

PERBANDINGAN MODEL MULTIPLE LINEAR REGRESSION DAN DECISION TREE REGRESSION (STUDI KASUS: PREDIKSI HARGA SAHAM TELKOM, INDOSAT, DAN XL)

¹Althof Thabibi, ²Raden Supriyanto

Fakultas Teknologi Industri Universitas Gunadarma
Jl. Margonda Raya No 100 Depok 16424, Jawa Barat
¹althof123.thabibi@gmail.com, ²supriyanto.r@gmail.com

Abstrak

Penggunaan machine learning menjadi salah satu bidang yang banyak digunakan seiring berkembangnya zaman. Melakukan prediksi dengan baik pada suatu data dapat dibantu dengan teknik dan model pada machine learning. Regresi menjadi salah satu teknik terbaik yang dapat digunakan dalam memprediksi suatu data. Tujuan penelitian ini yaitu untuk memanfaatkan machine learning menggunakan regresi dalam memprediksi sebuah data dengan baik. Tahapan penelitian dalam melakukan prediksi ini meliputi pengumpulan materi, analisis dan perancangan, uji coba, dan implementasi. Uji coba dilakukan dalam 3 skema pengujian, yaitu 80:20, 70:30, dan 60:40. Implementasi dilakukan dengan evaluasi error setiap model regresi untuk menentukan model terbaik dalam mengolah dataset. Hasil penelitian menggunakan model Multiple Linear Regression (MLR) dan Decision Tree Regression (DTR) memiliki perbedaan yang cukup besar. Model MLR memiliki nilai terbaik untuk MAE menggunakan skema 70:30 dengan nilai sebesar 0.008, sedangkan untuk RMSE dan MAPE menggunakan skema 60:40 dengan nilai sebesar 2.591 dan 0.852. Model DTR memiliki nilai terbaik menggunakan skema 80:20 dengan nilai MAE sebesar 0.013, nilai RMSE sebesar 4.014, dan nilai MAPE sebesar 1.287. Kesimpulan menunjukkan bahwa model MLR lebih bagus digunakan untuk dataset telekomunikasi selama 5 tahun terakhir karena memiliki nilai lebih kecil daripada model DTR walaupun dengan skema yang berbeda.

Kata Kunci: Decision Tree Regression, Machine learning, MAPE, Multiple Linear Regression, dan Prediksi Harga Saham.

Abstract

The use of machine learning is one of the fields that is widely used along with the times. Making good predictions on data can be assisted by techniques and models in machine learning. Regression is one of the best techniques that can be used to predict data. The purpose of this research is to take advantage of machine learning using regression in predicting data well. The research stages in making these predictions include material collection, analysis and design, trials, and implementation. The trials were carried out in 3 test schemes, namely 80:20, 70:30 and 60:40. Implementation is done by evaluating the errors of each regression model to determine the best model in processing the dataset. The results of the research using the Multiple Linear Regression (MLR) and Decision Tree Regression (DTR) models have quite a big difference. The MLR model has the best value for MAE using a 70:30 scheme with a value of 0.008, while for RMSE and MAPE it uses a 60:40 scheme with a value of 2.591 and 0.852. The DTR model has the best value using the 80:20 scheme with an MAE value of 0.013, an RMSE value of 4.014, and a MAPE value of 1.287. The conclusion shows that the MLR model is better used for telecommunications datasets for the last 5 years because it has a lower value than the DTR model even with a different scheme.

Keywords: Decision Tree Regression, Machine learning, MAPE, Multiple Linear Regression, & Stock Price Prediction.

PENDAHULUAN

Saat ini teknologi memiliki peran penting dalam melakukan aktifitas sehari-hari. Teknologi dapat membantu pekerjaan manusia menjadi lebih mudah karena sudah banyak teknologi yang sesuai untuk hampir semua bidang. Contohnya pada bidang ekonomi memanfaatkan teknologi *Artificial Intelligence (AI)* seperti *fraud detection* yang digunakan untuk pengamatan dan deteksi terhadap integritas keuangan yang tidak normal dan memicu penipuan. *Artificial intelligence* berhubungan dengan berkembangnya sistem komputer yang menggambarkan penalaran manusia ketika diterapkan pada domain pengetahuan rasional [1]. *Artificial intelligence* merupakan mesin yang mampu dalam mengenal bahasa manusia dan dapat menyelesaikan beberapa masalah yang kompleks dalam waktu yang singkat [2]. *AI* di dalamnya terdapat beberapa sub-bidang diantaranya yaitu *machine learning* dan *deep learning*. *Machine learning* berdasarkan definisinya adalah bidang ilmu komputer yang berkembang dari meninjau pengenalan pola dan teori pembelajaran komputasi dalam *artificial intelligence* [3]. *Machine learning* merupakan bagian dari *artificial intelligence* yang menyertakan sistem dengan manfaat dalam belajar secara otomatis dari konsep pengetahuan tanpa menjadi terprogram secara eksplisit. Pengalaman langsung mulai digunakan untuk mempersiapkan pola dan fitur data dan membuat hasil dan keputusan yang

lebih baik di masa depan [4]. *Machine learning* diperlukan untuk membuat komputer melakukan berbagai tugas sebagai pembelajaran dan diharapkan akan meningkatkan pengalamannya untuk memahami masalah yang lebih rumit.

