



# Forecasting Data Time Series Menggunakan MLP dan LSTM untuk Memprediksi Jumlah Produksi Bir

Muhammad Ibnu Choldun Rachmatullah

Fakultas Ilmu Terapan, D3 Rekayasa Perangkat Lunak Terapan, Universitas Telkom, Bandung, Indonesia

Email: muhammadibnuholdun@telkomuniversity.ac.id

**Abstrak**—Forecasting data deret waktu (*time series*) merupakan pendekatan penting dalam berbagai sektor seperti keuangan, energi, dan kesehatan. Seiring berkembangnya teknologi, metode *deep learning* seperti *Multi-Layer Perceptron (MLP)* dan *Long Short-Term Memory (LSTM)* semakin banyak digunakan untuk meningkatkan akurasi prediksi. Penelitian ini membandingkan kinerja kedua metode tersebut dalam melakukan forecasting terhadap dataset produksi bir bulanan di Australia yang bersifat *time series*. Model dilatih dan diuji menggunakan pembagian data 70% untuk pelatihan dan 30% untuk pengujian. Evaluasi performa dilakukan berdasarkan nilai *Root Mean Square Error (RMSE)* melalui 10 kali pengulangan eksperimen. Hasil menunjukkan bahwa MLP memiliki nilai RMSE yang lebih rendah dan deviasi standar yang lebih kecil dibandingkan LSTM, baik pada data pelatihan maupun pengujian. Hal ini mengindikasikan bahwa MLP lebih stabil dan efisien dalam menangani dataset dengan pola sederhana dan kompleksitas rendah, sedangkan LSTM cenderung memerlukan tuning lebih intensif serta memiliki risiko *overfitting* yang lebih tinggi. Oleh karena itu, MLP direkomendasikan sebagai metode forecasting alternatif yang lebih konsisten untuk skenario data serupa.

**Kata Kunci:** *Time Series; Forecasting; MLP; LSTM; RMSE*

**Abstract**—Time series data forecasting is an important approach in various sectors such as finance, energy, and healthcare. As technology advances, deep learning methods such as Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) are increasingly being used to improve prediction accuracy. This study compares the performance of these two methods in forecasting a time series dataset of monthly beer production in Australia. The model was trained and tested using a 70% training and 30% testing data split. Performance evaluation was based on the Root Mean Square Error (RMSE) value after 10 experimental repetitions. The results show that MLP has a lower RMSE value and a smaller standard deviation than LSTM, both on the training and testing data. This indicates that MLP is more stable and efficient in handling datasets with simple patterns and low complexity, while LSTM tends to require more intensive tuning and has a higher risk of overfitting. Therefore, MLP is recommended as a lighter and more consistent alternative forecasting method for similar data scenarios.

**Keywords:** Time Series; Forecasting; MLP; LSTM; RMSE

## 1. PENDAHULUAN

Data *time series* adalah jenis data yang dikumpulkan dan dicatat secara berurutan berdasarkan waktu, biasanya pada interval yang tetap seperti harian, mingguan, bulanan, atau tahunan. Setiap observasi dalam data *time series* memiliki urutan kronologis yang penting untuk dianalisis karena nilai di masa lalu dapat memengaruhi nilai di masa depan. Data ini sering digunakan untuk memodelkan dan meramalkan berbagai fenomena yang berkembang seiring waktu, seperti harga saham, suhu harian, penjualan produk, atau permintaan listrik. Ciri khas dari data *time series* adalah adanya pola-pola temporal seperti tren (peningkatan atau penurunan jangka panjang), musiman (fluktuasi periodik), dan siklus. Oleh karena itu, analisis *time series* memerlukan pendekatan khusus yang memperhitungkan keterkaitan antar waktu, seperti metode statistik (ARIMA) maupun pendekatan berbasis pembelajaran mesin seperti LSTM dan MLP yang mampu mengenali pola dalam urutan data.

