

Analisis Komparasi *Linear Regression* dan *Polynomial Regression* untuk Prediksi Harga Saham

Ni Wayan Sumartini Saraswati^{1*}, I Wayan Dharma Suryawan², I Made Andi Kertha Yasa³

¹*Teknologi dan Informatika, Institut Bisnis dan Teknologi Indonesia, Jl. Tukad Pakerisan No.97, Panjer, Denpasar Selatan, Denpasar, Indonesia*
E-mail: sumartini.saraswati@instiki.ac.id

²*Teknologi dan Informatika, Institut Bisnis dan Teknologi Indonesia, Jl. Tukad Pakerisan No.97, Panjer, Denpasar Selatan, Denpasar, Indonesia*
E-mail: dharma.suryawan@instiki.ac.id

³*Teknologi dan Informatika, Institut Bisnis dan Teknologi Indonesia, Jl. Tukad Pakerisan No.97, Panjer, Denpasar Selatan, Denpasar, Indonesia*
E-mail: andikertha7@gmail.com

*Penulis Korespondensi

Abstrak— Investasi memegang peranan penting dalam melawan inflasi dan mendorong pertumbuhan ekonomi. Di antara berbagai instrumen investasi, saham menawarkan potensi keuntungan yang tinggi, tetapi memerlukan analisis yang cermat untuk meminimalkan risiko dan memaksimalkan keuntungan. Penelitian ini berfokus pada prediksi harga penutupan yang disesuaikan (*Adj Close*) dari saham PT Mitra Energi Persada Tbk (KOPLJK), sebuah perusahaan energi yang terdaftar di Bursa Efek Indonesia (BEI), menggunakan teknik regresi *machine learning*, karena data historis yang mengalami fluktuasi signifikan. Dengan membandingkan *Linear Regression* dan *Polynomial Regression* yang dilengkapi dengan pengoptimalan regularisasi *Ridge* dan *Least Absolute Shrinkage and Selection Operator* (LASSO). Penelitian ini bertujuan untuk mengidentifikasi model yang paling efektif dalam memprediksi harga saham. Hasil analisis menunjukkan bahwa fitur *Low* dan *High* memiliki korelasi yang paling kuat dengan harga *Adj Close*, sementara *Volume* memiliki korelasi terendah. *Polynomial Regression* dengan *degree=3* dan pengoptimalan regularisasi *Ridge* memberikan performa terbaik. Hasil evaluasi mencapai *Mean Square Error* (MSE) 122.9618, *Root Mean Squared Error* (RMSE) 11.0888, dan *R-squared* (R^2) 0.9883. Pada pengoptimalan model menggunakan LASSO cenderung mengurangi relevansi fitur sehingga memberikan performa yang lebih buruk.

Kata Kunci — linear regression; polynomial regression; harga saham; LASSO regularization; ridge regularization.

Abstract— Investments play an important role in combating inflation and driving economic growth. Among the various investment instruments, stocks offer high profit potential but require careful analysis to minimize risk and maximize profit. This study focuses on predicting the adjusted closing price (*Adj Close*) of PT Mitra Energi Persada Tbk (KOPLJK), an energy company listed on the Bursa Efek Indonesia (BEI), using machine learning regression techniques, based on historical data that experience significant fluctuations. Linear Regression and Polynomial Regression equipped with Ridge regularization optimization and Least Absolute Shrinkage and Selection Operator (LASSO) were compared. This study aimed to identify the most effective model for predicting stock prices. The results of the analysis show that the Low and High features have the strongest correlation with the *Adj Close* price, whereas Volume has the lowest correlation. Polynomial Regression with degree = 3 and Ridge regularization optimization provided the best performance. The evaluation results reached an Mean Square Error (MSE) of 122.9618, an Root Mean Squared Error (RMSE) of 11.0888, and an R-squared (R^2) of 0.9883. Optimizing the model using LASSO tends to reduce the relevance of the features, thus providing worse performance.

Keywords — linear regression; polynomial regression; stock price; LASSO regularization; ridge regularization.

I. PENDAHULUAN

Investasi merupakan salah satu strategi yang digunakan untuk melawan tingkat inflasi yang dapat menggerus daya beli masyarakat. Investasi juga berperan sebagai pendorong utama pertumbuhan ekonomi suatu negara [1]. Dari berbagai instrumen investasi yang tersedia, saham menjadi salah satu pilihan dengan potensi keuntungan yang tinggi [2]. Saham di Indonesia diperdagangkan melalui BEI, di mana para investor perlu mempertimbangkan berbagai faktor sebelum memulai investasi agar dapat meminimalkan risiko dan memaksimalkan keuntungan. Dalam berinvestasi saham, beberapa aspek penting yang harus diperhatikan meliputi tujuan investasi, jangka waktu investasi, risiko, inflasi, dana,

wahana, likuidasi, dan keuntungan. Salah satu metode yang umum digunakan oleh investor dalam menganalisis saham adalah analisis teknikal. Analisis ini membantu investor dalam menentukan saham yang potensial serta waktu terbaik untuk menjual (*sell*), membeli (*buy*), atau mempertahankan (*hold*) saham tersebut [3], [4].

