



KELAS·WORK
by KELAS.COM

FINAL PROJECT DATA SCIENCE

Suci Rosyidatunnajah

ABOUT ME

Perkenalkan, nama saya Suci Rosyidatunnajah. Saya lulusan S1 Meteorologi dari Institut Teknologi Bandung yang baru saja lulus tahun 2024 ini. Selama kuliah saya sudah sering bekerja berhubungan dengan data, namun hanya seputar data meteorologi. Oleh karena itu saya mengikuti bootcamp data science ini untuk mempelajari lebih dalam tentang ilmu data science agar bisa diimplementasikan di dunia profesional nantinya.

Pada project ini, saya mengimplementasikan ilmu yang sudah saya dapat seperti membuat DataFrame, melakukan EDA hingga menghasilkan sebuah visualisasi data, sampai mengimplementasikan seluruh tahapan model machine learning dan deep learning. Project ini terdiri dari 4 case, yaitu sales force training, housing price, flight price prediction, dan Apple stock price prediction.

No	Kompetensi
1	Mampu Membuat Data Frame Berupa Baris dan Kolom
2	Mampu Melakukan Analisis Data Menggunakan Measures of Central Tendency dan Measures of Variability
3	Mampu Melakukan Hingga Menganalisis Hasil Uji T-Test pada Python
4	Mampu Menerapkan Proses Exploratory Data Analysis (EDA)
5	Mampu Menerjemahkan Data Menjadi Visualisasi Data
6	Mampu Mengimplementasikan Machine Learning Model Menggunakan Algoritma Supervised atau Unsupervised Learning
7	Mampu Mengimplementasikan Deep Learning Model Menggunakan Algoritma Artificial Neural Networks (ANN)

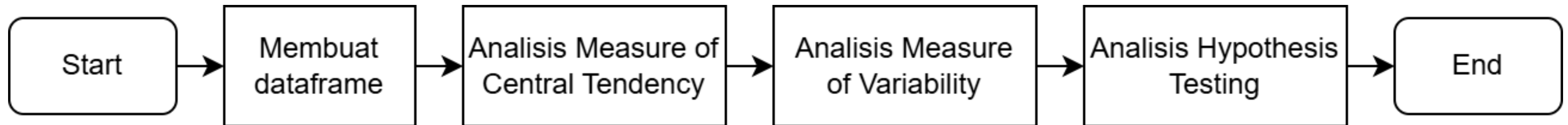
01: SALES FORCE TRAINING

Background Problem Project

- Perusahaan X ingin meningkatkan penjualan mereka dengan melakukan training kepada pekerja sales. Sebelumnya penjualan rata-rata perusahaan X adalah \$100 per transaksi
- Dengan data penjualan terbaru dari 25 sampel pekerja sales, akan dievaluasi apakah training kepada pekerja sales efektif dalam meningkatkan jumlah penjualan perusahaan.
- Langkah yang dilakukan untuk melihat efektivitas training ini dengan menganalisis data penjualan terbaru dengan Measure of Central Tendency, Measure of Variability, dan Hypothesis Testing.

01: SALES FORCE TRAINING

Flowchart



01: SALES FORCE TRAINING

Analisis measure of central tendency dan variability

```
mean: 102.6
median: 100
variansi: 972.2399999999997
srandar deviasi: 31.180763300471007
max: 150
min: 50
range: 100
quantile 0.25: 85.0
quantile 0.5: 100.0
quantile 0.75: 120.0
```

- Rata-rata penjualan terbaru hampir mirip dengan rata-rata penjualan sebelumnya.
- Range penjualan sebesar \$100 menunjukkan adanya variasi besar antara penjualan tertinggi dan terendah. Hal ini didukung dari nilai standar deviasi (31,18) yang menunjukkan adanya variasi data yang cukup besar di sekitar rata-rata

01: SALES FORCE TRAINING

Analisis hypothesis testing

hypothesis testing dilakukan dengan menggunakan T-test dengan nilai alpha 5%

Dengan H0 dan H1:

- H0: Penjualan rata-rata = \$100 per transaksi setelah dilakukan training kepada pekerja sales
- H1: Penjualan rata-rata > \$100 per transaksi setelah dilakukan training kepada pekerja sales

Didapatkan hasil berikut:

t-statistic: 0.4085001556802841

P-value: 0.6865284813438117

critical value: 1.7108820799094275

- Nilai pvalue (0.68652848) lebih besar dari nilai alpha (0.05), hal ini berarti tidak cukup bukti untuk menolak H0.
- Nilai t-statistic (0.41) lebih kecil dibanding nilai kritis (1.71), hal ini mendukung kesimpulan sebelumnya untuk tidak menolak H0.
- Hal ini menunjukkan training kepada pekerja sales kurang efektif karena rata rata penjualan masih sebesar \$100 sama seperti sebelumnya

01: SALES FORCE TRAINING

Kesimpulan

Training kepada pekerja sales kurang efektif dalam meningkatkan penjualan Perusahaan X menjadi di atas \$100 per transaksi.

[Link github](#)

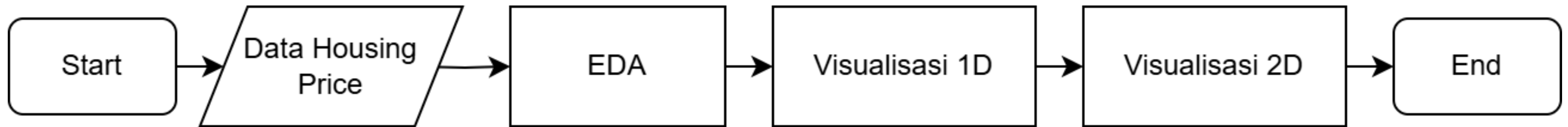
02: HOUSING PRICE

Background Problem Project

- Housing Price merupakan data dari properti yang sudah terjual yang berisi data harga hingga karakteristik properti seperti fasilitas, luas bangunan, dan sebagainya.
- Pada project ini akan dilakukan visualisasi data terhadap data Housing Price untuk menganalisis karakteristik properti yang sudah terjual serta menganalisis harga properti dan faktor yang memengaruhinya.
- Analisis dilakukan dengan menggunakan teknik visualisasi 1D dan 2D.

