

JURNAL DINDA

**Kelompok Keahlian Rekayasa Data
Institut Teknologi Telkom Purwokerto**

Vol. x No. x (20xx) xx - xx

ISSN Media Elektronik:

Prediksi Tren Harga Close Saham dengan Metode Regresi Linear (Studi Kasus Saham Telkomsel)

Rolanita Scenic Faravati¹

¹Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto

¹18102069@ittelkom-pwt.ac.id

Abstract

Covid-19 is a virus that was first detected in Wuhan, China which began to spread throughout the world. In anticipating this, the Indonesian government has extended the application of restrictions on community activities. This greatly affects the economy in Indonesia. Many Indonesians have gone out of business and experienced massive layoffs. This has made people start to switch to playing on stocks. However, during a pandemic like this, stock prices experience erratic increases and decreases. Therefore, in this experiment, a model is made using the Linear Regression method in predicting the trend of close stock prices in the case study of Telkomsel's shares. A confidence value of 83.28% was obtained, with predictions on the trend of close stock prices being quite accurate, but predictions on close stock prices not being too accurate.

Keywords: stock, telkomsel, linear regression, python

Abstrak

Covid-19 adalah virus yang terdeteksi pertama kali di Wuhan, Tiongkok yang mulai menyebar ke penjuru dunia. Dalam mengantisipasi hal tersebut, pemerintah Indonesia memperpanjang kembali Pemberlakuan Pembatasan Kegiatan Masyarakat. Hal ini sangat mempengaruhi perekonomian yang ada di Indonesia. Banyak masyarakat Indonesia yang mengalami gulung tikar hingga mengalami pemutusan hubungan kerja secara besar-besaran. Hal tersebut membuat masyarakat mulai beralih bermain pada saham. Akan tetapi, pada masa pandemi seperti ini, harga saham mengalami kenaikan dan penurunan yang tidak menentu. Oleh karena itu, pada percobaan kali ini dibuat sebuah model dengan menggunakan metode Regresi Linear dalam memprediksi tren harga saham close pada studi kasus saham Telkomsel. Didapatkan nilai confidence sebesar 83,31% dengan prediksi pada tren harga close saham cukup akurat, akan tetapi pada prediksi harga close saham belum terlalu akurat.

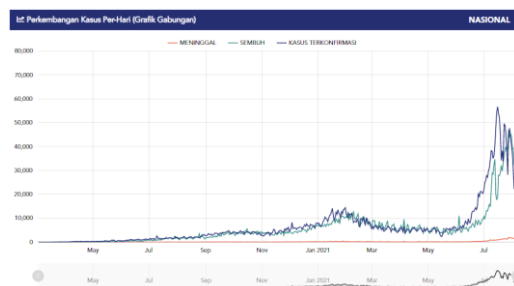
Kata kunci: saham, telkomsel, regresi linear, python

© 20xx Jurnal DINDA

1. Pendahuluan

Covid-19 adalah penyakit yang menyerang pada saluran pernapasan manusia. Virus ini terdeteksi pertama kali di Wuhan, Tiongkok yang mulai menyebar ke penjuru dunia. Berdasarkan data WHO per 5 Agustus 2021 tercatat total kasus positif mencapai 3.568.331 dengan jumlah 2.947.646 dinyatakan telah sembuh dan 102.375 dinyatakan meninggal dunia di Indonesia (<https://covid19.go.id/>). Dalam mengantisipasi hal tersebut, pemerintah Indonesia memperpanjang kembali Pemberlakuan Pembatasan Kegiatan Masyarakat Level 4 hingga 9 Agustus 2021 pada wilayah Jawa dan Bali

(<https://nasional.kompas.com/>). Hal ini sangat mempengaruhi perekonomian yang ada di Indonesia.



Gambar 1. Perkembangan Kasus Per-Hari Indonesia [1]

Diterima Redaksi : xx-xx-20xx | Selesai Revisi : xx-xx-20xx | Diterbitkan Online : xx-xx-20xx

Pada situasi seperti yang dijelaskan sebelumnya, menjadikan banyak masyarakat Indonesia yang kesulitan dengan finansial mereka. Menurut Shang, Hu & Ji (2020) dalam penelitian Purnaningrum (2020) menyebutkan bahwa ketidakpastian pada kondisi pandemi dan kerugian pada sektor ekonomi dapat menyebabkan pasar menjadi lebih fluktuatif dan sangat sulit dalam melakukan prediksi [2]. Hal tersebut dapat membuat para investor menjadi panik dalam melakukan penjualan dan pembelian saham dan dapat mengakibatkan penurunan pada harga saham [3].