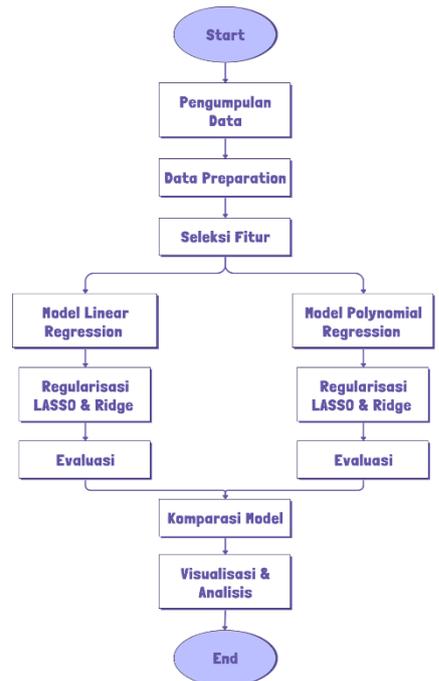
Salah satu perusahaan publik di BEI adalah PT. Mitra Energi Persada, Tbk, yang resmi tercatat sejak 23 April 2001 dengan kode perdagangan saham KOPIJK. Perusahaan ini bergerak di sektor infrastruktur energi serta minyak dan gas bumi di Indonesia. Berdasarkan data historis, harga saham PT. Mitra Energi Persada mengalami fluktuasi yang signifikan. Fluktuasi harga ini mencerminkan adanya peluang keuntungan, tetapi juga menunjukkan adanya risiko yang harus diwaspadai oleh investor. Adanya tantangan ini dapat menjadi risiko jika tidak diprediksi dengan tepat. Tantangan ini dapat diatasi jika terdapat rekomendasi yang tepat dalam pengambilan keputusan investasi. Salah satu pendekatan yang dapat digunakan adalah *machine learning*. *Machine learning* telah terbukti efektif dalam berbagai metode prediksi, terutama untuk kasus *time series*. Salah satu metode yang dapat diterapkan adalah regresi. Metode regresi yang umum digunakan antara lain, *Linear Regression* dan *Polynomial Regression*. *Linear Regression* memodelkan hubungan linear antara variabel independen dan dependen, sementara *Polynomial Regression* dapat menangani hubungan yang lebih kompleks dan non-linear [5], [6].

Dengan membandingkan kedua metode ini, model prediksi terbaik dapat dipilih untuk meningkatkan akurasi dalam memprediksi pergerakan saham. Penelitian oleh [7] menunjukkan bahwa *Polynomial Regression* lebih unggul dalam memprediksi COVID-19 dibandingkan dengan *Support Vector Machine* (SVM). Penelitian lain oleh [8] juga menunjukkan bahwa metode *regression* dapat mencapai akurasi tinggi dalam prediksi data finansial, dengan hasil terbaik mencapai 88.79%. Dalam *machine learning*, potensi *overfitting* dan multikolinearitas dapat menjadi tantangan dalam prediksi. Oleh karena itu, teknik regularisasi seperti *Ridge* dan *LASSO* dapat diterapkan untuk mengatasi masalah tersebut. Dengan menerapkan pendekatan yang tepat, investor dapat lebih memahami pola pergerakan saham dan mengambil keputusan investasi yang lebih optimal.

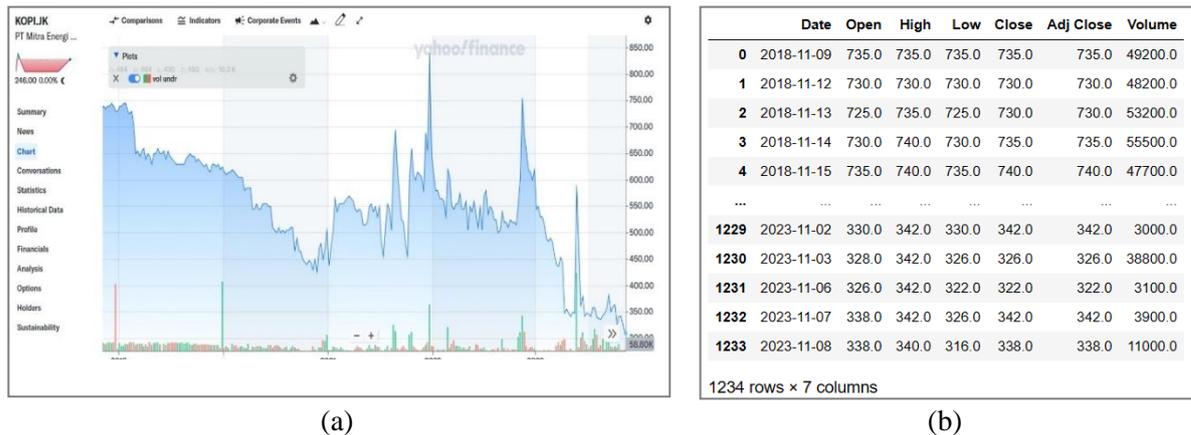
Prediksi dengan *regression* telah dilakukan oleh beberapa penelitian sebelumnya. Penelitian oleh [9] menggunakan *Linear Regression* dan *Polynomial Regression* untuk memprediksi produksi daging sapi nasional untuk mengetahui hasil produksi di tahun yang akan datang. Penelitian ini menunjukkan *Polynomial Regression* mendapatkan hasil yang lebih baik dibandingkan dengan model *Linear Regression*. Penelitian lain yang juga menggunakan *Linear Regression* dilakukan oleh [10]. *Linear Regression* mampu memprediksi dengan cukup baik dengan hasil evaluasi R^2 sebesar 0.99. Perbandingan antara *Linear Regression*, *Polynomial Regression*, dan SVM untuk prediksi harga rumah dilakukan oleh [11] menunjukkan bahwa *Polynomial Regression* mendapatkan hasil evaluasi paling baik dibandingkan dengan metode lainnya. Untuk studi kasus prediksi saham, penelitian oleh [12] menggunakan *Linear Regression* dan *Polynomial Regression* mendapatkan hasil bahwa kedua model tersebut mampu melakukan prediksi dengan cukup baik, dengan *Polynomial Regression* memberikan hasil yang lebih baik. Studi [13] menggunakan *Support Vector Regression* (SVR) untuk memprediksi harga minyak kelapa sawit di Indonesia dan nilai mata tukar uang EUR/USD. Penelitian ini menunjukkan bahwa SVR dengan kernel *Radial Basis Function* (RBF) mendapatkan hasil akurasi mendekati 100% ketika memprediksi nilai tukar EUR/USD, sedangkan untuk prediksi minyak kelapa sawit mendapatkan akurasi pada rentang 98%. Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, belum ada komparasi antara *Linear Regression* dan *Polynomial Regression* dengan pengoptimalan teknik regularisasi *Ridge* dan *LASSO* dalam mengurangi *overfitting* dan multikolinearitas. Penelitian ini menganalisis fluktuasi harga saham PT. Mitra Energi Persada Tbk serta faktor-faktor yang mempengaruhinya. Penelitian ini juga akan menerapkan metode analisis teknikal untuk memprediksi pergerakan harga saham perusahaan, guna membantu investor dalam pengambilan keputusan investasi yang lebih tepat. Selain itu, penelitian menilai efektivitas indikator teknikal dalam memprediksi harga saham serta memberikan rekomendasi strategi investasi yang optimal berdasarkan hasil analisis.