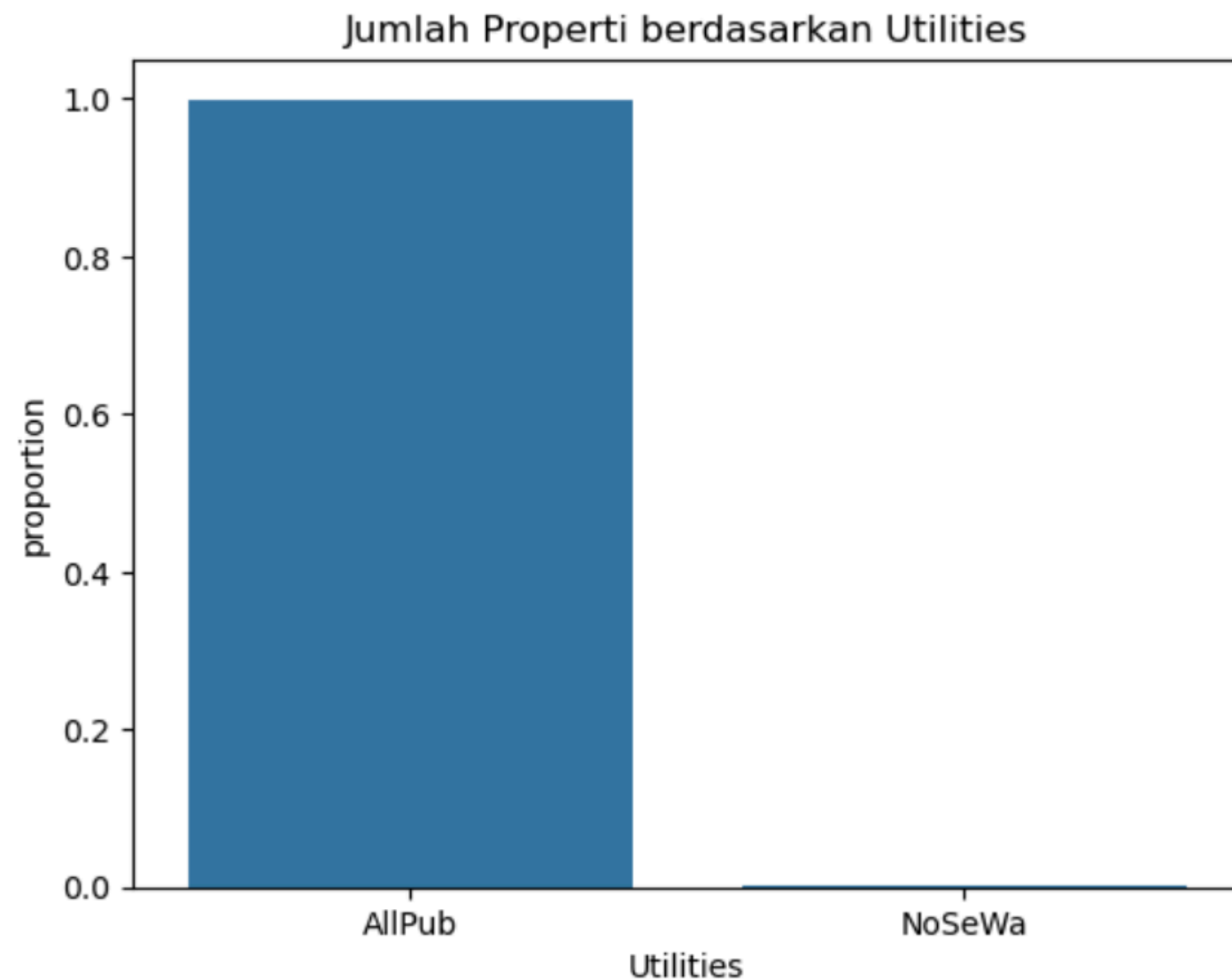
02: HOUSING PRICE

Flowchart



02: HOUSING PRICE

Visualisasi 1D

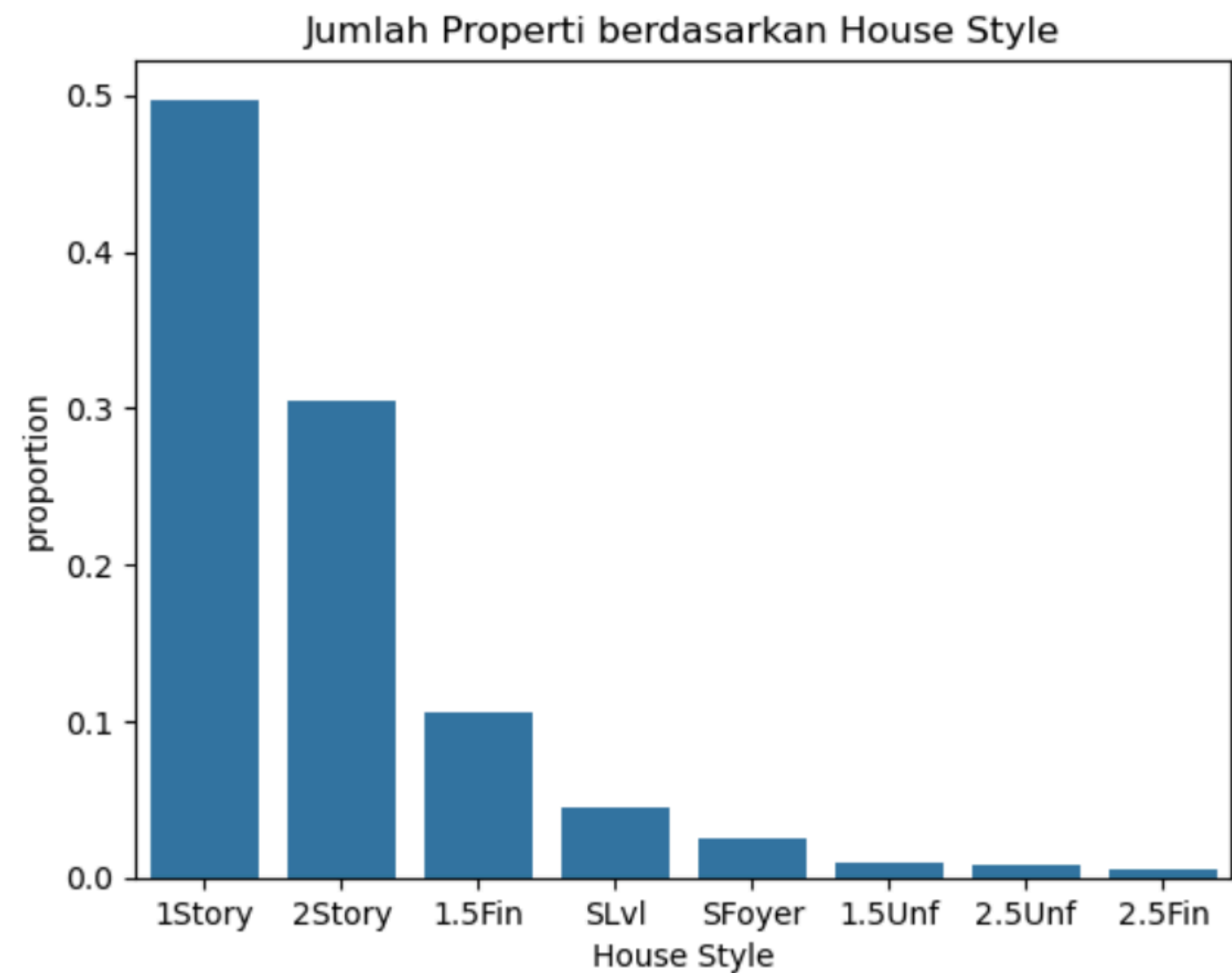


- Gambar di samping menunjukkan jumlah properti yang terjual berdasarkan fasilitasnya, yaitu AllPub (memiliki semua fasilitas) dan NoSeWa (hanya memiliki fasilitas listrik dan gas)
- Terlihat jika hampir 100% properti yang terjual memiliki semua fasilitas umum.

[Link github](#)

02: HOUSING PRICE

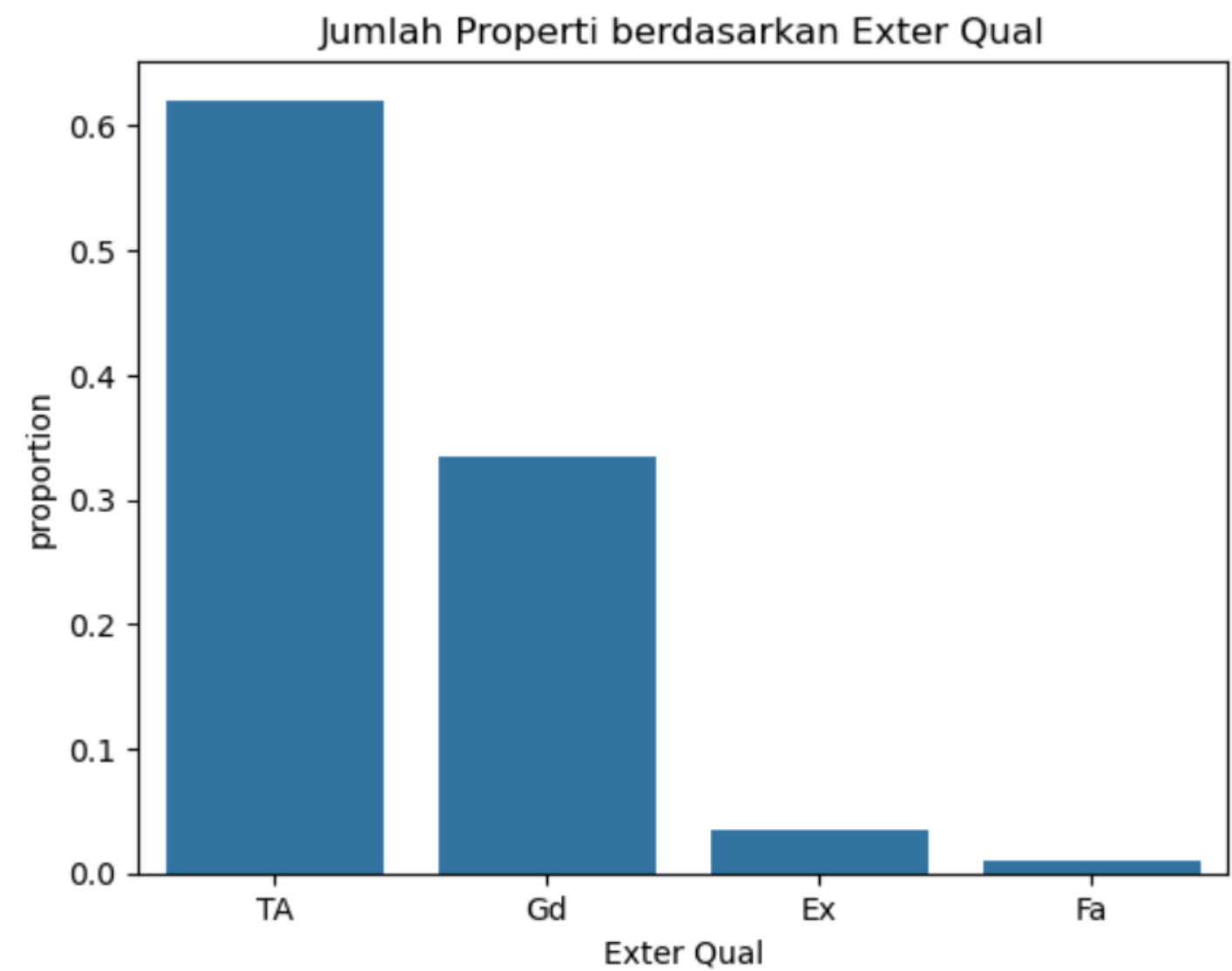
Visualisasi 1D



- Gambar di samping menunjukkan jumlah properti yang terjual berdasarkan gaya rumahnya.
- Hampir 50% properti yang terjual hanya memiliki 1 lantai lalu sekitar 30% properti yang terjual merupakan rumah 2 lantai.

