

LAPORAN PRAKTIKUM PEMBELAJARAN MESIN

**Metode Distribusi Gaussian dalam Memodelkan *Volatilitas* Harga Saham dan
Memprediksi Pergerakan Harga di Tahun 2018**



**Disusun Oleh :
Kelompok 4**

**Program Studi Sains Data
Jurusan Sains
INSTITUT TEKNOLOGI SUMATERA
Tahun Ajaran 2023/2024**

LAPORAN PRAKTIKUM PEMBELAJARAN MESIN

Metode Distribusi Gaussian dalam Memodelkan *Volatilitas* Harga Saham dan Memprediksi Pergerakan Harga di Tahun 2018

Muhammad Nabil Azizi¹), Rika Ajeng Finatih²), A Rafi Paringgom Iwari³), Anissa Luthfi Alifia⁴), Elsyah Sapyrah⁵).

Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera

Email : muhammad.120450090@student.itera.ac.id¹), rika.121450036@student.itera.ac.id²),
arafi.121450039@student.itera.ac.id³), anissa.121450093@student.itera.ac.id⁴),
elsyah.121450096@student.itera.ac.id⁵)

Abstrak

Metode Distribusi *Gaussian* menjadi pendekatan kunci dalam analisis volatilitas harga saham, memungkinkan para analis keuangan untuk memodelkan fluktuasi harga dengan akurat. Pendekatan ini didasarkan pada asumsi bahwa harga saham mengikuti distribusi normal atau *Gaussian*, yang memungkinkan para analis untuk mengukur risiko dan membuat prediksi mengenai pergerakan harga di masa depan. Distribusi *Gaussian* atau sering disebut juga sebagai distribusi normal merupakan salah satu jenis distribusi dengan variabel acak yang kontinu. Penelitian ini bertujuan untuk memodelkan volatilitas dan memprediksi pergerakan harga saham di masa depan. Referensi utama yang kami gunakan yaitu Modul 1 Praktikum Pembelajaran Mesin. Dengan menggunakan Metode Distribusi *Gaussian*, penelitian ini berusaha menyediakan wawasan yang lebih mendalam dalam analisis volatilitas harga saham untuk mendukung keputusan investasi yang lebih cerdas.

Kata Kunci : *Metode Distribusi Gaussian, Saham*

1. PENDAHULUAN

1.1 Latar Belakang

Saham merupakan surat berharga yang dikeluarkan oleh suatu perusahaan berbentuk perseroan terbatas atau yang biasa disebut emiten, yang menyatakan pemilik saham tersebut dengan demikian apabila seseorang membeli saham maka akan menjadi pemilik perusahaan [1].

Pasar keuangan merupakan lingkungan yang dinamis dan kompleks di mana harga saham dapat mengalami fluktuasi yang signifikan dalam jangka waktu singkat. Volatilitas harga saham merupakan salah satu aspek penting dalam analisis pasar keuangan, karena dapat memberikan informasi tentang risiko dan potensi keuntungan bagi para investor. Semakin besar *volatilitas return* saham suatu perusahaan, maka perusahaan tersebut akan dinilai berisiko tinggi, sehingga dapat menyebabkan *bid ask spread* yang semakin besar pula [2]. Oleh karena itu, penting bagi para analis keuangan untuk dapat memodelkan volatilitas harga saham dengan tepat guna membuat prediksi yang lebih akurat terkait pergerakan harga di masa depan.

Distribusi probabilitas merupakan suatu model peluang yang memungkinkan untuk mempelajari hasil eksperimen *random* yang nyata dan menduga hasil yang akan terjadi. Distribusi peluang tersebut merupakan distribusi populasi karena berhubungan dengan semua nilai yang mungkin terjadi dan populasinya merupakan variabel *random* [3].