Berkembangnya teknologi pada penggunaan *machine learning* saat ini, terdapat banyak teknik dan model yang dapat digunakan untuk menangani sebuah data. Untuk dapat menentukan model *machine learning* mana yang terbaik dalam menangani sebuah data yang akan diprediksi, regresi menjadi salah satu pilihan model yang dapat digunakan untuk memprediksi data dengan baik. Regresi adalah salah satu teknik penggunaan *Machine learning* yang berasal dari ilmu statistik untuk mendapatkan parameter-parameter dalam persamaan linear yang dapat menggambarkan *input* dan *output*. Persamaan linear dapat terdiri dari satu dimensi (linear) *input*, multi dimensi *input*, atau dengan persamaan polynomial [5].

Berdasarkan penjelasan diatas, karena itu penelitian ini dilakukan dengan membuat evaluasi model pada *machine learning* yang dapat memberikan prediksi dengan baik menggunakan model regresi. Terdapat banyak jenis model regresi pada *machine learning* yang digunakan untuk melakukan sebuah prediksi, tetapi pada penelitian ini akan fokus kepada 2 model regresi yaitu *Multiple Linear Regression* dan *Decision Tree Regression*. Berdasarkan penelitian terkait yang dilakukan oleh Abidatul Izzah dan Ratna Widyastuti,

Sachin Kamley, Shailesh Jaloree, dan R. S. Thakur model ini dipilih karena mudah digunakan, hasil prediksi tidak terlalu jauh dari aslinya, dan evaluasi error yang kecil. Setelah mendapatkan 2 model regresi, kemudian penelitian ini melakukan evaluasi pada data menggunakan teknik pengukur keakuratan model regresi dengan *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*.

Abidatul Izzah dan Ratna Widyastuti melakukan penelitian mengenai prediksi harga saham dengan model *Multiple Linear Regression* [6]. Keuntungan metode MLR ini menunjukkan error yang paling rendah, sedangkan dari segi kekurangannya adalah MLR K-Means tidak bagus dalam melakukan prediksi melainkan mengoreksi warna citra. Hasil penelitian yang didapat adalah model MLR menunjukkan pendekatan lebih baik dengan melihat dan membandingkan pengujian *error*-nya yaitu MSE, RMSE, dan MAPE.

Penelitian yang dilakukan Alaa Sheta, Sara Ahmed, dan Hossam Faris menggunakan model *Multiple Linear Regression* untuk memprediksi index pasar saham dan melakukan perbandingan dengan model ANN dan SVM [7]. Penelitian tersebut memperlihatkan hasil dari model MLR dengan grafik antara data asli dengan data hasil prediksi. Grafik MLR menunjukkan bahwa perbedaan antara nilai asli dengan hasil prediksi tidak terlalu jauh perbedaannya. Selain itu juga mendapatkan nilai R sebesar 0.9857 untuk

testing. Keunggulan model SVM ini lebih bagus dalam training dan prediksi, namun lebih cocok digunakan untuk menghindari jatuh atau turunnya data.

B. S. Bini dan Tessy Matthew melakukan penelitian terhadap prediksi saham menggunakan 2 teknik yaitu clustering dan regresi [8]. Setelah itu menggunakan model *Multiple Linear Regression* untuk melakukan prediksi harga saham yang akan datang. Keuntungan menggunakan teknik MLR dalam kasus ini adalah solusi terbaik untuk memprediksi harga saham yang akan datang, sedangkan kelemahannya yaitu dalam performa yang kurang dari teknik clustering. Gambar 3 menunjukkan hasil prediksi dengan menggunakan model *Multiple Linear Regression* yang berisikan 21 hari harga saham dan harga tutup saham. Hasil yang didapat dari penelitian tersebut dapat membantu para investor untuk memilih perusahaan yang tepat untuk investasi sehingga teknik *Multiple Linear Regression* dan teknik *Clustering* dapat dilakukan untuk memprediksi harga saham yang akan datang.