02: HOUSING PRICE

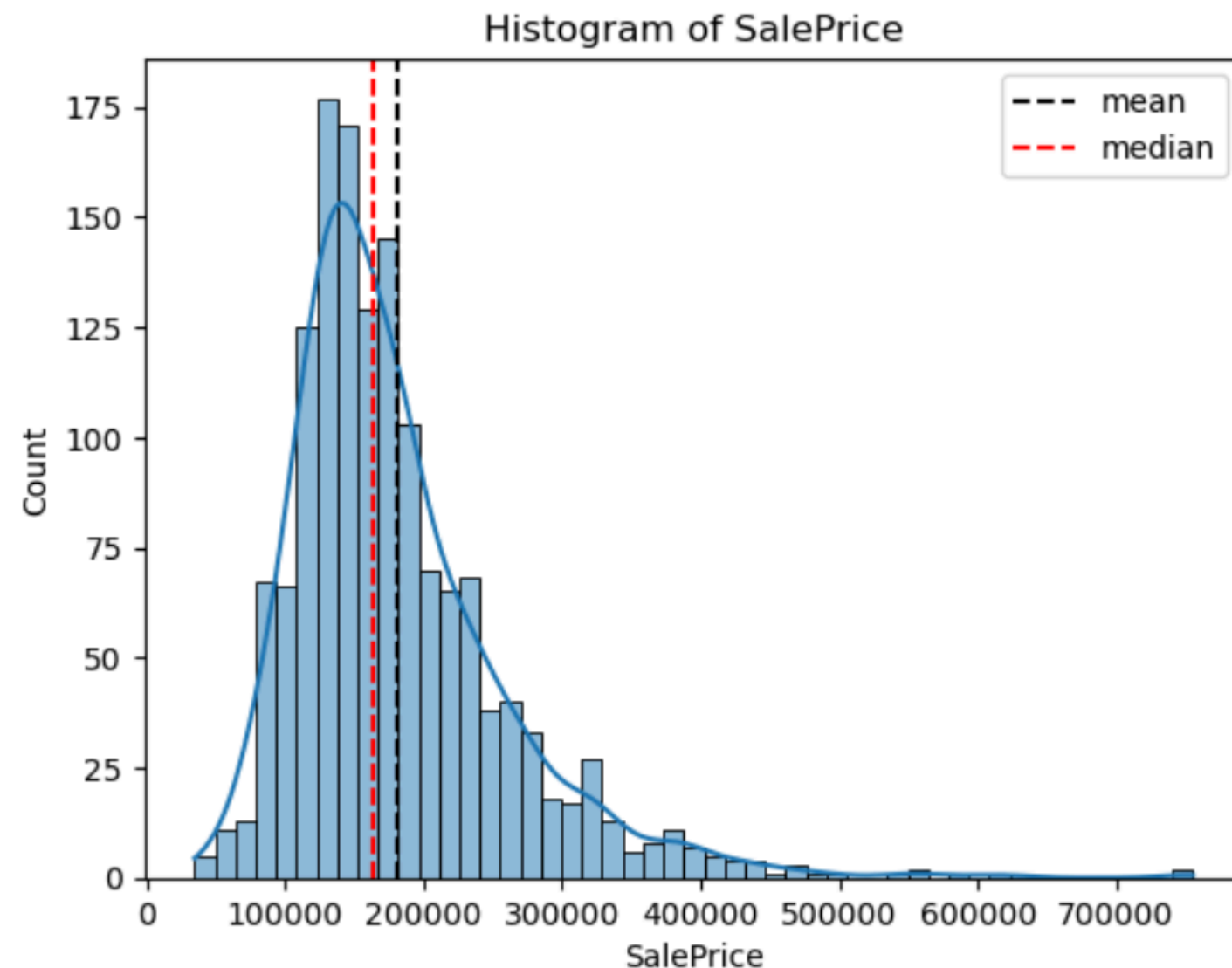
Visualisasi 1D



- Gambar di samping menunjukkan jumlah properti yang terjual berdasarkan kualitas material eksteriornya, yaitu TA (average/Typical), Gd (good), Ex (excellent), dan Fa (fair).
- Lebih dari 60% properti yang terjual memiliki material eksterior dengan kualitas rata-rata dan lebih dari 30% dengan kualitas bagus

02: HOUSING PRICE

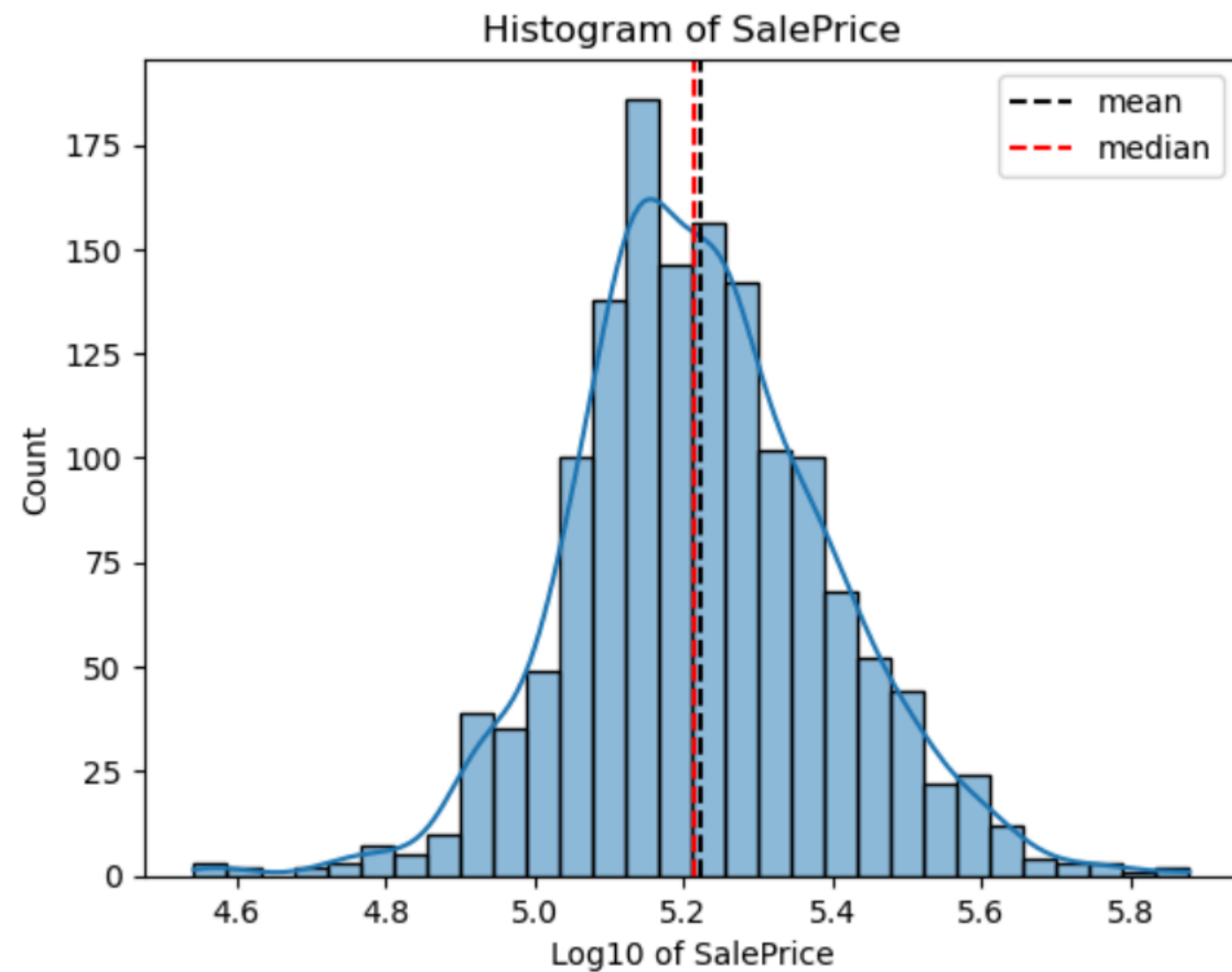
Visualisasi 1D



- Gambar di samping menunjukkan distribusi dari harga properti.
- Terlihat data terdistribusi right-skewed yang menunjukkan sebagian besar rumah yang terjual memiliki harga di rentang 100.000–250.000 lalu sebagian kecil rumah memiliki harga yang jauh lebih tinggi bahkan mencapai 700.000.
- Dari segi pasar hal ini menunjukkan rumah dengan harga lebih murah (100.000 sampai 250.000) sangat laku sedangkan rumah dengan harga jauh lebih mahal juga memiliki peminat namun cukup jarang..