Salah satu metode dari distribusi peluang yang dapat digunakan dalam memodelkan volatilitas harga saham adalah metode Distribusi *Gaussian*. Distribusi *Gaussian* atau sering disebut juga sebagai distribusi normal merupakan salah satu jenis distribusi dengan variabel acak yang kontinu. Pada distribusi normal terdapat kurva/grafik yang digambarkan menyerupai bentuk lonceng [3]. Metode ini didasarkan pada asumsi bahwa harga saham mengikuti distribusi normal atau *Gaussian*. Hal ini berarti bahwa fluktuasi harga dianggap sebagai variabel acak yang terdistribusi secara normal.

Dari penelitian yang telah kami lakukan bertujuan untuk dapat memodelkan volatilitas harga saham dan untuk dapat memprediksi pergerakan harga saham di masa depan. Berbagai referensi yang kami gunakan sebagai bahan pembelajaran pada penelitian kali ini yakni, Modul 1 Praktikum Pembelajaran Mesin, Jurnal Sosiohumaniora vol.3 No. 3., Mei 2012, ppt Fungsi Kepadatan Probabilitas, Jurnal Mean Squared Error (MSE) dan Penggunaannya.

2. METODE

2.1 Deskripsi Dataset

Dataset ini berisi data historis saham-saham yang terdaftar di IHSG dengan rentang waktu per menit, per jam, dan harian. Sumber dataset diambil dari data publik Yahoo Finance dan website IDX yang tercantum pada tab metadata [4]. Kumpulan data ini dibuat dengan tujuan untuk tujuan penelitian akademis dan bukan untuk dikomersilkan.

Dataset terdiri dari 14 kolom, yaitu :

1. Code = Kode Stok
2. Name = Nama Perusahaan
3. ListingDate = Tanggal pencatatan saham di Bursa Efek Indonesia
4. Shares= Jumlah saham
5. ListingBoard = Kategori Papan (Papan Utama, Papan Pengembangan, atau Akselerasi).
6. Sector = Kategori Sektor berdasarkan IDX-IC
7. LastPrice = Harga saham terakhir
8. MarketCap = Kapitalisasi Pasar.
9. MinutesFirstAdded = Tanggal data pertama kali diambil dalam rentang menit
10. MinutesLastAdded = Tanggal data terakhir diambil dalam rentang menit
11. HourlyFirstAdded = Tanggal data pertama kali diambil dalam rentang per jam
12. HourlyLastAdded = Tanggal data terakhir diambil dalam rentang per jam
13. DailyFirstAdded = Tanggal data pertama kali diambil dalam rentang harian
14. DailyLastAdded = Tanggal data terakhir diambil dalam rentang harian

2.2 Proses Pembersihan Data

Pada Praktikum kali ini ada beberapa tahap pembersihan data sebelum dapat diolah lebih lanjut.

1. **Pengecekan Kesalahan dan Ketidaksesuaian Data**
Memeriksa apakah ada nilai yang hilang (missing values) dalam setiap kolom dan mengambil tindakan yang sesuai, seperti mengisi nilai yang hilang atau menghapus baris yang memiliki nilai yang hilang.
2. **Penanganan Nilai yang Hilang**
Jika terdapat nilai yang hilang, strategi yang digunakan dapat berupa pengisian nilai yang hilang dengan mean, median, modus, atau menggunakan teknik imputasi data yang lebih canggih.
3. **Penghapusan Duplikat Data**
Mengidentifikasi dan menghapus entri yang merupakan duplikat dari data. Duplikasi data dapat mempengaruhi analisis dan hasil pembelajaran mesin.
4. **Penanganan *Outlier***
Mengidentifikasi dan menangani nilai ekstrim (*outliers*) yang tidak wajar yang dapat mengganggu analisis. Penggunaan metode statistik atau teknik khusus untuk mendeteksi dan menangani *outliers* bisa dilakukan.
5. **Validasi Data**
Memastikan bahwa data pada setiap kolom sesuai dengan format yang diharapkan, seperti jenis data yang benar (misalnya, angka pada kolom numerik, tanggal pada kolom tanggal), dan melakukan perbaikan jika diperlukan.
6. **Transformasi Data**
Melakukan transformasi data seperti normalisasi atau standarisasi pada kolom-kolom tertentu agar memiliki skala yang seragam.
7. **Pemilihan Fitur (*Feature Selection*)**
Menganalisis setiap fitur (kolom) dan memutuskan fitur mana yang akan digunakan dalam analisis atau pembelajaran mesin selanjutnya berdasarkan relevansi dan kepentingannya.
8. **Integrasi Data**
Mengintegrasikan data dari berbagai sumber atau format menjadi satu format yang konsisten sesuai dengan kebutuhan analisis.
9. **Pemantauan dan Evaluasi**
Melakukan pemantauan terus menerus terhadap data setelah proses pembersihan untuk memastikan kualitas dan integritas data.