Sachin Kamley, Shailesh Jaloree, dan R. S. Thakur membuat pengulasan penelitian pada performa dalam melakukan prediksi pasar saham menggunakan teknik *machine learning* [9]. Keuntungan *Decision Tree* disebabkan oleh kemudahan dan kemampuannya dalam menemukan contoh data yang bernilai besar atau kecil dan melakukan prediksi terhadap nilainya. Kekurangannya yaitu prediksi yang

akurasi yang paling kecil. Tetapi hasil menunjukkan bahwa teknik *Decision Tree* melakukan prediksi dengan akurasi yang lebih tinggi dibandingkan dengan model lain karena prediksi jangka pendek dan juga signifikan dengan variabel yang digunakan saat melakukan prediksi. Tabel yang terdapat 11 jumlah data menggunakan teknik *Decision Tree* dari beberapa peneliti yang menunjukkan teknik *Decision Tree* yang digunakan serta hasil keakuratan performanya. Kesimpulan yang dapat diambil dari analisis tersebut adalah bahwa teknik hybrid seperti *Hybrid Decision Tree* bisa digunakan dan menghasilkan hasil yang lebih baik dari teknik *Decision Tree* biasa.

Tujuan dari penelitian ini adalah untuk mengimplementasikan dan membandingkan kinerja dari model *Multiple Linear Regression* dan *Decision Tree Regression* pada *dataset* saham Telkom, Indosat, dan XL dalam melakukan prediksi selama 5 tahun terakhir. Implementasi model MLR dan DTR terhadap *dataset* saham Telkom, Indosat, dan XL menghasilkan data prediksi yang akan dibandingkan. Hasil prediksi dibandingkan untuk mengetahui seberapa akurat hasil prediksi terhadap data asli yang diuji dengan

MAE, RMSE, dan MAPE. Hasil perbandingan kedua model menunjukkan kedekatan atau minimnya kesalahan yang dilakukan oleh model MLR dan DTR terhadap data asli.

METODE PENELITIAN

Tahap ini memiliki metode penelitian yang terdapat pada Gambar 1. Pembuatan flowchart program bertujuan untuk mengetahui alur kerja program. Tahap berikutnya yaitu memilih *dataset* sebagai data yang akan diprediksi oleh model regresi MLR dan DTR. Tahapan berikutnya yaitu menentukan rasio skema pengujian yang akan digunakan untuk *training* dan *testing*. Setelah memilih skema pengujian dilanjutkan dengan implementasi model MLR dan DTR terhadap *dataset*. Tahapan berikutnya yaitu melakukan uji kesalahan terhadap hasil yang sudah didapat dari implementasi MLR dan DTR. Uji kesalahan dilakukan menggunakan MAE, RMSE, dan MAPE dari hasil implementasi model MLR dan DTR sebelumnya untuk mengetahui tingkat keakuratan prediksi model regresi. Tahap terakhir yaitu mendapatkan hasil dan membuat kesimpulan dari penelitian yang dilakukan.



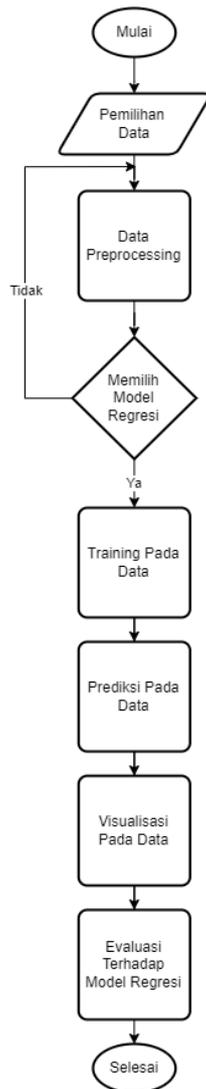
Gambar 1. Metode Penelitian

Flowchart Program

Penelitian ini akan dikerjakan melalui beberapa tahap seperti pengumpulan data, *data preprocessing*, dan *training model*. *Flowchart* program dapat membantu mengetahui alur kerja dari penelitian ini dengan memanfaatkan model regresi terhadap *dataset* yang akan diolah. *Flowchart* program terdapat pada Gambar 2.

Langkah pertama yang dilakukan adalah memilih data untuk menjadi bahan penelitian kemudian melakukan *data preprocessing*. *Data preprocessing* mencakup *import library* yang dibutuhkan, membaca *dataset*, melakukan pengecekan terhadap *dataset*, menghapus data pada *dataset*, dan membagi

dataset. Tahap selanjutnya adalah memilih model regresi yang akan digunakan untuk menangani *dataset*. *Training* tidak akan dilakukan jika belum memilih model regresi yang akan digunakan. Setelah melakukan *training* kemudian dilanjutkan dengan prediksi sesuai dengan model regresi yang akan menghasilkan data prediksi. Selanjutnya hasil prediksi akan dibuat dalam bentuk grafik *bar* dan *line* untuk membandingkannya dengan data asli. Tahap terakhir yaitu dengan melakukan evaluasi *error* terhadap model. Setiap model regresi akan dilakukan perhitungan untuk mengukur kinerjanya sehingga bisa menentukan model mana yang lebih bagus untuk melakukan prediksi.