II. METODE PENELITIAN

Beberapa tahapan yang dilakukan dalam penelitian ini untuk memprediksi harga saham ditunjukkan pada Gambar 1 yang diuraikan pada sub bab berikut.



Gambar 1. Flowchart Prediksi Saham



Gambar 2. (a) Grafik Data Aktual Harga Saham, (b) Sampel Data Harga Saham

Tabel 1. Detail Atribut Dataset

No	Data	Keterangan	Atribut
1	Date	Date adalah label <i>time-series</i> yang akan menandakan waktu setiap baris <i>record</i> pada <i>dataset</i>	Date
2	KOPIJK	KOPIJK merupakan nama kode saham (<i>ticker</i>) yang digunakan oleh PT Mitra Energi Persada Tbk	Open, High, Low, Close, Volume, Adj Close

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini menggunakan data saham PT. Mitra Energi Persada Tbk. Data harga saham yang diambil adalah dari rentang tanggal 9 November 2018 - 8 November 2023. Rentang tanggal tersebut diambil karena memiliki pergerakan harga saham yang signifikan. Dataset diambil dari *Yahoo Finance* (<https://finance.yahoo.com/>) dalam data *time series* harian harga saham dari perusahaan tersebut dengan satuan rupiah. Data yang diambil berjumlah 1233 *record* dan 7 atribut

dengan atributnya diuraikan pada Tabel 1. Gambar 2a merupakan grafik data aktual harga saham dari perusahaan. Gambar 2b merupakan sampel data dari harga saham.

Data Preparation

Data yang telah dikumpulkan masuk ke tahapan *cleaning data*. Proses *cleaning* ini dilakukan untuk mempersiapkan data sebelum masuk ke seleksi fitur, memastikan agar data tidak mengandung *missing value* dan duplikasi data. *Missing value* dan duplikasi data yang tidak diatasi akan menyebabkan kesalahan analisis. Selain itu, tahapan persiapan ini bertujuan untuk lebih untuk meningkatkan kualitas data yang akan digunakan [14]. Persiapan data mencakup semua kegiatan untuk membangun dataset yang akan diterapkan ke dalam alat pemodelan [15]. Hasil pengecekan menunjukkan bahwa data penelitian ini tidak mengandung *missing value* maupun duplikasi, sehingga data dinyatakan bersih dan siap digunakan untuk tahap analisis selanjutnya.

Seleksi Fitur

Sebelum masuk ke tahapan seleksi fitur, terdapat dua variabel yang ditentukan terlebih dahulu. Tersaji dua variabel utama yang dipergunakan dalam penelitian ini, yakni variabel bebas dan variabel terikat. Variabel bebas adalah variabel penyebab atau variabel yang berpengaruh terhadap variabel terikat [16]. Variabel yang digunakan dalam penelitian ini adalah harga *Open*, *High*, *Low*, *Close*, dan *Volume*. Variabel bebas ini berpotensi menjadi *feature* atau prediktor dari variabel terikat. Variabel terikat adalah variabel akibat atau variabel yang dipengaruhi variabel bebas [16]. Variabel terikat yang digunakan adalah harga penutupan atau *Adj Close* harga saham hari berikutnya. Variabel terikat akan menjadi target prediksi harga saham dalam penelitian.

Seleksi fitur merupakan teknik umum dalam *data mining* yang digunakan pada tahap *preprocessing*. Salah satu metode yang sering diterapkan untuk menentukan bobot fitur adalah *Mutual Information* (MI). MI mengukur sejauh mana keberadaan atau ketidakhadiran suatu *term* berkontribusi dalam meningkatkan akurasi keputusan klasifikasi [17]. Dengan kata lain, MI mengukur ketergantungan antar variabel multidimensi [18]. MI adalah salah satu cara untuk melakukan seleksi fitur yang digunakan pada penelitian ini. Beraha et al., dalam [19] menyatakan bahwa MI adalah sebuah indeks yang mengukur ketergantungan statistik antara dua variabel acak. Berbeda dengan indeks lain seperti koefisien korelasi Pearson, MI tidak hanya mampu mendeteksi hubungan linear, tetapi juga dapat menangkap ketergantungan non-linear serta tetap invarian terhadap transformasi pada variabel. Zhao et al., dalam [19] menyatakan bahwa nilai MI antara dua variabel acak selalu bernilai non-negatif. MI bernilai nol jika dan hanya jika kedua variabel acak tersebut benar-benar independen. Semakin besar nilai MI, semakin tinggi tingkat ketergantungan antara kedua variabel tersebut.