02: HOUSING PRICE

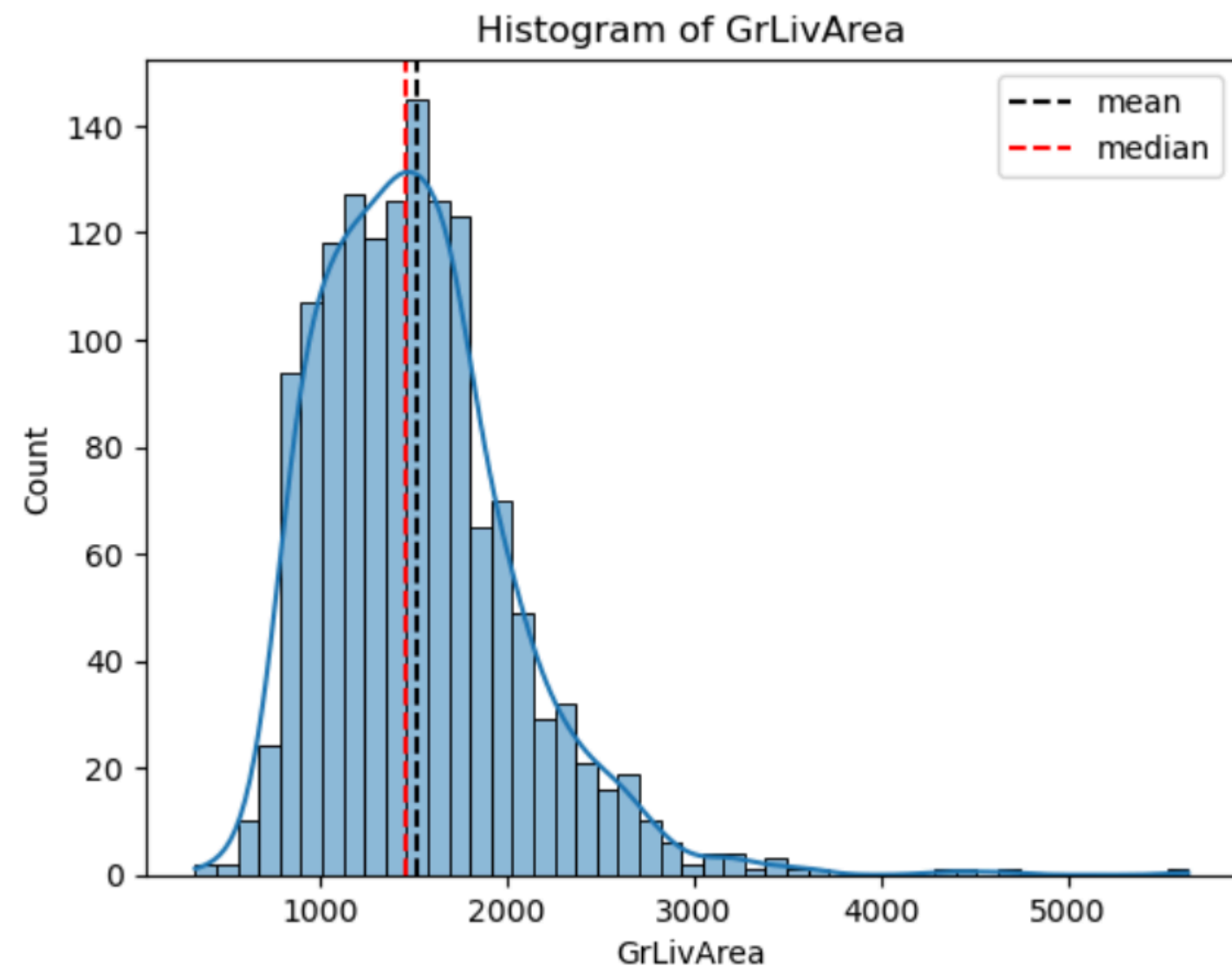
Visualisasi 1D



- Gambar di samping menunjukkan distribusi dari harga properti yang sudah dilakukan transformasi logaritmik.
- Distribusi data terlihat lebih simetris atau mendekati distribusi normal.

02: HOUSING PRICE

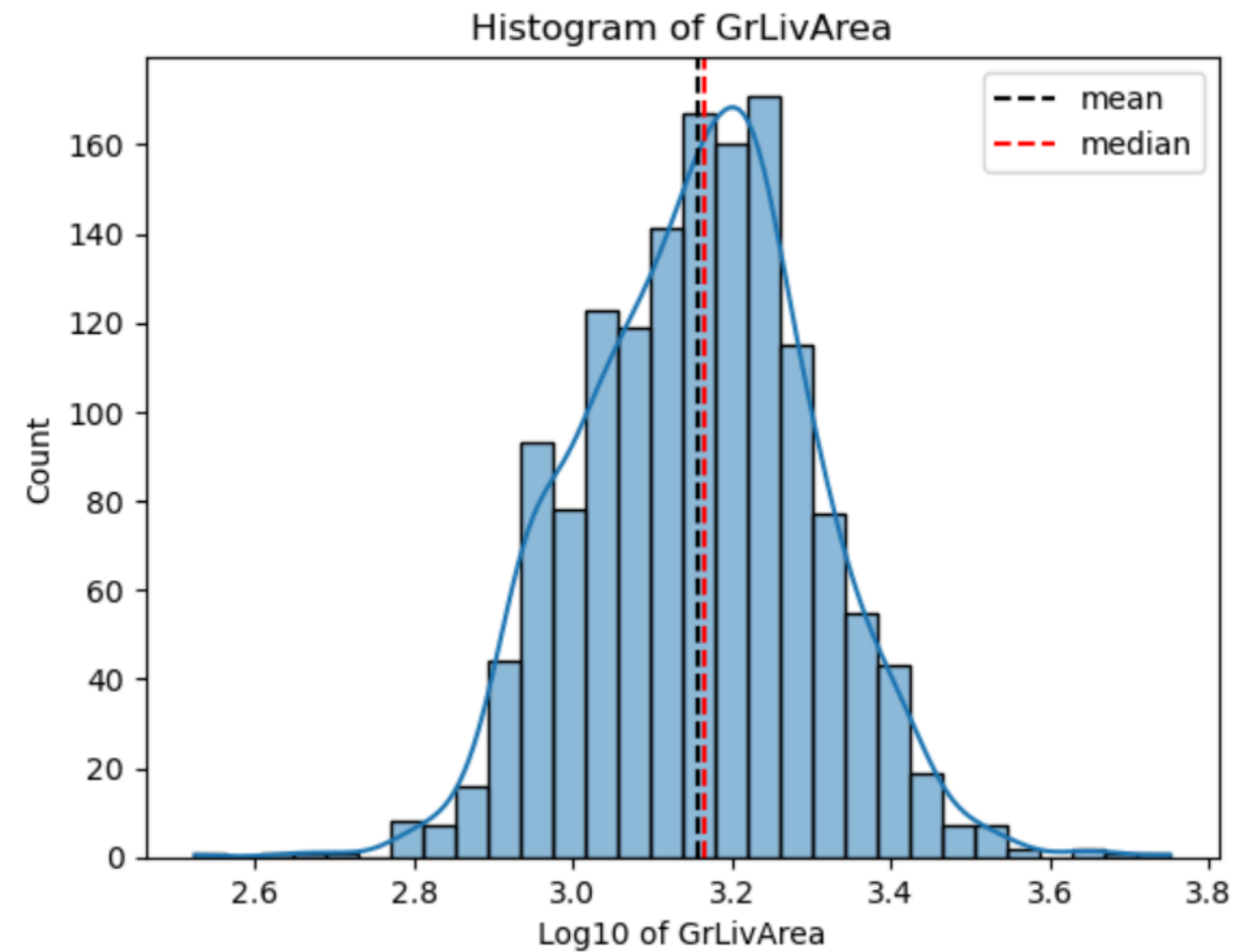
Visualisasi 1D



- Gambar di samping menunjukkan distribusi dari Luas bangunan properti.
- Terlihat data terdistribusi right-skewed yang menunjukkan sebagian besar rumah yang terjual memiliki luas di rentang 1000-2000 lalu sebagian kecil rumah yang terjual memiliki luas yang jauh lebih besar bahkan mencapai 5000.