Gambar 2. *Flowchart Program*

Pemilihan *Dataset*

Pencarian dan pemilihan *dataset* dilakukan melalui situs Yahoo Finance. Data dikumpulkan berdasarkan riwayat harga saham Telkom, Indosat, dan XL selama 5 tahun terakhir. Contoh untuk *dataset* saham Telkom dapat dilihat pada Tabel 1. Data dimulai dari tanggal 31 Agustus 2016 - 30 Agustus 2021. Terdapat 1260 jumlah data dan 7 *field* atau kolom yang berisikan Date, Open, High, Low, Close, Adj. Volume, dan Volume.

Fitur yang terdapat pada *dataset* sebagai berikut:

1. Date merupakan tanggal saham.
2. Open merupakan harga pembukaan saham.
3. High merupakan harga tertinggi saham.
4. Low merupakan harga terendah saham.
5. Close merupakan harga penutupan saham pada hari tersebut.
6. Adj Close merupakan harga penutupan yang disesuaikan.
7. Volume merupakan volume transaksi saham dalam jumlah lembar.

Tabel 1. *Dataset Saham Telkom*

Date	Open	High	Low	Close	Adj Close	Volume
2016-08-31	4160.0000 00	4240.0000 00	4150.0000 00	4210.0000 00	3611.5480 96	1190013 00
2016-09-01	4150.0000 00	4210.0000 00	4130.0000 00	4160.0000 00	3568.6555 18	7301940 0
2016-09-02	4120.0000 00	4210.0000 00	4120.0000 00	4200.0000 00	3602.9694 82	6779720 0
2016-09-05	4200.0000 00	4240.0000 00	4180.0000 00	4200.0000 00	3602.9694 82	4986040 0

Rasio Skema Pengujian

Pengujian pada penelitian ini dilakukan dengan membagi *dataset* menjadi *train set* dan *test set*. Tahap ini bertujuan untuk membangun model prediksi dalam menentukan model mana yang terbaik. Pembagian pada *dataset* dilakukan dengan 3 rasio, yaitu pertama menggunakan 80:20 dimana 80 persen dari data digunakan untuk *train set* dan 20 persen dari data digunakan untuk *test set*. Rasio kedua menggunakan 70:30, yaitu 70 persen dari data digunakan untuk *train set* dan 30 persen lainnya digunakan untuk *test set*. Rasio ketiga menggunakan 60:40, yaitu 60 persen dari data digunakan untuk *train set* dan 40 persen lainnya digunakan untuk *test set*.

Multiple Linear Regression

Multiple Linear Regression merupakan teknik regresi yang dapat digunakan dalam memprediksi sebuah data berdasarkan dari beberapa faktor atau variabel yang akan mempengaruhi data di masa depan. *Multiple Linear Regression* merupakan kebijakan tercatat yang menggunakan beberapa variabel independen untuk mengharapkan hasil dari variabel reaksi. Tujuan dari *Multiple Linear Regression* adalah untuk memodelkan hubungan lurus di tengah-tengah variabel independen tersebut selanjutnya variabel dependen [10]. Untuk dapat menghitung menggunakan model *Multiple Linear Regression* dapat dilihat pada persamaan (1).

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n \quad (1)$$

y adalah variabel dependen yang nilainya tetap sedangkan x_1 merupakan variabel independen yang mempengaruhi nilai dari variabel dependen. b_1 adalah koefisien untuk x_1 dan yang menghubungkannya dengan y , kemudian b_0 merupakan konstan. x_2 merupakan variabel independen lain pada *dataset* dan b_2 merupakan koefisiennya.

Decision Tree Regression

Setelah menerapkan model *Multiple Linear Regression* kepada *dataset*, selanjutnya menerapkan model *Decision Tree Regression*. *Decision Tree* adalah proses rekursif dari *binary tree* dalam membagi *dataset* sampai tersisa simpul daun yang sederhana yaitu data dengan 1 tipe golongan.

Decision Tree Regression bekerja dengan cara menguraikan proses pengambilan keputusan dari kompleks menjadi lebih sederhana. *Decision Tree Regression* akan mengelompokkan data yang akan digunakan dalam pemecahan masalah.