Perpanjangan pada PPKM level 4 ini menyebabkan banyak masyarakat Indonesia yang mengalami gulung tikar hingga mengalami pemutusan hubungan kerja secara besar-besaran. Hal tersebut membuat masyarakat mulai beralih bermain pada saham. Akan tetapi, pada masa pandemi seperti ini, harga saham mengalami kenaikan dan penurunan yang tidak menentu. Sehingga cukup sulit untuk melakukan prediksi secara manual.

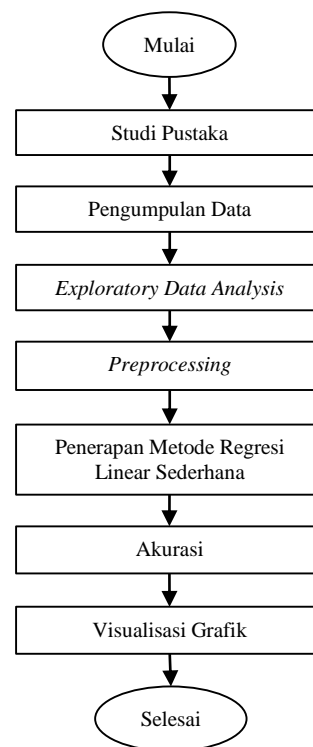
Machine learning sudah banyak digunakan dalam melakukan peramalan atau prediksi pada suatu kasus. Prediksi tren harga saham adalah topik yang cukup banyak dilakukan penelitian pada bidang machine learning. Salah satu metode yang digunakan pada peramalan atau prediksi tren harga saham adalah metode regresi linear.

Metode regresi linear sudah banyak digunakan pada penelitian yang membahas mengenai peramalan atau prediksi. Pada penelitian Muhartini, dkk (2021) melakukan analisis peramalan pada jumlah penerimaan mahasiswa baru dengan regresi linear sederhana dan mendapatkan tingkat akurasi sebesar 96,556% [4]. Kemudian pada penelitian Trianggana (2020) melakukan peramalan pada jumlah siswa dan siswi dengan pendekatan regresi linear dan mampu melakukan prediksi [5].

Berdasarkan penjelasan di atas, pada penelitian ini akan melakukan prediksi pada tren harga *close* saham dengan menggunakan metode regresi linear sederhana pada studi kasus saham telkomsel yang didapatkan pada laman finance.yahoo.com.

2. Metode Penelitian

Metode penelitian yang dilakukan melalui tahapan-tahapan sebagai berikut:



Gambar 2. Diagram Alir Penelitian

2.1. Studi Pustaka

Pada tahapan studi pustaka dilakukan pengumpulan studi literatur dari penelitian sebelumnya berupa jurnal, berita, buku dan lain sebagainya. Pada sumber bacaan yang menjadikan landasan dibuatnya penelitian ini adalah dengan penelitian terdahulu dengan metode yang sama. Sumber literatur ini diperoleh melalui *google scholar* dan *google books*.

2.2. Pengumpulan data

Pada tahapan pengumpulan data menggunakan dataset saham Telkomsel yang diambil melalui finance.yahoo.com dengan mengambil data dari tanggal 2 Januari 2015 hingga 30 Juli 2021 dengan data sebanyak 1647 *record* dengan data kolom yang digunakan adalah kolom Date, Open, High, Low, Close, Adj Close, dan Volume.

2.3. Exploratory Data Analysis

Pada tahapan *Exploratory Data Analysis* dilakukan untuk menampilkan data, kolom dan tipe data yang digunakan. Pada tahapan ini juga melakukan pemeriksaan pada missing value dengan menggunakan fungsi `info()`, statistik data dengan menggunakan fungsi `describe()` untuk menampilkan count (jumlah data), mean, std atau standar deviasi, nilai minimal dan maksimal serta kuartil 25%, 50% dan 75% pada data numerik yang ada pada dataframe. Kemudian dilakukan

pengecekan pada kurva skewness untuk menunjukkan distribusi atau persebaran data dengan melihat kemiringan yang divisualisasikan. Terakhir yaitu pengecekan outlier pada bagian kolom open, high, low, close, adj close, dan volume.

2.4. Preprocessing

Pada tahapan *preprocessing* data dilakukan data cleaning, kemudian menentukan jumlah dari sampel prediksi yang digunakan dan dimasukkan ke dalam dataframe baru. Selanjutnya dilakukan normalisasi data, menentukan data training dan testing serta menjadi nilai *confidence* pada *linear regression*.