2.3 Praproses Data

Metode distribusi Gaussian atau yang juga dikenal sebagai model GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) adalah salah satu pendekatan yang umum digunakan untuk memodelkan volatilitas harga saham dan memprediksi pergerakan harga di masa depan. Metode ini sering digunakan dalam analisis keuangan karena volatilitas harga saham cenderung bervariasi dari waktu ke waktu.

Berikut adalah langkah-langkah untuk menerapkan metode distribusi Gaussian menggunakan model GARCH:

1. Mengumpulkan Data Harga Saham

Mulai dengan mengumpulkan data historis harga saham yang akan digunakan untuk analisis.

2. Perhitungan *Return* Harian Saham

Hitung tingkat pengembalian harian atau periode tertentu berdasarkan data harga saham yang telah dikumpulkan. *Return* ini akan digunakan untuk memodelkan volatilitas.

3. Analisis Statistik Dasar

Lakukan analisis statistik dasar pada data return untuk memahami karakteristik dan distribusi datanya. Periksa apakah data return mengikuti distribusi Gaussian atau tidak.

4. Validasi Model

Validasi model gaussian apakah model ini sudah sesuai dengan kebutuhan data dan memiliki performa yang baik dalam memodelkan volatilitas harga

5. Prediksi Volatilitas pada tahun 2018

Menggunakan model yang paling sesuai untuk melakukan prediksi sesuai dengan keefisienan, hal ini berguna untuk membantu dalam menentukan volatilitas harga pada tahun 2018

6. Evaluasi dan Pemantauan

Evaluasi kinerja model dengan membandingkan prediksi volatilitas dan harga saham dengan data aktual. Lakukan pemantauan terus menerus dan lakukan penyesuaian jika diperlukan.

Penting untuk diingat bahwa distribusi Gaussian tidak selalu merepresentasikan sepenuhnya volatilitas harga saham di pasar keuangan yang sebenarnya. Beberapa aset keuangan dapat memiliki volatilitas yang tidak berdistribusi normal (fat-tailed), dan model-model non-parametrik seperti model GARCH dapat memberikan hasil yang lebih akurat dalam kasus ini.

2.4 Rumus atau Formula Metode atau Modeling Data

2.4.1 Distribusi Gaussian

Distribusi Gaussian, juga dikenal sebagai distribusi normal, adalah salah satu distribusi probabilitas yang paling umum digunakan dalam statistik dan ilmu data [3]. Rumus atau formula untuk distribusi Gaussian adalah sebagai berikut:

Fungsi Kepekatan Probabilitas (Probability Density Function, PDF) dari distribusi Gaussian:

$$f(x; \mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

Di mana:

- $f(x)$ adalah nilai kepekatan probabilitas pada titik (x).
- μ adalah nilai rata-rata (mean) dari distribusi.
- σ adalah deviasi standar (standard deviation), yang mengukur sebaran data.

Rumus ini digunakan untuk menghitung nilai kepekatan probabilitas pada titik tertentu (x) dalam distribusi Gaussian dengan parameter rata-rata μ dan deviasi standar σ [5]

Selain PDF, ada juga fungsi distribusi kumulatif (Cumulative Distribution Function, CDF) yang menggambarkan probabilitas bahwa variabel acak Gaussian akan kurang dari atau sama dengan nilai tertentu (x). CDF dari distribusi Gaussian tidak memiliki rumus tertutup yang sederhana, tetapi biasanya dihitung menggunakan tabel atau perangkat lunak komputer.