MAE

Evaluasi terhadap model regresi MLR dan DTR dilakukan dengan pengukuran kesalahan atau akurasi menggunakan MAE,

RMSE dan MAPE dari hasil implementasi model regresi. *Mean Absolute Error* (MAE) merupakan sebuah teknik pengukuran keakuratan sebuah model *machine learning* ketika berhasil mengerjakan *training* dan *predicting* pada sebuah data. Hasil yang didapat dari prediksi serta data asli akan digunakan untuk menunjukkan rata-rata kesalahan mutlak berdasarkan metode yang digunakan [11]. MAE memiliki persamaan sebagai berikut:

$$MAE = \frac{1}{N} \sum_{k=1}^N |y_k - \hat{y}_k| \quad (0)$$

y_k merupakan nilai asli, sedangkan \hat{y}_k merupakan nilai hasil prediksi, dan N merupakan jumlah sampel data.

RMSE

Evaluasi model untuk mengukur tingkat akurasi atau kesalahan selanjutnya dari hasil model regresi adalah RMSE. *Root Mean Squared Error* (RMSE) merupakan sebuah

teknik untuk melakukan pengukuran akar kuadrat dari rata-rata kuadrat dalam memastikan semua *error* pada prediksi numerik. Hasil dari RMSE akan semakin akurat jika mendekati 0, sedangkan jika menjauhi 0 maka prediksi semakin tidak akurat [12]. Pengukuran akurasi RMSE memiliki persamaan sebagai berikut:

$$RMSE = \sqrt{\frac{\sum (y_k - \hat{y}_k)^2}{N}} \quad (3)$$

y_k merupakan nilai asli, kemudian \hat{y}_k merupakan nilai hasil prediksi, dan N merupakan jumlah sampel data.

MAPE

Evaluasi model terakhir untuk diperhitungkan kesalahan atau akurasinya dari

hasil model regresi adalah MAPE. *Mean Absolute Percentage Error* (MAPE) berguna ketika ukuran atau besar variabel ramalan itu penting dalam menguji ketepatan ramalan.

MAPE menunjukkan seberapa besar kesalahan dalam meramal yang dibandingkan dengan nilai nyata dalam deret. [13]. Pengukuran akurasi MAPE memiliki persamaan sebagai berikut:

$$MAE = \frac{1}{N} \sum_{k=1}^N \left| \frac{y_k - \hat{y}_k}{y_k} \right| \times 100 \quad (4)$$

y_k merupakan nilai asli, kemudian \hat{y}_k merupakan nilai hasil prediksi, dan N merupakan jumlah sampel data.

test set menggunakan rasio skema pengujian dari 80:20, 70:30 dan 60:40.

HASIL DAN PEMBAHASAN

Pada bagian ini dijelaskan mengenai hasil *preprocessing* data, hasil dari *training* data, prediksi data, pengujian kesalahan, beserta kesimpulan. *Preprocessing* data dilakukan sebelum mengolah data seperti memasukan data, membersihkan data, kemudian membagi data menjadi *train set* dan

Hasil Training

Jumlah data dalam 5 tahun terakhir di setiap *dataset* berjumlah 1260 data. Untuk pembagian data *training* digunakan rasio skema pengujian 80:20, 70:30 dan 60:40. Hasil *training* data dapat dilihat pada Tabel 2. Jumlahnya jika di total akan menjadi 1259 tetapi karena bahasa python dimulai dari indeks ke-0 jadi total tetap 1260 data.

Tabel 2. Hasil *Training* Data

Rasio Skema Pengujian	Train Set	Test Set
80 : 20	1007	252
70 : 30	881	378
60 : 40	755	504

Tabel 0. Hasil Prediksi *Multiple Linear Regression* Telkomsel

Model	Indeks	Actual	Telkomsel		
			80 : 20 Predicted	70 : 30 Predicted	60 : 40 Predicted
<i>Multiple Linear Regression</i>	0	4210.0	4160.0	4160.0	4160.0
	1	3650.0	3620.0	3630.0	3630.0
	2	3790.0	3710.0	3710.0	3710.0
	3	3090.0	3090.0	3090.0	3090.0
	4	4350.0	4330.0	4330.0	4330.0
	5	4280.0	4240.0	4240.0	4240.0

Hasil Prediksi

Tahap ini akan menampilkan hasil prediksi dari 2 model *machine learning* yang digunakan yaitu *Multiple Linear Regression* dan *Decision Tree Regression*. Data yang diambil berjumlah 6 data dimulai dari indeks ke-0 hingga ke-5 dan menggunakan 3 rasio pembagian data, yaitu bagian pertama menggunakan perbandingan sebesar 80:20, bagian kedua sebesar 70:30, dan bagian terakhir sebesar 60:40. Tabel 3 merupakan hasil prediksi pada dataset Telkomsel yang dekat dengan data aslinya yaitu dengan skema 70:30 dan 60:40 pada indeks ke-1.