2.5. Penerapan Metode Regresi Linear Sederhana

Setelah dilakukan pembagian pada data training dan data testing selanjutnya dilakukan prediksi pada data sampel prediksi yang kemudian dimasukkan ke dalam dataframe baru.

2.6. Akurasi

Pada pengecekan akurasi atau nilai *confidence* pada model Regresi Linear menggunakan perhitungan score pada model Regresi Linear yang sudah dilakukan training pada dataset. Kemudian hasil ini akan dimasukkan ke dalam dataset awal dengan menambahkan kolom Prediksi di dalamnya.

2.7. Visualisasi Grafik Prediksi

Pada tahapan terakhir yaitu visualisasi grafik pada prediksi tren harga close saham telkomsel dengan menggunakan Regresi Linear dengan menggunakan fungsi plot().

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Pada penelitian ini menggunakan dataset berupa saham Telkomsel yang diunduh melalui laman <https://finance.yahoo.com/> dari riwayat tanggal 2 Januari 2015 hingga 30 Juli 2021 dengan jumlah data sebanyak 1647 record. Kolom yang terdapat pada unduhan saham Telkomsel adalah kolom Date, Close, High, Low, Close, Adj Close dan Volume.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2015-01-02	2885.0	2885.0	2850.0	2860.0	2430.012939	18992100.0
1	2015-01-05	2850.0	2850.0	2820.0	2835.0	2408.771484	49940700.0
2	2015-01-06	2810.0	2835.0	2800.0	2815.0	2391.778809	47892100.0
3	2015-01-07	2865.0	2870.0	2795.0	2810.0	2387.530518	70076600.0
4	2015-01-08	2840.0	2845.0	2825.0	2835.0	2408.771484	56582500.0
...
1642	2021-07-26	3140.0	3200.0	3140.0	3190.0	3190.000000	55525800.0
1643	2021-07-27	3210.0	3270.0	3160.0	3270.0	3270.000000	62673400.0
1644	2021-07-28	3250.0	3250.0	3150.0	3190.0	3190.000000	103282300.0
1645	2021-07-29	3210.0	3290.0	3200.0	3230.0	3230.000000	126542400.0
1646	2021-07-30	3200.0	3290.0	3200.0	3240.0	3240.000000	95143100.0

Gambar 3. Dataset Saham Telkomsel

3.2. Exploratory Data Analysis

Pada tahapan *Exploratory Data Analysis* menampilkan terlebih dahulu dataset, kemudian melakukan pengaksesan data serta tipe data. Dalam melakukan tahapan ini menggunakan bahasa pemrograman python yang dikerjakan pada *google colab*. Library yang digunakan pada tahapan ini adalah *pandas* dan *numpy*.

Program Jurnal

```
dataset_telkom = pd.read_csv("TLKM.JK.csv")
print(dataset_telkom.columns)
dataset_telkom.dtypes
```

Selanjutnya dilakukan pengecekan pada missing value pada dataset. Pada gambar dibawah ini terlihat bahwa dataset pada saham Telkomsel tidak memiliki data yang kosong atau null.

```
Informasi Variabel Saham Telkomsel

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1647 entries, 0 to 1646
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Date        1647 non-null   object
1   Open        1646 non-null   float64
2   High        1646 non-null   float64
3   Low         1646 non-null   float64
4   Close       1646 non-null   float64
5   Adj Close   1646 non-null   float64
6   Volume      1646 non-null   float64
dtypes: float64(6), object(1)
memory usage: 90.2+ KB
```

Gambar 4. Pengecekan missing value pada dataset Saham Telkomsel

Kemudian mengecek statistik data serta pengecekan kurva skewness dan outlier. Pada tahap pengecekan kurva skewness digunakan fungsi *fillna* untuk mengisi kolom 'not a number' sehingga sistem akan mengidentifikasi sebagai outlier dan bisa diabaikan.