02: HOUSING PRICE

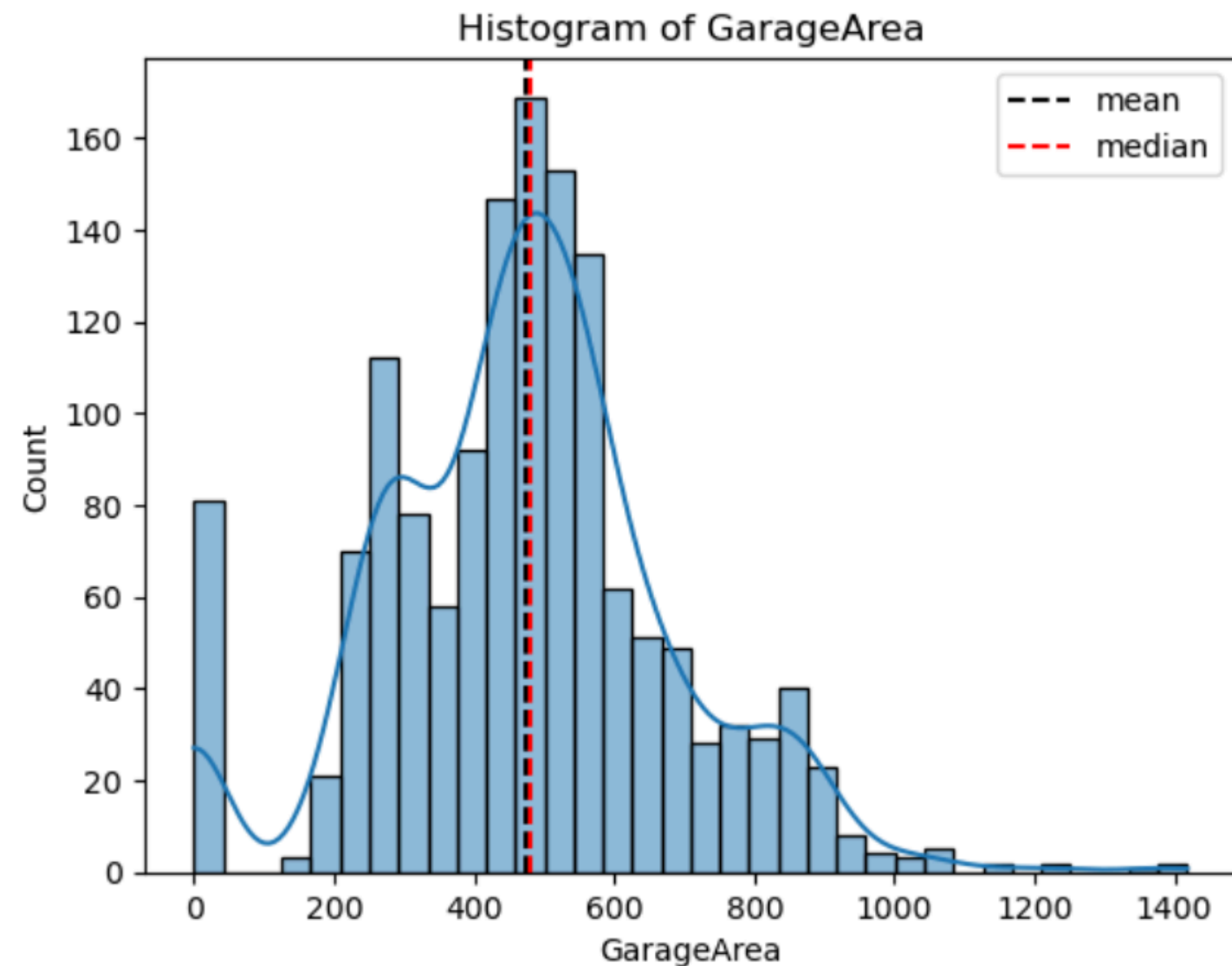
Visualisasi 1D



- Gambar di samping menunjukkan distribusi dari luas bangunan properti yang sudah dilakukan transformasi logaritmik.
- Distribusi data terlihat lebih simetris atau mendekati distribusi normal.

02: HOUSING PRICE

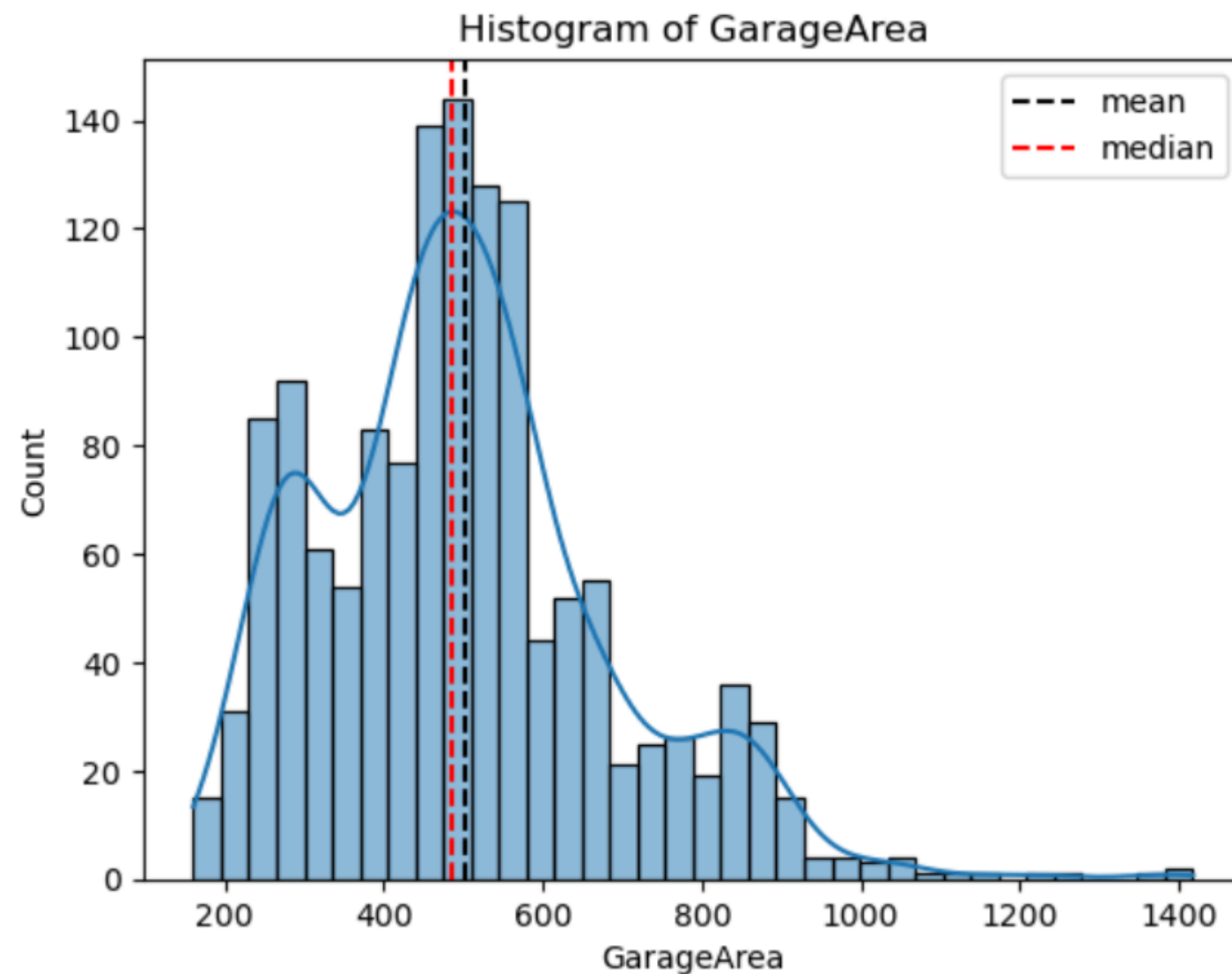
Visualisasi 1D



- Gambar di samping menunjukkan distribusi data dari Luas garasi properti.
- Terlihat data juga tidak terdistribusi normal, namun disini terdapat hal menarik karena sebagian data bernilai 0 yang artinya tidak memiliki garasi. dari grafik dapat dilihat jika sekitar 80 rumah tidak memiliki area garasi.

02: HOUSING PRICE

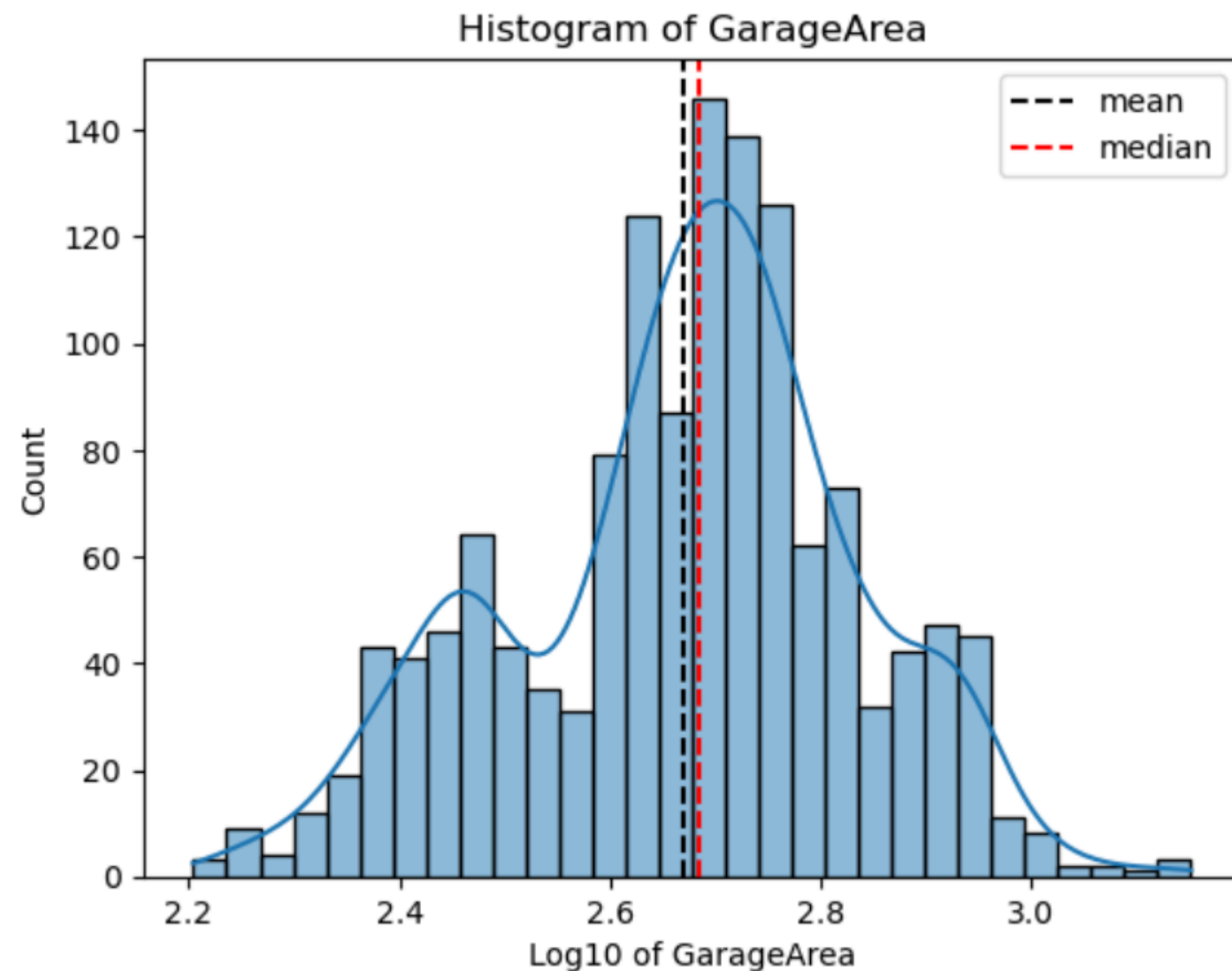
Visualisasi 1D



- Gambar di samping menunjukkan distribusi data dari Luas garasi pada properti yang memiliki garasi.
- Terlihat data terdistribusi right-skewed yang menunjukkan sebagian besar rumah yang terjual memiliki garasi seluas 300-600 lalu sebagian kecil rumah yang terjual memiliki garasi dengan luas yang jauh lebih besar bahkan mencapai 5000.