Tabel 4 menunjukkan hasil prediksi pada

dataset Indosat menggunakan MLR yang memiliki pendekatan terbaik di semua skema karena semuanya mendekati data asli dengan nilai yang serupa dan tidak jauh dari data aslinya.

Tabel 5 menghasilkan prediksi dari MLR pada *dataset* XL yang menunjukkan pendekatan terbaik dengan semua skema dikarenakan hasil prediksi yang sama dan tidak berbeda jauh dari data aslinya.

Tabel 6 memperlihatkan hasil prediksi menggunakan model regresi DTR pada *dataset* Telkomsel dan memiliki pendekatan terbaik pada semua skema yang terdapat di indeks ke-3 yaitu dengan nilai 3090.0.

Tabel 4. Tabel Hasil Prediksi *Multiple Linear Regression* Indosat

Indosat					
Model	Indeks	Actual	80 : 20	70 : 30	60 : 40
			Predicted	Predicted	Predicted
<i>Multiple Linear Regression</i>	0	5500.0	5480.0	5480.0	5480.0
	1	3490.0	3510.0	3510.0	3510.0
	2	6100.0	6150.0	6150.0	6150.0
	3	2570.0	2600.0	2600.0	2600.0
	4	6775.0	6820.0	6820.0	6820.0
	5	5625.0	5610.0	5610.0	5610.0

Tabel 5. Hasil Prediksi *Multiple Linear Regression* XL

XL					
Model	Indeks	Actual	80 : 20	70 : 30	60 : 40
			Predicted	Predicted	Predicted
<i>Multiple Linear Regression</i>	0	2770.0	2770.0	2770.0	2770.0
	1	2850.0	2800.0	2800.0	2800.0
	2	2280.0	2280.0	2280.0	2280.0
	3	2940.0	2940.0	2940.0	2940.0
	4	2960.0	2990.0	2990.0	2990.0
	5	2500.0	2490.0	2490.0	2490.0

Tabel 6. Hasil Prediksi *Decision Tree Regression* Telkomsel

Telkomsel					
Model	Indeks	Actual	80 : 20	70 : 30	60 : 40
			Predicted	Predicted	Predicted
<i>Multiple Linear Regression</i>	0	4210.0	4210.0	4210.0	4190.0
	1	3650.0	3620.0	3580.0	3580.0
	2	3790.0	3660.0	3660.0	3700.0
	3	3090.0	3090.0	3090.0	3090.0
	4	4350.0	4330.0	4330.0	4330.0
	5	4280.0	4280.0	4250.0	4250.0

Tabel 7. Hasil Prediksi *Decision Tree Regression* Indosat

Indosat					
Model	Indeks	Actual	80 : 20	70 : 30	60 : 40
			Predicted	Predicted	Predicted
<i>Multiple Linear Regression</i>	0	5500.0	5500.0	5500.0	5500.0
	1	3490.0	3370.0	3480.0	3440.0
	2	6100.0	6100.0	6100.0	6100.0
	3	2570.0	2660.0	2650.0	2550.0
	4	6775.0	6750.0	6750.0	6750.0
	5	5625.0	5575.0	5700.0	5575.0

Tabel 8. Hasil Prediksi *Decision Tree Regression* XL

XL					
Model	Indeks	Actual	80 : 20	70 : 30	60 : 40
			Predicted	Predicted	Predicted
<i>Multiple Linear Regression</i>	0	2770.0	2730.0	2770.0	2760.0
	1	2850.0	2770.0	2770.0	2730.0
	2	2280.0	2290.0	2340.0	2230.0
	3	2940.0	2910.0	2870.0	2930.0
	4	2960.0	2940.0	2980.0	2990.0
	5	2500.0	2470.0	2470.0	2470.0

Hasil prediksi pada Tabel 7 menunjukkan prediksi menggunakan model DTR dalam semua skema dan memiliki pendekatan terbaik yang sama terhadap data asli di semua skema pada indeks ke-0 dan sebesar 5500.0 dan indeks ke-2 sebesar 6100.0.

Model DTR dalam melakukan prediksi terhadap *dataset* XL pada Tabel 8 memiliki nilai pendekatan berbeda-beda dari sebelumnya. Prediksi dengan hasil yang sama di semua skema terdapat pada indeks ke-5 dengan nilai 2470.0, tetapi rata-rata pendekatan terbaik ada

pada skema 80:20.

Pengujian Error

Setelah melakukan prediksi kemudian melakukan perbandingan hasil pengukuran *error* pada model menggunakan *dataset* saham Telkom, Indosat, dan XL yang terdapat pada Tabel 9. Model MLR memiliki rata-rata MAE sebesar 0.009, RMSE sebesar 2.686, dan MAPE sebesar 0.906. Model DTR memiliki rata-rata MAE sebesar 0.013, RMSE sebesar 4.014, dan MAPE sebesar 1.287.