3.3. Preprocessing

Pada tahapan *preprocessing* tidak semua data akan digunakan dalam melakukan training dan testing. Sehingga pada tahapan ini dialkuakn manipulasi data dalam mendapatkan kategori yang baru. Sebelumnya, pada jual beli saham dikenal sebutan %change atau berapa persen perubahan suatu harga pembukaan hingga harga penutupan pada hari tertentu. %change ini memiliki rumus yaitu harga penutupan (close) – harga pembukaan (open) dibagi dengan harga pembukaan kemudian hasilnya dikalikan dengan 100. Kemudian terdapat istilah lain yaitu spread atau volatility, yaitu jangkauan antara nilai maksimum (High) dengan nilai minimum (Low) kemudian dilakukan pembagian dengan harga penutupan (Close). Data baru ini kemudian dimasukkan ke dalam dataframe yang baru.

```

Program Jurnal
#%%Change
dataset_telkom['%Change'] = (dataset_telkom['Close'] -
dataset_telkom['Open']) / dataset_telkom['Open']
* 100.0

#Spread
dataset_telkom['Spread'] = (dataset_telkom['High'] -
dataset_telkom['Low']) / dataset_telkom['Close']
* 100.0

#Membuat dataframe baru
data_baru = dataset_telkom[['Close', 'Spread',
'%Change', 'Volume']]
data_baru

```

	Close	Spread	%Change	Volume	Output Prediksi
0	2860.0	1.223776	-0.866551	18992100.0	2830.0
1	2835.0	1.058201	-0.526316	49940700.0	2845.0
2	2815.0	1.243339	0.177936	47892100.0	2860.0
3	2810.0	2.669039	-1.919721	70076600.0	2830.0
4	2835.0	0.705467	-0.176056	56582500.0	2805.0
...
1642	3190.0	1.880878	1.592357	55525800.0	NaN
1643	3270.0	3.363914	1.869159	62673400.0	NaN
1644	3190.0	3.134796	-1.846154	103282300.0	NaN
1645	3230.0	2.786378	0.623053	126542400.0	NaN
1646	3240.0	2.777778	1.250000	95143100.0	NaN

1647 rows x 5 columns

Gambar 5. Dataframe data_baru

Setelah dilakukan pembuatan dataframe baru, selanjutnya dataframe tersebut akan dilakukan pembersihan data atau data cleaning dengan melakukan pergantian data pada data 'not a number' menjadi angka -99999 agar dibaca nilai ekstrem atau outlier dan dapat dilakukan pengabaian. Kemudian ditentukan jumlah dari sampel prediksi. Pada percobaan ini penulis menggunakan jumlah sampel prediksi sebanyak 1% dari seluruh data. Data sampel prediksi ini digunakan untuk melakukan prediksi. Penentuan ini digunakan fungsi shifting yaitu cara untuk menggeser atau menghapus 1% data terakhir dari data asli. Pada percobaan ini dilakukan shifting pada data harga Close saham dan akan dilakukan prediksi pada tren harga Close saham nantinya.

```

Program Jurnal
data_baru.fillna(value= -99999, inplace=True)
jml_sampel_prekdiksi = int(math.ceil(0.01*len(
data_baru)))

# menggeser tabel kolom "Close" ke atas sehingga beberapa nilai terakhir menjadi NaN
data_baru['Output Prediksi'] = data_baru['Close'].shift(-jml_sampel_prekdiksi)
data_baru = data_baru.shift(-jml_sampel_prekdiksi)
data_baru

```

```

data_telkom_X = np.array(data_baru[['Close', 'Volume', '%Change', 'Spread']])
data_telkom_X

```

Sehingga didapatkan hasil seperti di bawah ini.

	Close	Spread	%Change	Volume	Output Prediksi
0	2860.0	1.223776	-0.866551	18992100.0	2830.0
1	2835.0	1.058201	-0.526316	49940700.0	2845.0
2	2815.0	1.243339	0.177936	47892100.0	2860.0
3	2810.0	2.669039	-1.919721	70076600.0	2830.0
4	2835.0	0.705467	-0.176056	56582500.0	2805.0
...
1642	3190.0	1.880878	1.592357	55525800.0	NaN
1643	3270.0	3.363914	1.869159	62673400.0	NaN
1644	3190.0	3.134796	-1.846154	103282300.0	NaN
1645	3230.0	2.786378	0.623053	126542400.0	NaN
1646	3240.0	2.777778	1.250000	95143100.0	NaN

1647 rows x 5 columns

Gambar 6. Hasil Shifting

Pada tahapan selanjutnya sudah masuk ke dalam fungsi persamaan linier. Pada persamaan linier sederhana dengan menggunakan variabel output (y) dan variabel input (x). Percobaan ini menggunakan data 'Close', 'Spread', '%Change' dan 'Volume' sebagai variabel inputan sedangkan pada variabel output menggunakan 'Output Prediksi'.