Dalam konteks pemodelan data, distribusi Gaussian sering digunakan untuk menggambarkan data yang terdistribusi secara normal atau mendekati normal. Ini merupakan salah satu konsep dasar dalam statistik inferensial dan analisis data.

2. 4.2 Return Harian Saham

Return Saham disebut juga sebagai pendapatan saham dan merupakan perubahan nilai harga saham periode t dengan $t-1$. Dan berarti bahwa semakin tinggi perubahan harga saham maka semakin tinggi return saham yang dihasilkan [6].

Kemudian, dalam menentukan return per hari dari saham Tesla (TSLA), diperlukan data harga pembukaan (opening price) dan harga penutupan (closing price) saham Tesla pada setiap hari perdagangan. Return per hari dapat dihitung menggunakan rumus berikut:

$$Return = \frac{Harga\ Penutupan - Harga\ Pembukaan}{Harga\ Pembukaan} \times 100\%$$

Dalam rumus ini:

- "Harga Penutupan" adalah harga saham Tesla pada akhir hari perdagangan.
- "Harga Pembukaan" adalah harga saham Tesla pada awal hari perdagangan.

2. 4.3 Mean Squared Error (MSE)

Mean Squared Error (MSE) adalah metrik evaluasi yang umum digunakan dalam statistik dan machine learning untuk mengukur seberapa akurat sebuah model regresi dalam memprediksi nilai numerik. MSE menghitung selisih antara nilai prediksi model dan nilai sebenarnya dari data, kemudian mengkuadratkan selisih tersebut agar tidak ada selisih yang

bernilai negatif. Kemudian, selisih kuadrat dijumlahkan dan diambil rata-rata dari semua sampel data [7].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

di mana:

- y_i adalah nilai observasi ke- i
- \hat{y}_i adalah nilai prediksi ke- i
- n adalah jumlah observasi

2.5 Variasi Percobaan

Berikut merupakan langkah-langkah variasi percobaan yang dilakukan melalui penelitian ini:

1. Mengimpor dataset harga saham Tesla dan melakukan pembersihan data untuk memastikan tidak ada data yang hilang.
2. Selanjutnya, menghitung "return" harian, yang merupakan persentase perubahan harga penutupan dari hari ke hari.
3. kemudian menghitung rata-rata (mean) dan standar deviasi dari return harian ini, dan menggunakan nilai-nilai ini untuk membuat distribusi Gaussian (normal). Distribusi ini digambarkan dalam plot.
4. Untuk prediksi harga saham di masa depan, menggunakan model random walk. Ini berarti bahwa perubahan harga saham dianggap sebagai proses stokastik (random), dan perubahan harga dari hari ke hari diasumsikan mengikuti distribusi Gaussian yang telah dihitung.
5. Harga saham untuk "besok" diprediksi dengan mengambil harga penutupan terakhir dan menambahkan return yang diprediksi (dihasilkan dari distribusi Gaussian).
6. Akhirnya, mengevaluasi model prediksi dengan menghitung Mean Squared Error (MSE) antara harga saham historis dan harga saham yang diprediksi.

3. HASIL

Pada bab ini dilakukan penyajian informasi terkait data harga saham yang telah dianalisis, serta proses prediksi harga saham dengan menggunakan metode Gaussian.