Tabel 9. Tabel Evaluasi Model Pada Saham (80:20)

Model	PT.	MAE	RMSE	MAPE (%)
<i>Multiple</i>	Telkom	0.006	1.777	0.590
<i>Linear</i>	Indosat	0.011	3.970	1.138
<i>Regression</i>	XL	0.010	2.312	0.989
Rata-Rata		0.009	2.686	0.906
Decision	Telkom	0.008	2.731	0.831
Tree	Indosat	0.016	6.024	1.605
Regression	XL	0.014	3.288	1.424
Rata-Rata		0.013	4.014	1.287
Selisih Rata-Rata		0.004	1.328	0.381

Tabel 10. Tabel Evaluasi Model Pada Saham (70:30)

Model	PT.	MAE	RMSE	MAPE (%)
<i>Multiple</i>	Telkom	0.006	1.779	0.582
<i>Linear</i>	Indosat	0.010	3.902	1.087
<i>Regression</i>	XL	0.009	2.206	0.948
Rata-Rata		0.008	2.629	0.872
Decision	Telkom	0.008	2.798	0.836
Tree	Indosat	0.016	5.939	1.606
Regression	XL	0.014	3.316	1.429
Rata-Rata		0.013	4.018	1.299
Selisih Rata-Rata		0.005	1.389	0.427

Tabel 11. Tabel Evaluasi Model Pada Saham (60:40)

Model	PT.	MAE	RMSE	MAPE (%)
<i>Multiple</i>	Telkom	0.006	1.744	0.565
<i>Linear</i>	Indosat	0.011	3.908	1.079
<i>Regression</i>	XL	0.009	2.122	0.913
Rata-Rata		0.009	2.591	0.852
Decision	Telkom	0.008	2.790	0.853
Tree	Indosat	0.017	6.522	1.678
Regression	XL	0.014	3.187	1.384
Rata-Rata		0.013	4.166	1.305
Selisih Rata-Rata		0.004	1.575	0.453

Evaluasi pada Tabel 10 untuk Model MLR memiliki rata-rata MAE sebesar 0.008, RMSE sebesar 2.629, dan MAPE sebesar 0.872. Model DTR memiliki rata-rata MAE sebesar 0.013, RMSE sebesar 4.018, dan MAPE sebesar 1.299. Selisih rata-rata untuk model MLR dan DTR pada MAE sebesar 0.005, kemudian pada RMSE sebesar 1.389, dan untuk MAPE sebesar 0.427.

Evaluasi pada Tabel 11 untuk Model MLR memiliki rata-rata MAE sebesar 0.009, RMSE sebesar 2.591, dan MAPE sebesar 0.852. Model DTR memiliki rata-rata MAE sebesar 0.013, RMSE sebesar 4.166, dan MAPE sebesar 1.305. Selisih rata-rata dari model MLR dan DTR pada MAE sebesar 0.004, kemudian pada RMSE sebesar 1.575, dan untuk MAPE sebesar 0.453.

Tabel 12. Evaluasi Keseluruhan

No.	Skema Pengujian			Model		Rata-Rata		
	80:20	70:30	60:40	MLR	DTR	MAE	RMSE	MAPE
1.	✓			✓		0.009	2.686	0.906
2.		✓		✓		0.008	2.629	0.872
3.			✓	✓		0.009	2.591	0.852
4.	✓				✓	0.013	4.014	1.287
5.		✓			✓	0.013	4.018	1.299
6.			✓		✓	0.013	4.166	1.305

Setelah itu melakukan evaluasi secara keseluruhan dengan hasil yang sudah diperoleh sebelumnya. Tabel 12 menampilkan skema pengujian, model *machine learning*, dan rata-rata MAE, RMSE, dan MAPE dari tabel sebelumnya. Terdapat 6 data yang ditampilkan mulai dari skema pengujian 80 : 20 hingga 60 : 40.

Kesimpulan yang didapat dari hasil evaluasi adalah rata-rata MAE, RMSE, dan MAPE menggunakan model *Multiple Linear Regression* lebih kecil dibandingkan menggunakan model *Decision Tree Regression*. Model MLR pada skema 80:20 memiliki nilai MAE sebesar 0.009, RMSE sebesar 2.686, dan MAPE sebesar 0.906. Model DTR pada skema yang sama yaitu 80:20 memiliki nilai MAE sebesar 0.013, RMSE sebesar 4.014, dan MAPE sebesar 1.287.