$$y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + c$$

$$output_prediksi = m_1Close + m_2SPV + m_3CHG + m_4Volume + c$$

Gambar 7. Rumus Persamaan Linier Sederhana

Selanjutnya dilakukan normalisasi untuk mempercepat dan mempermudah perhitungan pada machine learning dengan menggunakan library sklearn dengan mengimport library preprocessing. Normalisasi data ini merubah nilai data menjadi nilai dengan rentang -1 sampai 1. Setelah dilakukan preprocessing, akan dibuat dataframe baru untuk menampung nilai x dan y.

```

Program Jurnal
from sklearn import preprocessing
#Normalisasi dengan mengubah nilai menjadi rentang -1 hingga 1
data_telkom_X = preprocessing.scale(data_telkom_X)

data_telkom_X_prediksi = data_telkom_X[-jml_sampel_prekdiksi:]
data_telkom_X = data_telkom_X[:-jml_sampel_prekdiksi] #data 99% elemen data_telkom_X

data_baru.dropna(inplace=True) #menghapus nilai yang 'not a number'/NaN
data_telkom_Y = np.array(data_baru[['Output Prediksi']]) # variabel ini berisi nilai output Prediksi

```

data_telkom_Y

Kemudian menentukan data training dan data testing. Pada percobaan ini menggunakan data training sebesar 80% dan data testing sebesar 20%.

3.4. Penerapan Model Regresi Linear

Pada tahapan penerapan model Regresi Linear sebagai classifier pada proses training data.

Program Jurnal

```
from sklearn.linear_model import LinearRegression
# regresi linier sebagai classifier dengan proses training
linreg = LinearRegression(n_jobs=-1)
linreg.fit(X_train, y_train) # lakukan training
```

Kemudian dilakukan prediksi pada data prediksi untuk 17 hari ke depan dan didapatkan hasil sebagai berikut:

Program Jurnal

```
set_prediksi = linreg.predict(data_telkom_X_prediksi) # prediksi nilai y (output)
set_prediksi
```

```
array([[2998.12132431, 3043.21921378, 3013.78561056, 3199.66403666,
        3230.66205003, 3004.65569667, 3065.04969009, 3094.4151695 ,
        3107.41846675, 3127.29176882, 3139.46477511, 3291.30305596,
        3887.3463654 , 3207.0669848 , 3270.95800731, 3180.05753605,
        3171.84892409])
```

Gambar 8. Hasil Prediksi 13 Hari ke depan

Kemudian dilakukan perbandingan pada harga close saham Telkomsel yang sesungguhnya dimulai dari tanggal terakhir pada dataset yaitu pada tanggal 30 Juli 2021.

Tanggal	Harga Close Saham Asli	Harga Close Saham Prediksi	Tren
1-8-2021	-	2998	-
2-8-2021	3310	3043	Naik
3-8-2021	3330	3013	Turun
4-8-2021	3370	3199	Naik
5-8-2021	3340	3230	Naik

Pada percobaan ini dilakukan 13 hari ke depan akan tetapi, data pada saham hanya ada sampai data tanggal 5 Agustus 2021. Dari sampel tersebut, didapatkan bahwa prediksi tren hampir semuanya benar, akan tetapi pada prediksi harga masih kurang akurat dalam melakukan prediksinya. Karena pada harga saham tidak sepenuhnya linear atau eksponensial.

3.5. Akurasi

Hasil akurasi dilakukan dengan menghitung nilai dari *confidence* pada model Regresi Linear. Dengan menggunakan program dibawah ini.

Program Jurnal

```
confidence = linreg.score(X_test, y_test) # hitung score
print(confidence) # Nilai akurasi model
```

```
[953] confidence = linreg.score(X_test, y_test) # hitung score
print(confidence) # Nilai akurasi model
0.8331479979359973
```

Gambar 9. Nilai Confidence

Hasil *confidence* mendapatkan nilai sebesar 83,31%. Nilai *confidence* sudah cukup tinggi dengan menggunakan model Regresi Linear.