02: HOUSING PRICE

Visualisasi 1D



- Gambar di samping menunjukkan distribusi dari data luas garasi yang sudah dilakukan transformasi logaritmik.
- Distribusi data terlihat lebih simetris atau mendekati distribusi normal.

02: HOUSING PRICE

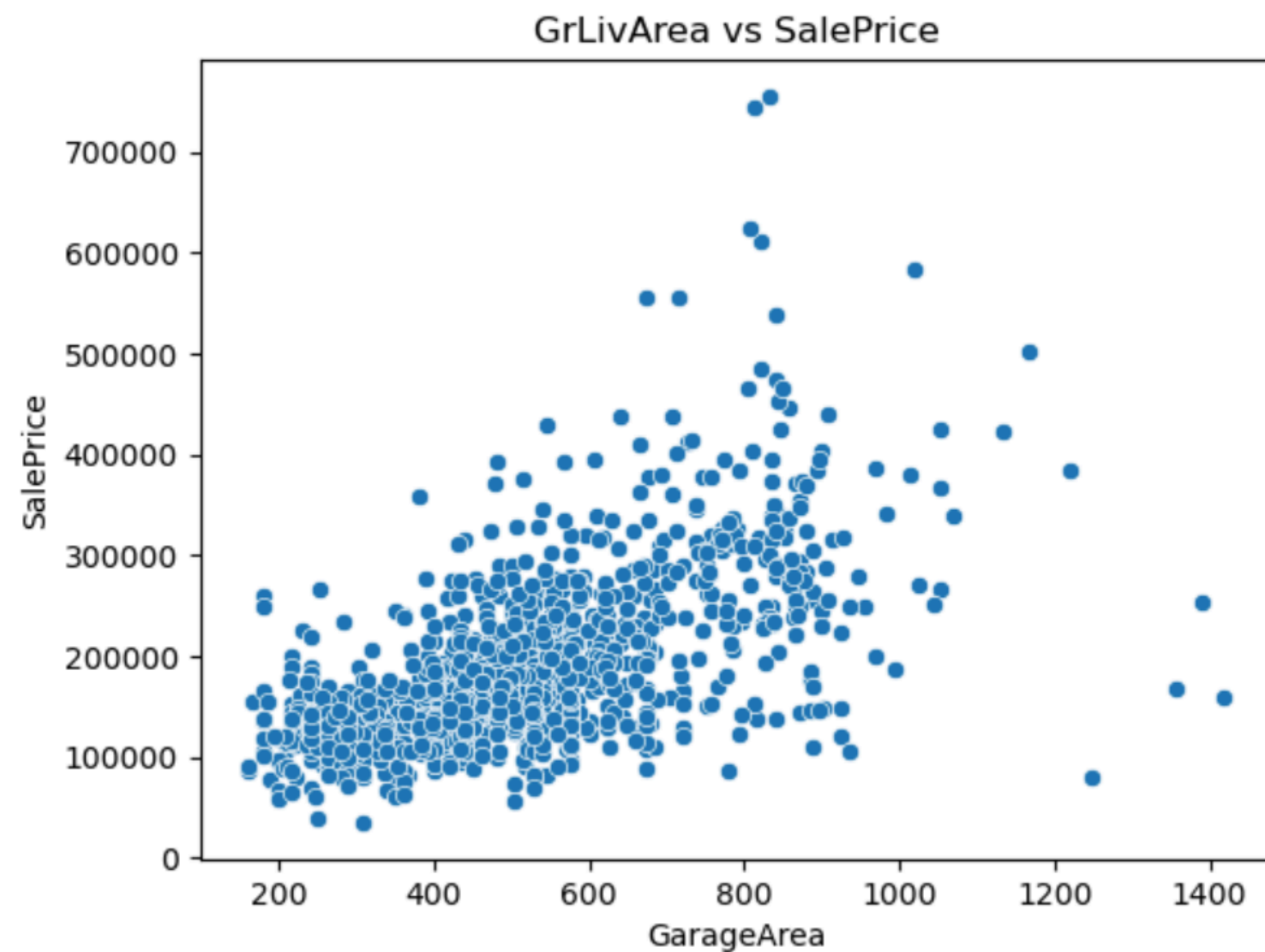
Visualisasi 2D



- Gambar di samping menunjukkan scatter plot hubungan antara harga properti dengan luas bangunan properti.
- Terlihat ada pola korelasi positif antara luas bangunan dan harga rumah. Artinya, semakin luas suatu bangunan maka semakin tinggi harganya.
- Namun ada beberapa outlier di sebelah kanan grafik yang menunjukkan bangunan dengan luas yang besar namun harganya tidak terlalu tinggi.

02: HOUSING PRICE

Visualisasi 2D



- Gambar di samping menunjukkan scatter plot hubungan antara harga properti dengan luas garasi properti.
- Terlihat ada pola korelasi positif juga antara luas garasi dan harga rumah. Sama seperti sebelumnya, hal ini menunjukkan semakin besar luas garasi maka harga rumahnya akan semakin mahal.
- Namun juga terlihat adanya outlier di sebelah kanan dan atas grafik yang menunjukkan rumah dengan garasi yang sangat luas namun harganya tidak terlalu mahal dan ada pula rumah dengan garasi tidak terlalu luas namun harganya sangat tinggi.

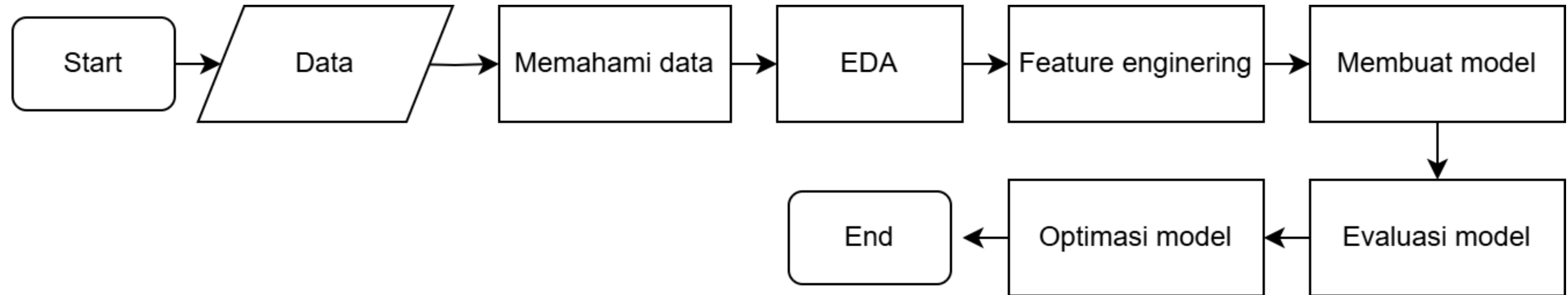
03: FLIGHT PRICE PREDICTION

Background Problem Project

- Pesawat terbang merupakan salah satu transportasi umum yang sering digunakan oleh masyarakat. Harga dari tiket pesawat sangat bervariasi sesuai dengan fasilitas pesawat yang ditawarkan.
- Pada project ini akan dibuat model machine learning untuk memprediksi harga dari tiket pesawat.
- Data yang digunakan merupakan data pemesanan tiket pesawat terbang di India yang terdiri dari harga tiket beserta rinciannya, seperti maskapai pesawat, kelas pesawat, lama durasi perjalanan, lokasi keberangkatan, dan sebagainya.
- Model machine learning akan dibuat dengan menggunakan metode KNN, decision tree, dan random forest. Setelah itu model akan dievaluasi untuk menentukan metode terbaik.