Linear Regression dan Polynomial Regression

Tahap selanjutnya adalah membangun model prediksi dengan *Linear Regression* dan *Polynomial Regression*. *Linear Regression* bekerja dengan teknik peramalan kuantitatif statistik yang pada umumnya menggunakan data historis yang menitikberatkan pada pola, perubahan pola, dan faktor gangguan yang disebabkan oleh pengaruh acak. Selain itu, metode ini memberikan hasil *output* prediksi dengan melakukan pengembangan hubungan matematis antar variabel. Terdapat dua jenis *Linear Regression* yaitu *Linear Regression* satu variabel dan *Linear Regression* multi variabel. *Linear Regression* satu variabel hanya dicari korelasi antara variabel x dengan variabel bebas y. *Linear Regression* multi variabel mencari hubungan antara banyak variabel [20]. Model *Linear Regression* penelitian ini ditunjukkan pada persamaan (1).

$$AdjClose_{t+1} = \beta_0 + \beta_1 \cdot Open_t + \beta_2 \cdot High_t + \beta_3 \cdot Low_t + \beta_4 \cdot Volume_t + \beta_5 \cdot Close_t + \varepsilon \quad (1)$$

Di mana t adalah adalah waktu dengan t+1 adalah prediksi hari berikutnya. β_0 adalah intersep, $\beta_1 - \beta_5$ adalah *slope* atau koefisien yang dikurang koefisien regresi. *Open*, *High*, *Low*, *Close*, dan *Volume* adalah X atau variabel bebas, *Adj Close* adalah Y atau variabel terikat yang ingin diprediksi, serta ε adalah *error*. *Polynomial Regression* adalah pengembangan dari *Linear Regression* untuk mengatasi data-data non-linear. Secara formula matematis, yang membedakan antara *Linear Regression* dengan *Polynomial Regression* adalah untuk menghitung nilai koefisien regresi β hanyalah di matriks

desain X. *Polynomial Regression* adalah jenis regresi khusus yang bekerja pada hubungan lengkung (*curvilinear*) antara nilai dependen dan nilai independen. *Polynomial Regression* merupakan model *Linear Regression* yang dibentuk dengan menjumlahkan pengaruh masing-masing variabel prediktor (X) yang dipangkatkan meningkat sampai orde ke-n [9]. *Polynomial Regression* order 2 ditunjukkan pada persamaan (2).

$$AdjClose_{t+1} = \beta_0 + \beta_1 \cdot Open_t + \beta_2 \cdot High_t + \beta_3 \cdot Low_t + \beta_4 \cdot Volume_t + \beta_5 \cdot Close_t + \beta_6 \cdot Open_t^2 + \beta_7 \cdot High_t^2 + \beta_8 \cdot Low_t^2 + \beta_9 \cdot Volume_t^2 + \beta_{10} \cdot Close_t^2 + \varepsilon \quad (2)$$

Regularisasi

Setelah dibangun model *Linear Regression* dan *Polynomial Regression*, model diregularisasi dengan LASSO dan *Ridge*. LASSO merupakan metode regresi berganda yang digunakan untuk melakukan *shrinkage*, yaitu memperkecil nilai koefisien estimasi mendekati nol, serta *selection operator*, yaitu memilih variabel independen yang paling berpengaruh. Dengan adanya kendala tersebut, LASSO mampu mengecilkan koefisien hingga mendekati nol atau bahkan mengaturnya menjadi nol, sehingga secara efektif dapat melakukan seleksi variabel [21], [22]. *Ridge* merupakan metode yang digunakan untuk meningkatkan kestabilan penduga kuadrat terkecil ketika terdapat masalah multikolinearitas. Metode ini bekerja dengan menambahkan koefisien bias positif dalam proses estimasi parameter, sehingga menghasilkan penduga dengan ragam yang lebih kecil dibandingkan dengan metode kuadrat terkecil biasa [23].

Evaluasi Model

Evaluasi model pada penelitian ini menggunakan tiga jenis model evaluasi. Evaluasi model terdiri dari MSE, RMSE, dan R^2 . Setiap model evaluasi mendapatkan nilai hasil masing-masing. MSE adalah metode dengan teknik lain untuk menguji tingkat kesalahan dari metode peramalan. Setiap kesalahan dikuadratkan. Teknik pendekatan ini menetapkan kesalahan peramalan besar karena kesalahan kuadrat. Dengan kata lain, MSE adalah rata-rata perbedaan kuadrat antara nilai yang diprediksi dan nilai yang sebenarnya [24]. Rumus dari MSE ditunjukkan pada persamaan (3).

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \quad (3)$$

Dengan n adalah jumlah data, y_i adalah nilai actual dari data, serta \hat{y}_i adalah nilai yang diprediksi oleh model. RMSE adalah cara umum dipergunakan untuk mengukur kesalahan model dari prediksi data yang bersifat kuantitatif. Semakin kecil nilai RMSE, semakin baik model memprediksi harga saham [25]. RMSE dipergunakan untuk mengetahui ukuran sebaran simpangan titik data dari garis *Linear Regression* atau untuk mengetahui konsentrasi data di sekitar garis *Linear Regression* [26]. RMSE merupakan akar kuadrat dari MSE [27]. Rumus dari RMSE ditunjukkan pada persamaan (4).

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (4)$$

Koefisien determinasi (R^2) merupakan proporsi atau persentase dari total variasi Y yang dijelaskan oleh garis regresi [28]. Koefisien determinasi adalah kuadrat koefisien korelasi. Koefisien determinasi ini digunakan untuk mengetahui presentase pengaruh yang terjadi dari variabel bebas terhadap variabel terikat. Dalam konteks *Linear Regression*, R^2 dihitung dengan persamaan (5).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

Model yang telah dievaluasi kemudian divisualisasikan. Visualisasi data menampilkan grafis dari informasi dan data [29]. Visualisasi data menggunakan elemen visual berupa bagan, grafik, dan peta [30]. Visualisasi data memungkinkan penyajian informasi dalam bentuk yang lebih mudah diakses untuk memahami pola dalam suatu dataset. Proses ini mengubah kumpulan data yang kompleks menjadi representasi yang lebih sederhana dan intuitif.