3.1. Pre-procesing Data

Pada tahap pre-processing ini, dilakukan pemeriksaan terhadap data harga saham Tesla untuk menentukan apakah terdapat data yang hilang (missing) atau duplikat.

```
# Mengecek data yang missing
Data_Saham.isnull().sum()

Date      0
Open      0
High      0
Low       0
Close     0
Volume    0
Ex-Dividend 0
Split_Ratio 0
Adj. Open 0
Adj. High 0
Adj. Low  0
Adj. Close 0
Adj. Volume 0
dtype: int64
```

Gambar 3.1: Proses Cleaning Data

Terlihat pada **Gambar 3.1** hasil yang ditemukan adalah bahwa tidak ada data yang hilang ataupun duplikat, sehingga dilanjutkan ke tahap selanjutnya dengan data yang sudah bersih.

3.2. Return Harian Saham

Pada proses ini dilakukan proses perhitungan terhadap return harian, dimana perhitungan return harian dihitung melalui selisih antara data hari ini dengan data sebelumnya pada kolom close yang kemudian ditampilkan dalam bentuk persentase menggunakan fungsi `pct.change()`.

```
Data_Saham['Return'] = Data_Saham['Close'].pct_change() * 100
print(Data_Saham['Return'])

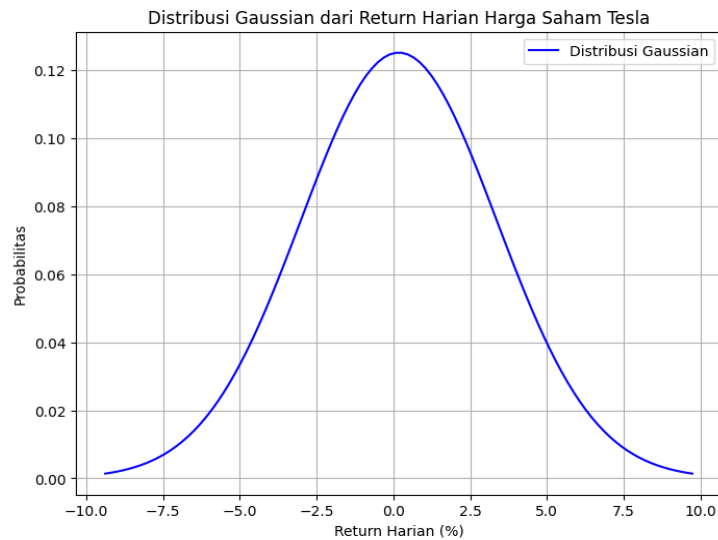
1948      NaN
1947    -0.251151
1946    -7.847251
1945   -12.568306
1944   -16.093750
...
4         1.925616
3        -2.347329
2        -2.445810
1         0.875506
0        -8.218818
Name: Return, Length: 1949, dtype: float64
```

Gambar 3.2: Perhitungan Return Harian

Terlihat dari **Gambar 3.2** bahwa pada lima data teratas dan lima data terbawah nilai return harian lebih sering menunjukkan nilai yang negatif, artinya pada tahun awal dan tahun terkini saham pada perusahaan Tesla lebih cenderung menurun.

3.3. Perhitungan Statistik Distribusi Gaussian Berdasarkan Return Harian Saham

Pada perhitungan statistik return harian digunakan melalui pendekatan distribusi Gaussian untuk melihat persentase return harian yang mungkin didapatkan. Distribusi Gaussian akan menempatkan return harian sebagai variabel x dan persentase sebagai nilai y , kemudian data amatan tersebar dengan rentang dari -10 hingga 10.



Gambar 3.3: Grafik Distribusi Gaussian dari *Return Harian*

Berdasarkan **Gambar 3.3** grafik diatas terlihat bahwa rata-rata return harian berada pada nilai 0.17657%. Artinya nilai return harian yang akan didapatkan per harinya adalah keuntungan sebesar 0.17657% dengan peluang sebesar 0.12%, sehingga saham pada perusahaan Tesla memiliki resiko kerugian yang cukup besar serta peluang keuntungan yang tidak terlalu besar.