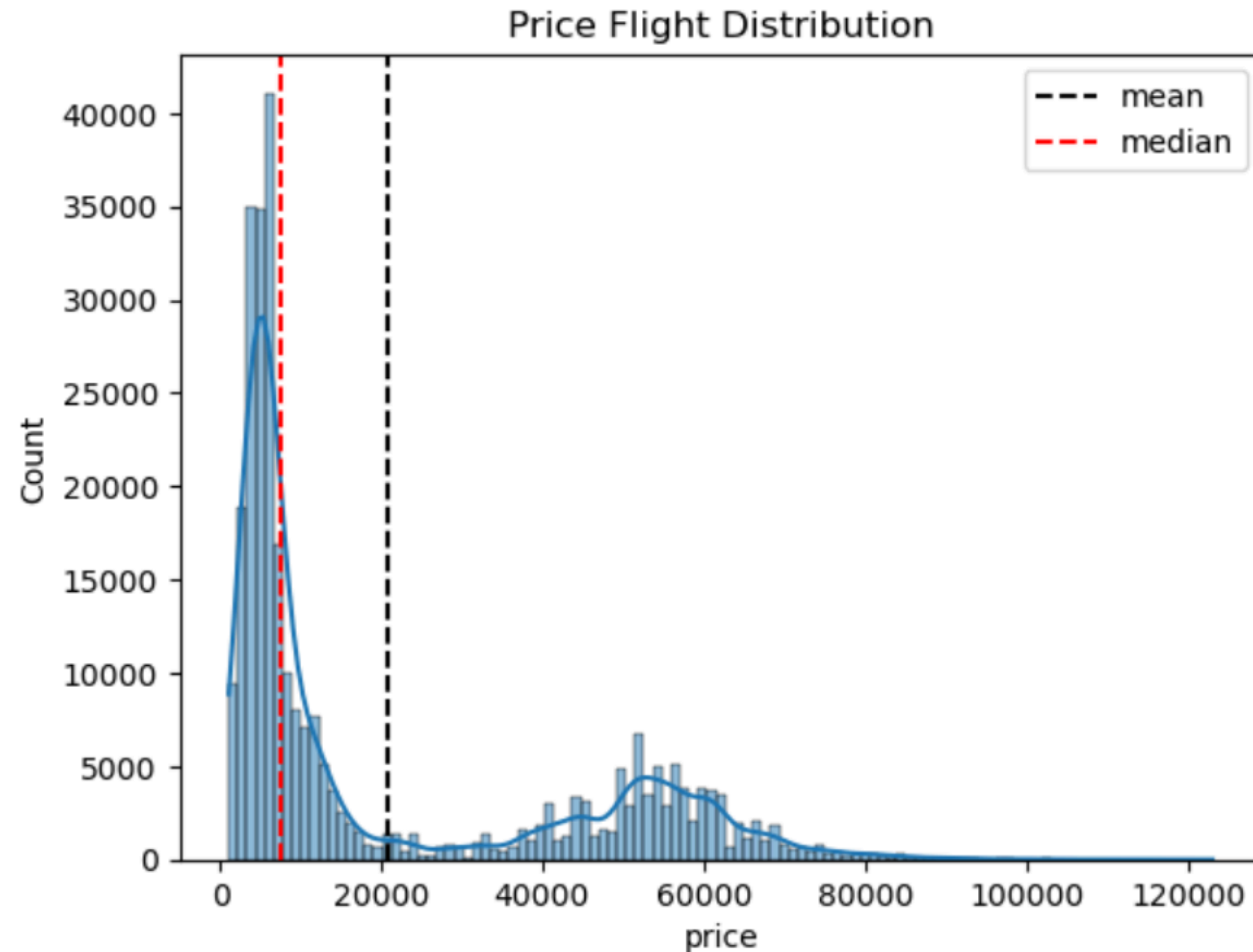
03: FLIGHT PRICE PREDICTION

Flowchart



03: FLIGHT PRICE PREDICTION

EDA



- Gambar di samping menunjukkan distribusi dari harga tiket pesawat.
- Terlihat jika distribusi harga tiket sangat miring ke kanan (right-skewed) yang menunjukkan mayoritas harga tiket berkisar di nilai yang lebih rendah, dan beberapa tiket memiliki harga yang jauh lebih tinggi. Hal ini bisa dipahami karena mayoritas masyarakat pasti akan memilih tiket dengan harga lebih terjangkau.

[Link github](#)

03: FLIGHT PRICE PREDICTION

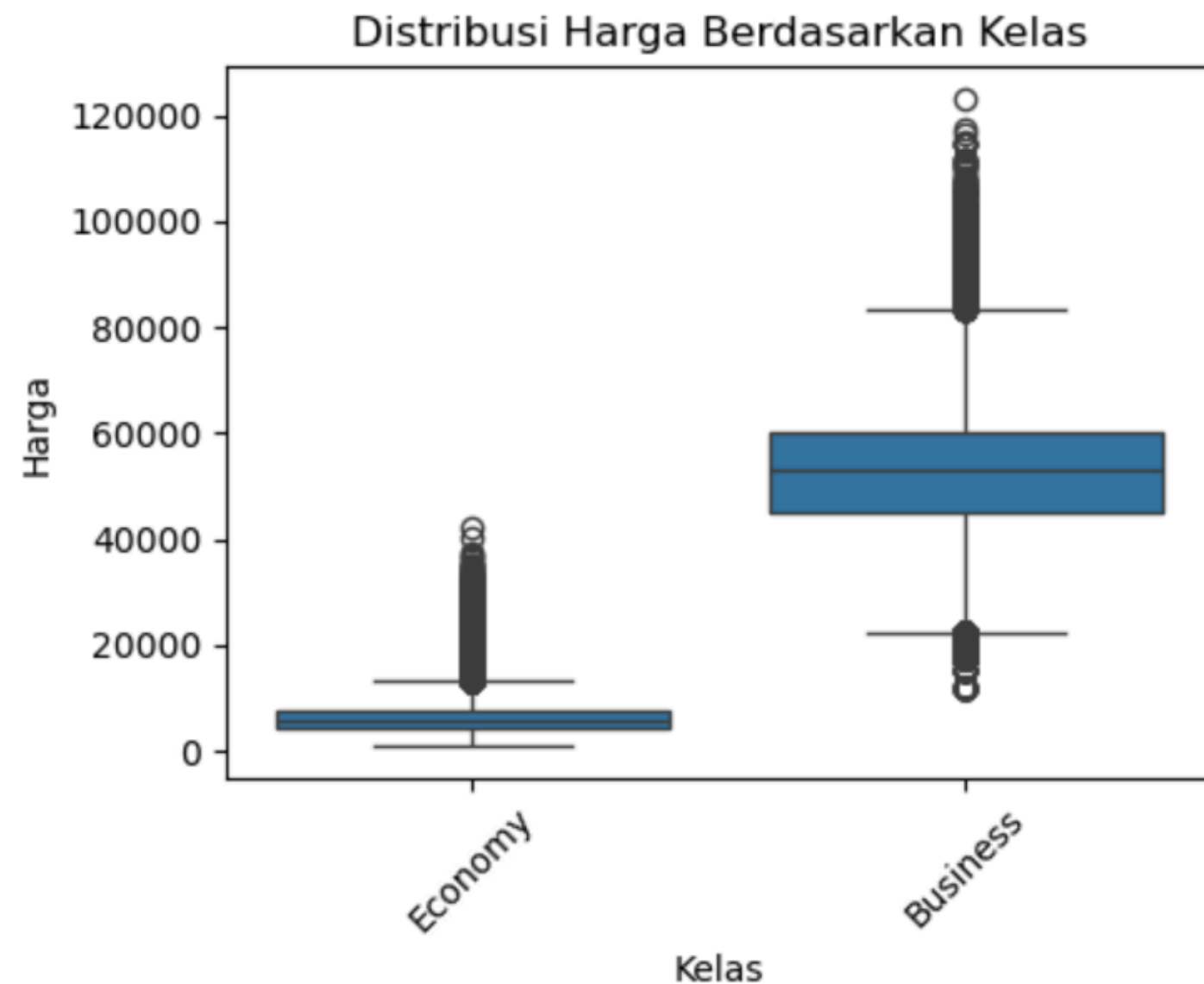
EDA



- Gambar di samping menunjukkan correlation heatmap dari rincian pemesanan tiket pesawat.
- Terlihat kategori yang memengaruhi harga pesawat adalah kelas pesawat karena memiliki korelasi yang kuat sebesar -0.94.
- Selain kelas pesawat tidak ada yang memengaruhi harga pesawat dengan signifikan karena korelasinya di bawah 0.5 yang menunjukkan korelasinya lemah.

03: FLIGHT PRICE PREDICTION

EDA



- Gambar di samping menunjukkan distribusi dari harga tiket pesawat berdasarkan kelas pesawatnya.
- Terlihat adanya perbedaan rentang harga tiket antara pesawat kelas ekonomi dan pesawat kelas bisnis.

03: FLIGHT PRICE PREDICTION

Model machine learning

Model machine learning dibuat dengan menggunakan metode KNN, decision tree, dan random forest. Setelah itu model dilakukan evaluasi untuk menentukan model terbaik. Berikut hasil evaluasinya:

K-Nearest Neighbors Performance:

MAE: 0.08

MSE: 0.03

R^2 : 0.97

Decision Tree Performance:

MAE: 0.05

MSE: 0.02

R^2 : 0.98

Random Forest Performance:

MAE: 0.05

MSE: 0.01

R^2 : 0.99

- Hasil evaluasi terbaik ditunjukkan dengan nilai MAE dan MSE mendekati 0 dan nilai R^2 mendekati 1.
- Terlihat model Random Forest memiliki hasil evaluasi terbaik sehingga model inilah yang dipilih.

03: FLIGHT PRICE PREDICTION

Model machine learning



- Gambar di samping menunjukkan scatterplot dari residual (selisih dari harga tiket aktual dan harga tiket prediksi) vs harga tiket prediksi.
- Terlihat residual tersebar acak di sekitar garis nol. Hal ini menunjukkan model yang dibuat cocok dengan data.

04: APPLE STOCK PRICE PREDICTION

Background Problem Project

- Dalam dunia investasi, kemampuan untuk memprediksi harga saham menjadi hal yang sangat diinginkan oleh investor, analis pasar, dan institusi keuangan. Sekarang, kita bisa memanfaatkan perkembangan teknologi untuk membantu memprediksi harga saham.
- Pada project ini akan dibuat model untuk memprediksi harga saham.
- Data yang digunakan adalah data saham Apple dari Januari 2014 sampai Oktober 2023.
- Model akan dibuat dengan metode LSTM dan GRU lalu akan dievaluasi untuk menentukan model terbaik.