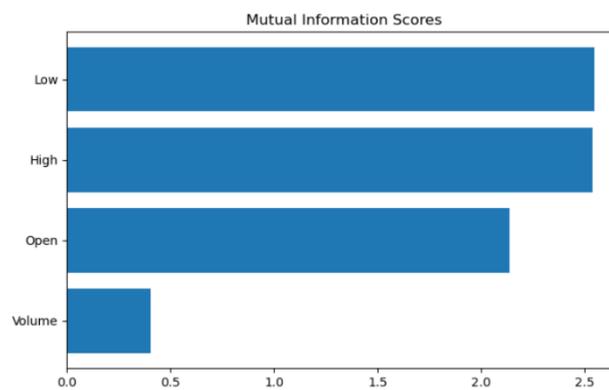
III. HASIL DAN PEMBAHASAN

Hasil dari seleksi fitur menggunakan MI terhadap nilai *Adj Close* harga saham ditunjukkan pada Gambar 3. Dari hasil tersebut, fitur *Low* dan *High* merupakan atribut yang memiliki skor tertinggi, yaitu

2.5 menunjukkan bahwa kedua fitur ini memiliki korelasi paling tinggi terhadap nilai *Adj Close*. Fitur *Open* memiliki skor yang lebih rendah dibandingkan *Low* dan *High*, tetapi masih signifikan pada angka 2.0. *Volume* memiliki skor terendah di antara keempat fitur, berada di bawah angka 0.5. Hal ini menunjukkan bahwa nilai fitur *Volume* memiliki korelasi yang paling rendah dengan variabel target dan memberikan informasi yang paling sedikit dibandingkan dengan fitur lainnya.

Tabel 2 menunjukkan hasil evaluasi antara *Linear Regression* dan *Polynomial Regression*. Dari masing-masing evaluasi, model *Linear Regression* memiliki nilai RMSE yang sama dengan *Polynomial Regression* dengan *degree*=1, tetapi MSE untuk *polynomial* memiliki nilai yang lebih kecil dan R^2 yang lebih besar dibandingkan dengan *Linear Regression*. Jika pada model polinomial, semakin tinggi nilai *degree* semakin baik nilai evaluasinya, dengan *degree*=3 memiliki hasil evaluasi yang paling baik. Namun, ketika model menggunakan *degree*=4 hasil evaluasi jauh memburuk secara signifikan. Hal ini dapat menunjukkan *degree*=3 merupakan nilai yang paling optimal dalam menangkap pola pada data, sedangkan *degree* yang tinggi seperti *degree*=4 menunjukkan model menjadi terlalu kompleks terlalu menyesuaikan diri pada data sehingga dapat menyebabkan *overfitting* dan kehilangan kemampuan untuk melakukan generalisasi pada data baru.

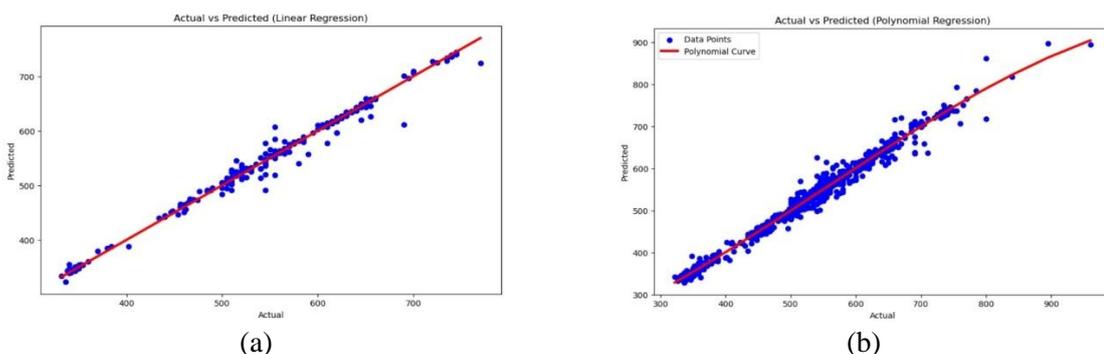
Hasil prediksi beserta nilai aktual divisualisasikan pada Gambar 4. Gambar 4a adalah visualisasi model *Linear Regression* dan Gambar 4b adalah visualisasi untuk *Polynomial Regression*. Dari visualisasi tersebut menunjukkan bahwa *Polynomial Regression* lebih baik dalam menangkap pola terlihat dari lengkungan, terutama untuk pola non-linear. Selain itu, *Polynomial Regression* lebih mampu dalam menyesuaikan diri dengan fluktuasi data yang kompleks.



Gambar 3. Hasil Mutual Information Scores

Tabel 2. Hasil Evaluasi *Linear Regression* dan *Polynomial Regression*

Metode	Degree	RMSE	MSE	R2
<i>Linear Regression</i>		13.8128417787	190.794598003304	0.9819065719
	1	13.8128417787	190.7945980021	0.9819068836
<i>Polynomial Regression</i>	2	12.5991735065	158.7391730473	0.9849467104
	3	11.3923259465	129.7850904722	0.9876924359
	4	38.6151342136	1491.1285903378	0.8585957700



Gambar 4. (a) Evaluasi *Linear Regression* Tanpa Regularisasi, (b) Evaluasi *Polynomial Regression* Tanpa Regularisasi

Tabel 3 menunjukkan nilai evaluasi model *Linear Regression*, *Polynomial Regression*, beserta kedua model yang mengimplementasikan *regularization* L1 (*LASSO*) dan L2 (*Ridge*). Terlihat pada nilai perbandingan evaluasi prediksi *Linear Regression* dengan *Polynomial Regression*, nilai pada *Polynomial Regression* mendapatkan nilai MSE dan RMSE yang lebih rendah serta nilai R² yang lebih tinggi daripada metode *Linear Regression* menandakan metode *Polynomial Regression* memberikan performa yang lebih baik dengan akumulasi persentase sebesar 50.9%. Kemudian dengan menambahkan fungsi L1 metode *Polynomial Regression* mendapat nilai akumulasi persentase sebesar 26.98% lebih baik daripada model *Linear Regression*. Dengan L2 nilai metode *Polynomial Regression* mendapatkan nilai akumulasi persentase sebesar 55.93% yang lebih baik daripada model *Linear Regression*.

