10.33020/saintekom.v13i2.419

dari RMSE sebelumnya yang terbilang cukup tinggi.

4. KESIMPULAN

Berdasarkan hasil dari penelitian dan pengujian yang dilakukan dapat ditarik kesimpulan bahwa penggunakan model *machine learning multiple regression* terbilang efektif untuk memprediksikan harga saham Netflix dalam penggunakan 3 *variable* yaitu *open, high, low* yang mempunyai *score model* 0.9 . dan juga mempunyai nilai error evaluasi RMSE sebesar 3,73 dan MAE sebesar 2,80 dari 1334 test data dan 2000 train data.

Adapun saran bagi penelitian selanjutnya Data mining mempunyai banyak sekali algoritma lainnya selain *multiple regression.* Misalnya *decision tree, KNN, K-means, Random Forest* dan lain sebagainya yang bisa digunakan sebagai penelitian selanjutnya sebagai perbandingan apakah prediksi lebih efektif atau tidaknya.

DAFTAR PUSTAKA

Hammad, R., Hardita, V. C., Zulfikri, M., & Sholeha, E. W. (2022). Penerapan metode apriori sebagai sistem pendukung keputusan pembentukan paket bibit buah . SAINTEKOM, 58-68.https://doi.org/10.33020/saintekom.v12i1.240

Indarwati, T., Irwanti, T., & Rimawati, E. (2018).Penggunaan Metode Linear Regression untuk Prediksi Penjualan Smartphone. Jurnal TIKomSiN, 1-6. https://doi.org/10.30646/tikomsin.v6i2.369

Irawan, F., Sumijan, & Yuhandri. (2021).Prediksi Tingkat Produksi Buah Kelapa Sawit dengan Metode Single Moving Average.Jurnal Informasi dan Teknologi, 251-256. https://doi.org/10.37034/jidt.v3i4.162

Pambudi, A., Abidin, Z., & Permata. (2023). Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt.Telkom Indonesia. JDMSI, 1-14. https://doi.org/10.33365/jdmsi.v4i1

Patriya, E. (2020).Implementasi Support Machine pada Prediksi Harga Saham Gabungan (IHSG). Jurnal Ilmiah Teknologi dan Rekayasa,24-38. http://dx.doi.org/10.35760/tr.2020.v25i1.2571

Priyadi, I., Santony, J., & Na'am, J. (2019). Data Mining Predictive Modeling for Prediction of Gold Prices . Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM), 93-100. http://dx.doi.org/10.24014/ijaidm.v2i2.6864

Rahmatullah,S., Juningsih, E. H., & Rachmawati, S. (2023). Prediksi nilai akademik peserta didik di masa pandemi covid 19 dengan regresi linier berganda. JISAMAR,112-123. https://doi:10.52362/jisamar.v7i1.1012

DOI: https://doi.org/10.33020/saintekom.v13i2.419

Suryanto, A. A., Muqtadir A. (2019).Penerapan Metode Mean Absolute Error(MAE) Dalam Algoritma Regresi linear Untuk prediksi Produksi Padi.Jurnal Sains dan Teknologi,78-83. https://doi.org/10.32764/saintekbu.v11i1.298

Triyanto, E., Sismoro, H., & Laksito, A. D. (2018).Implementasi Algoritma Regresi Linear Berganda untuk Memprediksi Produksi Padi di Kabupaten Bantul.Jurnal Teknologi dan sistem Informasi Univrab, 73-86. https://doi.org/10.36341/rabit.v4i2.666

Untoro, A. B. (2020).Prediksi harga saham Dengan menggunakan Jaringan syaraf Tiruan. Jurnal Teknologi Informatika dan Komputer,24-38.https://doi.org/10.37012/jtik.v6i2.212

Yanto, R. (2018). Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Regresi Linear. JURNAL RESTI, 2, 3.https://doi.org/10.29207/resti.v2i1.282

DOI: https://doi.org/10.33020/saintekom.v13i2.419