Skema 70:30 pada model MLR menghasilkan nilai sebesar 0.008, RMSE sebesar 2.626, dan MAPE sebesar 0.872. Model DTR pada skema 70:30 memiliki nilai MAE sebesar 0.013, RMSE sebesar 4.018, dan MAPE sebesar 1.299. Hasil rata-rata dengan skema 60:40 memperoleh nilai MAE sebesar

0.009, RMSE sebesar 2.591, dan MAPE sebesar 0.852. Model DTR pada skema 60:40 memiliki nilai MAE sebesar 0.013, RMSE sebesar 4.166, dan MAPE sebesar 1.305.

KESIMPULAN DAN SARAN

Kesimpulan yang didapat dari penelitian ini adalah model *Multiple Linear Regression* menghasilkan rata-rata *error* yang kecil daripada model *Decision Tree Regression*. Hasil evaluasi rata-rata MAE, RMSE, dan MAPE menggunakan model *Multiple Linear Regression* lebih kecil dibandingkan menggunakan model *Decision Tree Regression*. Perbandingan 2 model, yaitu Model MLR menghasilkan rata-rata MAE terbaik menggunakan skema pengujian 70 : 30 dengan nilai 0.008, sedangkan RMSE dan MAPE terbaik menggunakan skema pengujian 60 : 40 dengan nilai 2.591 dan 0.852. Model DTR menghasilkan rata-rata MAE, RMSE, dan MAPE terbaik dengan skema pengujian 80 : 20, yaitu dengan 0.013, 4.014, dan 1.287. Hasil perbandingan antara model MLR dan DTR menunjukkan bahwa model MLR memiliki

nilai lebih kecil dibandingkan dengan model DTR sehingga menjadikan model MLR sebagai model yang terbaik untuk penelitian ini.

Saran yang dapat diberikan pada penelitian ini adalah dengan melakukan perkembangan lebih lanjut menggunakan model lainnya seperti *Long Short Term Memory* dalam *Recurrent Neural Network* (RNN) yang dapat memprediksi harga saham untuk hari berikutnya berdasarkan beberapa hari terakhir. Selain itu juga dapat dilakukan percobaan pada *dataset* yang lebih luas dan kompleks.

DAFTAR PUSTAKA

- [1] A. Cardon, *Beyond Artificial Intelligence: From Human Consciousness to Artificial Consciousness*, ISTE Ltd, 2018.
- [2] A. V. Joshi, *Machine Learning and Artificial Intelligence*, Redmond, WA, USA: Springer, 2020.
- [3] D. Nagar, S. Raghay, A. Bhardwaj, R. Kumar, P. L. Singh and R. Sindhwani, "Machine learning: Best Way to Sustain the Supply Chain in the Era of Industry 4.0," *Material Today : Proceedings*, hal. 1-7, 2021.
- [4] S. Dargan, M. Kumar, M. R. Ayyagari and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, hal. 1071-1092, 2020.
- [5] I. Siradjuddin, "Machine Learning: Linear Regression (#1)," hal. 1-7, 2018.
- [6] A. Izzah and R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 2, no. 3, hal. 141-150, 2017.
- [7] Z. Berrardi, M. Lazaar, H. Omara and O. Mahboub, "Effect of Architecture in Recurrent Neural Network Applied on the Prediction of Stock Price," *IAENG International Journal of Computer Science*, vol. 47, no. 3, hal. 199-204, 2020.
- [8] M. Biswas, A. Shome, M. A. Islam, A. Nova and S. Ahmed, "Predicting Stock Market Price: A Logical Strategy Using Deep Learning," *Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, vol. 11, hal. 218-223, 2021.
- [9] F. H. Abdulraheem, M. Y. Al-Ridha and R. R. O. Al-Nima, "Predicting Death and Confirmed Cases of Coronavirus," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, hal. 380-385, February 2022.

- [10] D. K. Sharma and A. Sharma, "Prediction of Health Insurance Emergency using Multiple Linear Regression Technique," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 4, hal. 98-105, 2020.
- [11] A. A. Suryanto and A. Mustaqdir, "Penerapan Metode Mean Absolute Error (MEA) dalam Algoritma Regresi Linear untuk Prediksi Produksi Padi," *SAINTEKBU*, vol. 11, no. 1, hal. 78-83, 2019.
- [12] M. Zakwan, M. Muzzammil and J. Alam, "Application of Data Driven Techniques in Discharge Rating Curve - An Overview," *Aquademia*, vol. 1, no. 1, hal. 1-8, 2017.
- [13] M. B. S. Junianto, "Fuzzy Inference System Mamdani dan the Mean Absolute Percentage Error (MAPE) untuk Prediksi Permintaan Dompot Pulsa pada XL Axiata Depok," *Jurnal Informatika Universitas Pamulang*, vol. 2, no. 2, hal. 97-102, 2017.