Jika dibandingkan berdasarkan regularisasi, *Ridge* memberikan hasil evaluasi yang sama pada MSE dan RMSE, serta menghasilkan R² yang lebih baik meskipun tidak secara signifikan. Hal ini menunjukkan bahwa penambahan regularisasi *Ridge* tidak memberikan dampak yang signifikan pada model. *Linear Regression* dengan penambahan *LASSO* menurunkan performansi pada RMSE dan MSE dalam jumlah yang tidak terlalu signifikan, tetapi R² menghasilkan nilai yang lebih baik dengan jumlah yang tidak terlalu signifikan. Adapun pada *Polynomial Regression*, *LASSO* tidak meningkatkan performansi dari model. Berbeda dengan *Ridge* yang meningkatkan performansi model. Hal ini dapat terjadi karena berdasarkan konsep kerja dari *LASSO* yang menyusutkan koefisien mendekati nol dan dapat menjadi seleksi fitur, menyebabkan model kehilangan informasi penting dan relevan bagi model.

Gambar 5a merupakan sebagian kecil nilai hasil prediksi *Adj Close* dari metode *Linear Regression*. Model memberikan hasil mendekati hasil aktual pada baris data akhir (1229-1233), terutama pada fungsi dari *Predicted Adj Close* dan L2. Gambar 5b merupakan sebagian kecil nilai hasil prediksi *Adj Close* dari metode *Polynomial Regression*. Model memberikan hasil prediksi mendekati data aktual pada baris data awal (0-4), terutama pada fungsi dari *Predicted Adj Close* dan L2.

Tabel 3. Hasil Evaluasi Model beserta Regularisasi

	<i>Linear Regression</i>	<i>Polynomial Regression</i>	Jumlah Selisih Nilai	Persentase Selisih
MSE	190.7945980033	129.7850904722	61.0095075311	31.98%
RMSE	13.8128417787	11.3923259465	2.4205158322	17.52%
R ²	0.9819065719	0.9876924359	0.0057858639	0.59%
L1 LASSO:				
MSE	190.7946531146	157.4188177753	33.3758353393	17.49%
RMSE	13.8128437736	12.5466656039	1.2661781698	9.17%
R ²	0.9819068836	0.9850719201	0.0031650366	0.32%
L2 Ridge:				
MSE	190.7945980033	122.9618536501	67.8327443532	35.55%
RMSE	13.8128417787	11.0888166028	2.7240251759	19.72%
R ²	0.9819068836	0.9883394857	0.0064326022	0.66%

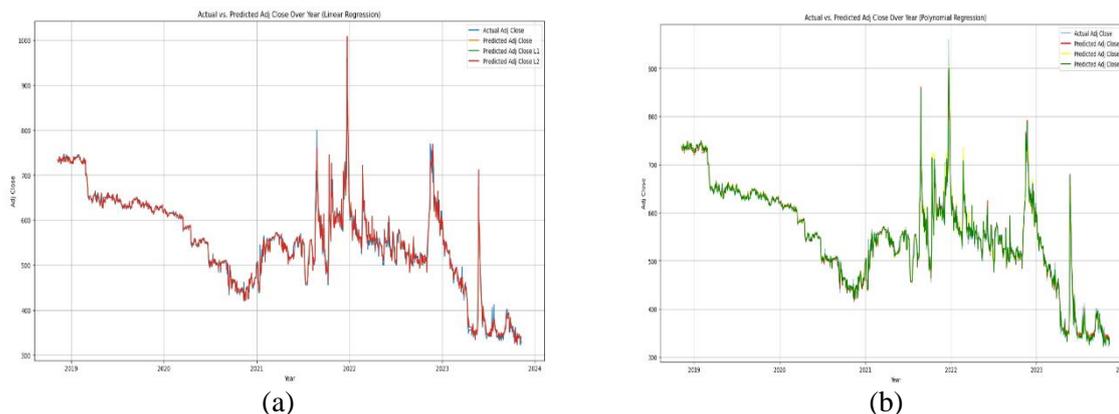
	Actual Adj Close	Predicted Adj Close	L1 Predicted Adj Close	L2 Predicted Adj Close
0	735.0	734.169663	734.174609	734.169646
1	730.0	729.254630	729.259512	729.254614
2	730.0	728.725382	728.727127	728.725359
3	735.0	733.641204	733.643044	733.641181
4	740.0	736.359820	736.363128	736.359800
...
1229	342.0	341.331748	341.328381	341.331732
1230	326.0	338.695854	338.692737	338.695842
1231	322.0	336.016513	336.012000	336.016504
1232	342.0	336.279831	336.279625	336.279849
1233	338.0	327.566191	327.567211	327.566234

(a)

	Actual Adj Close	Predicted Adj Close	L1 Predicted Adj Close	L2 Predicted Adj Close
0	735.0	734.682139	737.550466	738.396472
1	730.0	729.924905	732.387374	733.161007
2	730.0	728.999116	731.782943	729.283656
3	735.0	733.887461	737.176032	734.515080
4	740.0	735.806444	738.836890	738.266092
...
1229	342.0	345.007030	339.151996	341.761806
1230	326.0	339.919157	335.692116	335.042965
1231	322.0	341.542208	336.220943	339.089352
1232	342.0	339.397635	335.992846	335.383741
1233	338.0	331.774832	329.586357	327.373945

(b)

Gambar 5. (a) Prediksi Metode *Linear Regression*, (b) Prediksi *Polynomial Regression*



Gambar 6. Visualisasi Metode (a) *Linear Regression*, (b) *Polynomial Regression*

Gambar 6a dan 6b menampilkan hasil visualisasi prediksi model untuk harga penutupan yang disesuaikan (*Adj Close*) dari waktu ke waktu. Model *Linear Regression* menunjukkan prediksi yang mendekati nilai aktual. Garis-garis prediksi hampir selalu berdekatan dengan garis nilai aktual, menunjukkan bahwa model ini mampu menangkap pola harga dengan cukup baik. Kemudian menggunakan regularisasi L1 dan L2 memberikan hasil yang sangat mirip dengan prediksi *Linear Regression* standar, menunjukkan bahwa penambahan regularisasi tidak banyak mengubah hasil prediksi. Model *Polynomial Regression* cukup responsif terhadap lonjakan dan penurunan yang drastis pada data aktual, terutama di sekitar tahun 2022. Hal ini menunjukkan bahwa model *Polynomial Regression* yang digunakan mampu mengadaptasi perubahan drastis dalam tren harga. Selama periode stabil, prediksi model mengikuti pola data aktual dengan cukup baik, meskipun mungkin ada sedikit perbedaan pada beberapa titik.