04: APPLE STOCK PRICE PREDICTION

Flowchart



04: APPLE STOCK PRICE PREDICTION

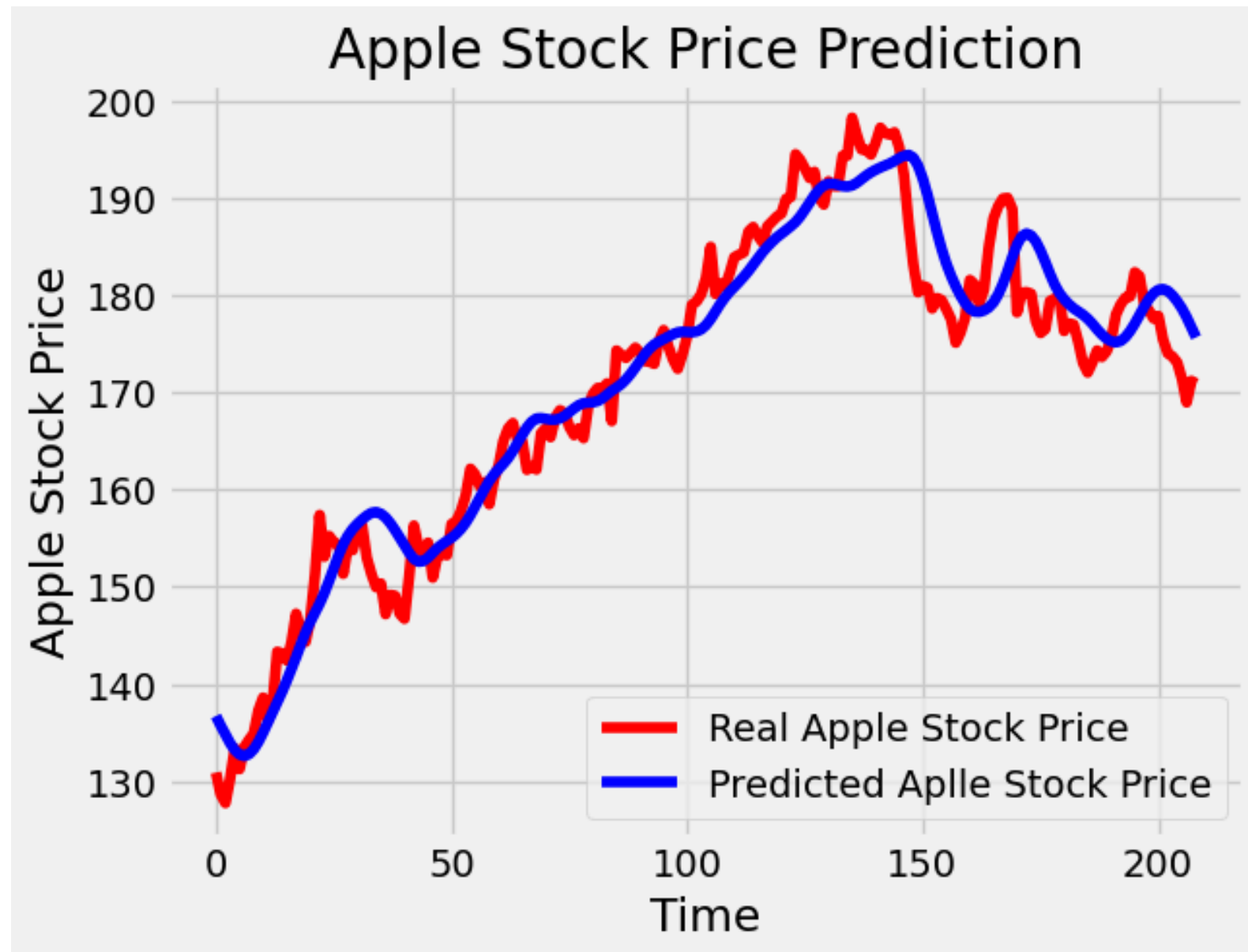
EDA



- Gambar di samping menunjukkan timeseries dari data harga oenutupan saham Apple.
- Grafik ini menunjukkan bahwa saham Apple mengalami pertumbuhan yang signifikan dalam jangka panjang. Tren utama cenderung naik, yang menunjukkan pertumbuhan perusahaan atau sentimen positif dari pasar.

04: APPLE STOCK PRICE PREDICTION

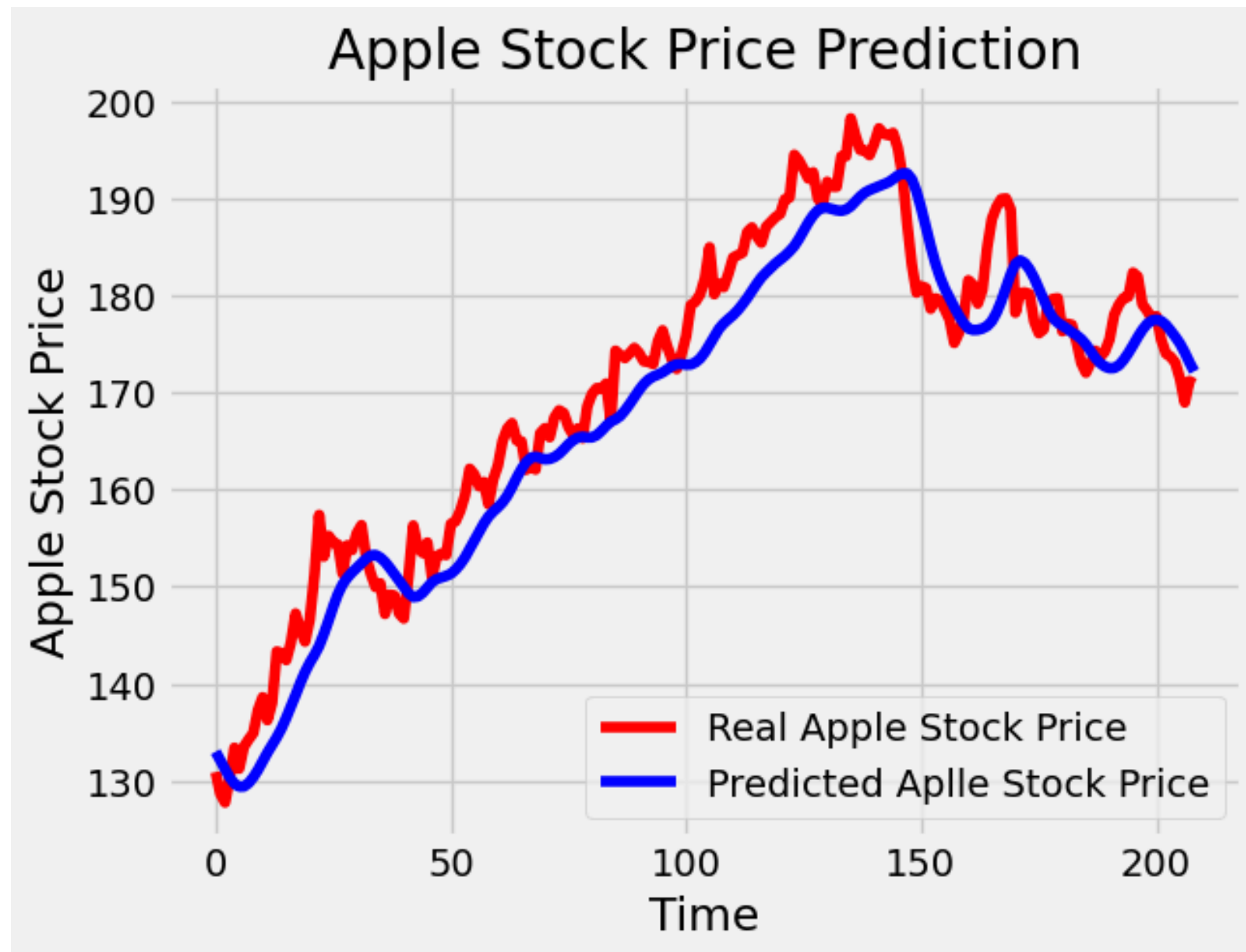
Modelling



- Gambar di samping menunjukkan timeseries dari prediksi harga saham penutupan Apple dengan menggunakan metode LSTM.
- Model dilakukan evaluasi menggunakan RMSE dan didapat nilai RMSE 4.26. Hal ini menunjukkan akurasi yang cukup baik dari model

04: APPLE STOCK PRICE PREDICTION

Modelling



- Gambar di samping menunjukkan timeseries dari prediksi harga saham penutupan Apple dengan menggunakan metode GRU.
- Model dilakukan evaluasi menggunakan RMSE dan didapat nilai RMSE 4.89. Hal ini tidak jauh berbeda dengan hasil evaluasi metode LSTM dan menunjukkan akurasi yang cukup baik dari model

04: APPLE STOCK PRICE PREDICTION

Kesimpulan

- Model yang telah dibuat memiliki akurasi yang cukup baik dalam memprediksi harga saham Apple.
- Baik metode LSTM maupun GRU sama-sama memiliki akurasi yang cukup baik, namun metode LSTM memiliki akurasi yang lebih baik dibandingkan metode GRU.

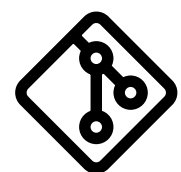
[Link github](#)



<https://www.linkedin.com/in/suci-rosyidatunnajah/>



sucirosyidatunnajah03@gmail.com



<https://github.com/SuciRosyida03>



CV



THANK YOU