3.4. Analisis Prediksi Saham di Masa Depan berdasarkan return harian

```
[ ] # Prediksi harga saham besok

last_close = Data_Saham['Close'].iloc[-1]
predicted_return = np.random.normal(mean_return, std_dev_return)
predicted_price = last_close * (1 + predicted_return / 100)

print(f'Harga saham besok diperkirakan: ${predicted_price:.2f}')

plt.show()

Harga saham besok diperkirakan: $277.59
```

Gambar 3.4: Prediksi Harga Saham Besok

Pada **Gambar 3.4** ini dilakukan prediksi harga saham Tesla untuk hari berikutnya berdasarkan distribusi Gaussian yang telah dihitung sebelumnya. Harga return harian tersebut diperoleh dari harga penutupan (Close) dari dataset harga saham Tesla sebagai nilai awal untuk prediksi. Dengan menggunakan distribusi Gaussian yang telah dihitung sebelumnya (mean_return_ dan std_dev_return), mengambil sampel dari distribusi Gaussian dengan menggunakan np.random.normal. Ini menghasilkan nilai return yang diestimasi berdasarkan distribusi Gaussian. Harga saham besok dihitung dengan mengalikan harga penutupan terakhir dengan faktor pertumbuhan. Faktor pertumbuhan ini dihitung dengan menambahkan return harian yang diestimasi (predicted_return) ke harga penutupan terakhir. Kemudian diperolehlah estimasi harga dengan ouputan menunjukan bahwa harga saham Tesla besok diperkirakan sekitar \$277,59.

3.5. Evaluasi Model

Pada tahap ini dilakukannya evaluasi model dengan menggunakan MSE (Mean Square Error). MSE digunakan untuk mengukur sejauh mana prediksi harga saham berbeda dari harga saham sebenarnya. Semakin rendah nilai MSE, semakin baik model prediksi.

```
[ ] # Menghitung nilai MSE dari prediksi
mse = ((historical_data - predicted_data) ** 2).mean()
print(f'Mean Squared Error (MSE): {mse:.2f}')

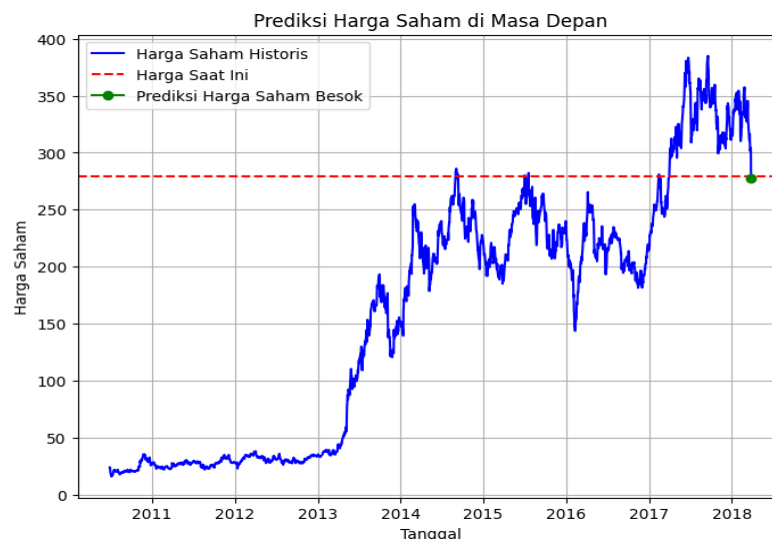
Mean Squared Error (MSE): 37461.67
```

Gambar 3.5: MSE dari Prediksi

Pada **Gambar 3.5** hasil Evaluasi model menunjukkan Mean Square Error (MSE) sebesar 37,461.67. Hasil yang ditampilkan bahwa MSE dari model relatif tinggi, hal ini mengidentifikasikan bahwa prediksi harga saham yang dihasilkan oleh model memiliki deviasi yang signifikan dari harga saham yang sebenarnya. Hal ini menunjukkan bahwa harga model prediksi sederhana yang didasarkan pada distribusi Gaussian mungkin tidak cukup akurat untuk memprediksi pergerakan harga saham Tesla.