IV. KESIMPULAN

Dari hasil analisis, fitur *Low* dan *High* memiliki korelasi paling tinggi terhadap nilai *Adj Close*, sedangkan *Volume* memiliki korelasi terendah. Model *Polynomial Regression* dengan *degree* = 3 terbukti memberikan performa terbaik, mampu menangkap pola data yang kompleks tanpa *overfitting*, yang mulai muncul pada *degree* = 4. Dibandingkan *Linear Regression*, *Polynomial Regression* menghasilkan nilai MSE dan RMSE yang lebih rendah, serta R^2 lebih tinggi, terutama dengan tambahan regularisasi *Ridge* yang meningkatkan performa yaitu mencapai MSE 122.9618, RMSE 11.0888, R^2 0.9883. Sementara itu, regularisasi *LASSO* tidak memberikan pengaruh signifikan pada model *Polynomial Regression* dan cenderung mengurangi efektivitas dengan menyusutkan koefisien fitur penting.

UCAPAN TERIMA KASIH

Kami ingin mengucapkan terima kasih kepada Institut Bisnis dan Teknologi Indonesia (INSTIKI) atas hibah INSTIKI *Research And Development Program* (IRDP) sehingga penelitian ini dapat dilaksanakan.

DAFTAR PUSTAKA

- [1] A. Rosyidta Pratiwi Octasyilva *et al.*, “Analisis Teknikal Saham Kontruksi di Indonesia,” *J. ILMU Pengetah. DAN Teknol.*, vol. 6, no. 2, pp. 23–32, Aug. 2022, doi: 10.31543/JII.V6I2.211.
- [2] T. Febiyola, R. S. Utari, B. T. Panggabean, and R. Agustina, “Analisis Surat Berharga Sebagai Alat Investasi,” *Depos. J. Publ. Ilmu Huk.*, vol. 2, no. 3, pp. 75–86, Jun. 2024, doi: 10.59581/DEPOSISI.V2I3.3688.
- [3] M. Fiddin and P. A. Wulandari, “Analisis Teknikal Sebagai Dasar Pengambilan Keputusan Dalam Transaksi Saham,” *EKALAYA J. Ekon. Akunt.*, vol. 2, no. 1, pp. 84–96, Mar. 2024, doi: 10.59966/EKALAYA.V2I1.711.
- [4] M. Mustaqim, D. C. Putrihadiningirum, and N. Wahyuningtiyas, “Analisis Teknikal Saham Kalbe Farma dengan Menggunakan Moving Average Convergence Divergence dan Stochastic Oscillator Selama Pandemi Covid-19 Periode 2020-2021,” *Al-Muhasib J. Islam. Account. Financ.*, vol. 2, no. 1, pp. 98–118, Sep. 2022, doi: 10.30762/ALMUHASIB.V2I1.186.
- [5] D. Maulud, D. Maulud, and A. M. Abdulazeez, “A Review on Linear Regression Comprehensive