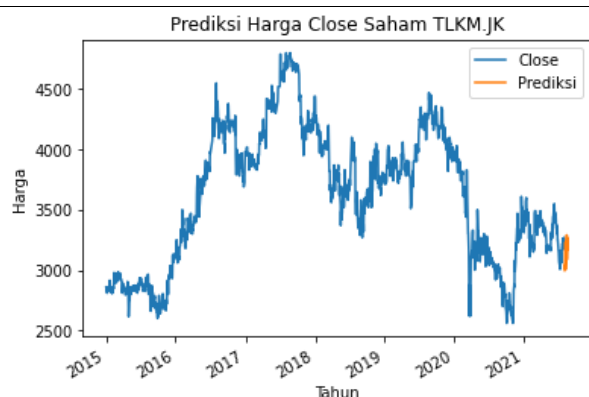
3.6. Visualisasi Grafik Prediksi

Sebelum dilakukan visualisasi pada grafik prediksi, dilakukan manipulasi pada dataframe data_baru dengan menjadikan kolom 'Date' menjadi index. Kemudian diambil tanggal terakhir sebelum data prediksi dan kemudian melakukan penghitungan pada tanggal selanjutnya. Kemudian dilakukan penambahan elemen nilai prediksi yang telah didapatkan ke dalam dataframe data_baru.

Pada penampilan visualisasi grafik prediksi harga Close saham Telkomsel dilakukan dengan fungsi plot(). Plot pada dataframe Close adalah harga Close saham Telkomsel dan kemudian plot pada dataframe 'Prediksi' untuk 13 hari ke depan.

Program Jurnal

```
data_baru['Close'].plot()
data_baru['Prediksi'].plot()
plt.legend()
plt.xlabel('Date')
plt.ylabel('Harga')
plt.title('Prediksi Harga Close Saham TLKM.JK')
plt.show()
```



Gambar 10. Grafik Hasil Prediksi

Kemudian dilakukan perbandingan dengan data yang aslinya.



Gambar 11. Grafik Saham Telkom

Disini terlihat bahwa tren prediksinya masuk ke dalam akurat apabila melihat grafik. Akan tetapi, pada prediksi harga masih terbilang kurang akurat. Dan pada percobaan kali ini, waktu prediksi hingga 17 hari ke depan apabila dihitung dari tanggal 1 Agustus – 17 Agustus 2021. Dikarenakan pada data saham Telkomsel baru sampai tanggal 5 Agustus 2021 sehingga visualisasinya belum bisa terlihat dengan jelas.

4. Kesimpulan

Metode linear regression dapat melakukan prediksi tren harga Close saham dengan nilai *confidence* sebesar 83,31%. Walaupun pada prediksi harga Close saham kurang tepat dikarenakan harga saham tidak linear, akan tetapi pada prediksi tren harga open saham cukup akurat. Saran untuk penelitian selanjutnya adalah dilakukan penelitian pada metode yang lain agar mendapatkan nilai *confidence* atau nilai akurasi yang lebih tinggi dan melakukan *preprocessing* data dengan sangat teliti agar tidak mengalami *overfitting* pada data.

Ucapan Terimakasih

Penulis dengan segenap hati mengucapkan banyak terima kasih kepada banyak pihak yang telah membantu penulis melakukan penelitian ini. Penulis sangat berterima kasih kepada orang tua penulis yang selalu mendukung penulis setiap saat. Penulis juga berterima kasih kepada Ibu Siti Khomsah selaku dosen pengajar Data Mining dan kepada teman-teman yang

telah membantu dan menjadi *support system* dikala waktu yang sulit. Semoga kebaikan yang telah penulis terima dapat bertimbal balik kepada orang tersebut. Aamiin.

Daftar Rujukan

- [1] “Peta Persebaran,” *covid19.go.id*, 2021. [Online]. Available: <https://covid19.go.id/peta-sebaran>. [Accessed: 06-Aug-2021].
- [2] E. Purnaningrum and V. Ariyanti, “Pemanfaatan Google Trends Untuk Mengetahui Intervensi Pandemi Covid-19 Terhadap Pasar Saham Di Indonesia,” *J. Unipasby*, vol. 25, no. 1411, pp. 93–101, 2020.
- [3] A. S. Ahmar and E. Boj, “Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information,” *Sci. Direct*, no. January, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.138883>.
- [4] A. A. Muhartini *et al.*, “Analisis peramalan jumlah penerimaan mahasiswa baru dengan menggunakan metode regresi linear sederhana,” *J. Bayesian*, vol. 1, no. 1, pp. 17–23, 2021.
- [5] D. A. Trianggana, “Peramalan Jumlah Siswa-Siswi Melalui Pendekatan Metode Regresi Linear,” *J. Media Infotama*, vol. 16, no. 2, pp. 115–120, 2020, doi: 10.37676/jmi.v16i2.1149.
- [6] W. Van Der Aalst, *Process Mining Data Science In Action*. Springer Heidelberg New York Dordrecht London, 2016.