3.6. Visualisasi Hasil

Melalui pemodelan yang telah dilakukan didapatkan hasil akhir berupa plot dari volatilitas dari saham perusahaan Tesla. Plot yang dihasilkan memiliki variabel x berupa tahun dan variabel y adalah harga satuan, dimana rentang untuk variabel x adalah dari awal tahun 2010 hingga pertengahan tahun 2018 dan rentang variabel y antara 0 hingga 400.



Gambar 3.6: Visualisasi Harga Saham

Pada **Gambar 3.6** didapatkan informasi bahwa pola pada saham harga satuan Tesla dari tahun 2010 hingga 2019 berpola trend atau berpola dengan kecenderungan naik, meski sempat beberapa kali tidak konsisten dengan penurunan yang cukup signifikan. Kemudian dirapatakan

pula hasil dari pemodelan program dalam memprediksi harga satuan saham Tesla pada besok hari adalah pada harga 277.59\$. Harga tersebut tentu bernilai buruk jika dibandingkan dengan harga saham historis pada awal tahun tersebut, harga tersebut terbilang sangat buruk pada tahun tersebut jika melihat kenaikan dari harga historis pada tahun sebelumnya yang memiliki kecenderungan naik.

4. KESIMPULAN

Berdasarkan hasil pemodelan harga saham pada perusahaan Tesla dari tahun 1947 hingga 2018 dengan menggunakan pendekatan distribusi Gaussian didapatkan rata-rata return harian yang mungkin didapatkan pada saham perusahaan Tesla adalah sebesar 0.17657 dengan peluangnya yaitu 0.12%, sehingga disimpulkan bahwa keuntungan yang mungkin didapatkan cukup kecil disertai dengan resiko yang cukup besar.

Kemudian jika melihat pola tren saham pada perusahaan Tesla yang cenderung naik, meski terdapat lonjakan penurunan yang harga yang cukup signifikan. Namun setelah dilakukan analisis pada data harga terakhir adalah harga terburuk untuk harga pada tahun 2018. Kemudian, didapatkan pula hasil prediksi untuk harga pada besok hari adalah tidak jauh pada harga sebelumnya yaitu dengan nilai prediksi adalah 277.42\$. Sehingga, dengan mengetahui informasi ini, dapat dilakukan pengambilan keputusan kedepannya terhadap tindakan yang akan dilakukan untuk meminimalkan kerugian atau dapat memperkirakan kemungkinan harga jika menunggu beberapa hari atau beberapa bulan kedepan.

REFERENSI

- [1] M. BUDIANTARA, "PENGARUH TINGKAT SUKU BUNGA, NILAI KURS, DAN INFLASI TERHADAP INDEKS," *Jurnal Sosiohumaniora* vol.3 No. 3., Mei 2012, p. 59, 2012.
- [2] A. A. G. S. Ni Made Wahyuliantini, "PENGARUH HARGA SAHAM, VOLUME PERDAGANGAN SAHAM,," *Jurnal Manajemen, Strategi Bisnis dan Kewirausahaan*, no. 9, p. 148, 2015.
- [3] P. S. S. D. ITERA, "Modul Praktikum Pembelajaran Mesin," Bandar Lampung, 2023, p. 1.
- [4] M. KHADAFI, "kaggle.com/datasets/muamkh/ihsgstockdata," 09 01 2023. [Online]. Available: <https://www.kaggle.com/datasets/muamkh/ihsgstockdata>. [Diakses 21 09 2023].
- [5] A. Basuki, "Fungsi Kepadatan Probabilitas," pp. 7-8, 2004.
- [6] A. Halim, dalam *Analisis Investasi*, Jakarta: Salemba, 2005.

[7] H. H. Nuha, “Mean Squared Error (MSE) dan Penggunaannya,” vol. 2, p. 1, 2023.

LAMPIRAN

Berikut merupakan lampiran dari hasil kode program beserta dengan dataset yang digunakan:

<https://drive.google.com/drive/folders/1WZ44WVGIFdtBfqOdqE33YCW1kT48Csh>