- in Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 1, no. 2, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.
- [6] J. Penelitian, P. Masyarakat, M. A. Pratama, and W. Priatna, “Utilizing Linear Regression for Predicting Sales of Top-Performing Products,” *Int. J. Inf. Technol. Comput. Sci. Appl.*, vol. 1, no. 3, pp. 174–180, Sep. 2023, doi: 10.58776/IJITCSA.V1I3.92.
- [7] B. Purnomosidi, D. Putranto, M. A. Kholik, M. A. Nugroho, and D. Kriestanto, “Polynomial Regression Method and Support Vector Machine Method for Predicting Disease Covid-19 in Indonesia,” *J. Intell. Softw. Syst.*, vol. 2, no. 1, pp. 18–23, Jul. 2023, doi: 10.26798/JISS.V2I1.931.
- [8] S. Sanyal, S. Kumar Biswas, D. Das, M. Chakraborty, and B. Purkayastha, “Boston House Price Prediction Using Regression Models,” *2022 2nd Int. Conf. Intell. Technol. CONIT 2022*, 2022, doi: 10.1109/CONIT55038.2022.9848309.
- [9] A. Eka, A. Juarna, T. Informatika, F. T. Industri, and U. Gunadarma, “Prediksi Pro duksi Daging Sapi Nasional dengan Meto de Regresi Linier dan Regresi Polinomial,” *J. Ilm. Komputasi*, vol. 20, no. 2, pp. 209–215, 2021, doi: 10.32409/jikstik.20.2.2722.
- [10] D. I. Mulyana and Marjuki, “Optimasi Prediksi Harga Udang Vaname Dengan Metode Rmse Dan Mae Dalam Algoritma Regresi Linier,” *J. Ilm. Betrik*, vol. 13, no. 1, pp. 50–58, 2022, doi: 10.36050/betrik.v13i1.439.
- [11] I. P. Putra and I. Ketut Gede Suhartana, “Perbandingan Akurasi Algoritma Regresi Linier, Regresi Polinomial, dan Support Vector Regression Pada Model Sistem Prediksi Harga Rumah,” *J. Nas. Teknol. Inf. dan Apl.*, vol. 1, no. 1, pp. 147–154, 2022, [Online]. Available: <https://www.kaggle.com/datasets/wisnuanggara/daftar-hargarumah>
- [12] S. A. L. Satriyo, Adi Rizky Pratama, and Rahmat, “Perbandingan metode linear regresi dan polynomial regresi untuk memprediksi harga saham studi kasus Bank BCA,” *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 59–70, 2023, doi: 10.37373/infotech.v4i1.602.
- [13] S. Saadah, F. Z. Z, and H. H. Z, “Support Vector Regression (SVR) Dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR/USD,” *J. Comput. Sci. Informatics Eng.*, vol. 5, no. 1, pp. 85–92, 2021, doi: 10.29303/jcosine.v5i1.403.
- [14] J. Pebralia, “Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook,” *J. Ilmu Fis. dan Pembelajarannya*, vol. 6, no. 2, pp. 23–30, 2022, doi: 10.19109/jifp.v6i2.13958.
- [15] M. Rafi Muttaqin, T. Iman Hermanto, M. Agus Sunandar, P. Studi Teknik Informatika, and S. Tinggi Teknologi Wastukencana, “Penerapan K-Means Clustering dan Cross-Industry Standard Process For Data Mining (CRISP-DM) untuk Mengelompokan Penjualan Kue,” *Journal.Unpak.Ac.Id*, vol. 191. Rafi, no. 1, pp. 38–53, 2022, [Online]. Available: <http://journal.unpak.ac.id/index.php/komputasi/article/view/3976>
- [16] A. A. Azahra, “Analisis Prediksi Jumlah Penerimaan Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana,” *Bull. Appl. Ind. Eng. Theory*, vol. 3, no. 1, pp. 75–78, 2022.
- [17] Wilantapoera dkk, “Analisis Sentimen Kategori Aspek Pada Ulasan Produk Menggunakan Metode KNN Dengan Seleksi Fitur Mutual Information Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung Lembar Persetujuan Analisis Sentimen Kategori Aspek Pada U,” vol. 10, no. 2, pp. 1673–1681, 2022.
- [18] W. Han, H. Chen, and S. Poria, “Improving Multimodal Fusion with Hierarchical Mutual Information Maximization for Multimodal Sentiment Analysis,” *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, no. Mi, pp. 9180–9192, 2021, doi: 10.18653/v1/2021.emnlp-main.723.
- [19] D. S. Ramadhansyah, “Perbandingan Metode Seleksi Fitur Filter, Wrapper, Dan Embedded Prediksi Kandungan Vitamin C Pada Buah Mangga Menggunakan Metode Linear Regression Dan ...,” no. mm, 2022.
- [20] I. Mamlua’atul Mufidah and H. Basuki, “Analisis Regresi Linier Berganda Untuk Mengetahui Faktor Yang Mempengaruhi Kejadian Stunting Di Jawa Timur,” *Indones. Nurs. J. Educ. Clin.*, vol. 3, no. 3, pp. 51–59, 2023.
- [21] S. Wang, Y. Chen, Z. Cui, L. Lin, and Y. Zong, “Diabetes Risk Analysis based on Machine Learning LASSO Regression Model,” *J. Theory Pract. Eng. Sci.*, vol. 4, no. 1, pp. 58–64, 2024, doi: 10.53469/jtpes.2024.04(01).08.

- [22] A. Lesmana, S. Rina, F. Sari, and R. Widyasari, “Penerapan Lasso Kematian Pasien Komorbid Yang Terinfeksi Covid-19,” *J. Penelit. Mat. dan Pendidik. Mat.*, vol. 7, no. 1, pp. 290–296, 2024.
- [23] S. Suritman, R. Raupong, A. Kalondeng, D. Statistika, and F. Matematika dan Ilmu Pengetahuan Alam, “Pemodelan Mixed Geographically Weighted Regression yang Mengandung Multikolinearitas dengan Regresi Ridge,” *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, pp. 2721–379, 2023, doi: 10.20956/ejsa.vi.25426.
- [24] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, “Prediksi Penjualan Kendaraan Menggunakan Regresi Linear : Studi Kasus pada Industri Otomotif di Indonesia,” vol. 4, pp. 1199–1216, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative%0APrediksi>
- [25] W. Andriani, Gunawan, and A. E. Prayoga, “Prediksi Nilai Emas Menggunakan Algoritma Regresi Linear,” *J. Ilm. Inform. Komput.*, vol. 28, no. 1, pp. 27–35, 2023, doi: 10.35760/ik.2023.v28i1.8096.
- [26] A. S. B. Karno, “Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory),” *J. Inform. Inf. Secur.*, vol. 1, no. 1, pp. 1–8, 2020, doi: 10.31599/jiforty.v1i1.133.
- [27] A. Wibowo and E. Setiadi, “Uji Kelayakan Konstruksi Balok Pt Daeyung Dari Aplikasi Sap2000 Menggunakan Machine Learning Dengan Model Linier Regression,” *J. Tek. Sipil Cendekia*, vol. 5, no. 1, pp. 799–814, 2024, doi: 10.51988/jtsc.v5i1.185.
- [28] J. Karch, “Improving on Adjusted R-Squared,” *Collabra Psychol.*, vol. 6, no. 1, p. 45, Sep. 2020, doi: 10.1525/collabra.343.
- [29] E. D. Madyatmadja, M. N. Ridho, A. R. Pratama, M. Fajri, and L. Novianto, “Penerapan Visualisasi Data Terhadap Klasifikasi Tindak Kriminal Di Indonesia,” *Infotech J. Technol. Inf.*, vol. 8, no. 1, pp. 61–68, 2022.
- [30] I. Setya, D. Jayanti, and M. A. Khadija, “Business Intelligence Dashboard Data Pengunjung Kearsipan Di Dinas Perpustakaan Dan Kearsipan Kota Surakarta,” vol. 4, pp. 8629–8638, 2024.