Peramalan deret waktu (*time series forecasting*) telah menjadi elemen vital dalam berbagai sektor seperti keuangan, energi, meteorologi, rantai pasok, kesehatan, dan industri, berfungsi untuk memproyeksikan nilai masa depan berdasarkan pola historis temporal (Kim et al., 2024) (Li & Law, 2024). Model statistik klasik seperti ARIMA/SARIMA, *Exponential Smoothing (ETS)*, dan *Moving Average (MA)* tetap banyak digunakan karena keunggulan interpretabilitas dan performa baik pada data linear dan stasioner, meskipun terbatas dalam menangani non-linearity, musiman kompleks, dan noise tinggi. Oleh karenanya, sejak 2019, perhatian penelitian berkembang pesat ke metode Machine Learning (ML) seperti Random Forest, XGBoost, Support Vector Machines (SVM), dan regresi penalized yang mampu menangani interaksi kompleks tanpa prasyarat stasioneritas (Sina et al., 2023) (Oliveira & Ramos, 2024). Namun lompatan terbesar terjadi pada aplikasi Deep Learning (DL), dengan arsitektur seperti LSTM, CNN, Transformer, N-BEATS, dan Diffusion Models yang semakin populer (Kong et al., 2025) (Casolaro et al., 2023). Dalam penelitiannya, Kong dkk. menunjukkan LSTM unggul dalam akurasi meski dengan varians tinggi, sementara CNN menawarkan performa hampir setara dan lebih stabil (Kong et al., 2025). Sementara itu, (Masini et al., 2023) dan (Kim et al., 2024) menyajikan survei komprehensif yang mengeksplorasi arsitektur-diferensiasi (MLP, RNN, CNN, GNN, Transformer), serta munculnya model generatif (GAN, Diffusion, Mamba, foundation models), menyoroti tantangan seperti distribution shift, channel dependencies, dan feature extraction. Montet dkk. secara khusus mengkaji diffusion models, menunjukkan potensi penggunaan model ini, meski masih menghadapi tantangan tuning dan computational cost (Montet et al., 2024) (Lin et al., 2024).

Di Indonesia, metode Fuzzy Time Series (FTS) menjadi alternatif populer, terutama untuk dataset dengan ketidakpastian tinggi dan asumsi stasioneritas lemah. Studi (Okia Dinda Kelana et al., 2023) membandingkan model Chen dan Singh untuk revenue forecasting dan menemukan Singh lebih unggul (MAPE 5.17% vs 19.13%) (Okia Dinda

Kelana et al., 2023). Aplikasi FTS lainnya, seperti model Chen dengan Markov chain pada data kurs Rupiah–USD, mencatat MAPE sangat rendah (~0.36%) (Rahmad Revi Fadillah et al., 2023) (Hariyanto et al., 2023), sedangkan (Damayanti et al., 2023) menerapkan FTS Chen untuk impor minyak dan gas Indonesia dengan MAPE ~20%. Pendekatan hybrid juga mulai naik daun: (Sina et al., 2023) menyatakan ARIMA-RF, ARIMA-SVM, dan ARIMA-KNN konsisten mengungguli metode tunggal pada data energi dan trafik, memberikan bukti empiris bahwa penggabungan kelebihan statistik klasik dan ML menghasilkan performa yang lebih robust.

Dalam skala global, evolusi ini juga ditandai oleh kompetisi forecasting seperti M-Competitions (M3–M6), yang menunjukkan prediktor berbasis ML dan ensemble—seperti LightGBM dan model DL—mengungguli benchmark statistik. Meningkatnya kompleksitas arsitektur angin waktu nyata, seperti Transformer-specialized (Autoformer, PatchTST), foundation models, serta struktur hybrid dan generatif, menunjukkan masa depan forecasting semakin ditandai oleh scale, interpretability, dan efisiensi komputasi (Kim et al., 2024) (Kong et al., 2025).

Walau ML dan DL menawarkan akurasi superior, tantangan praktis seperti kebutuhan dataset besar, computational cost tinggi, risiko overfitting, dan interpretabilitas masih signifikan (Kim et al., 2024) (Kong et al., 2025). Sebagai perbandingan, FTS menyediakan solusi ringan dan mudah diimplementasi untuk domain tertentu, sementara hybrid menawarkan keseimbangan namun dengan kompleksitas implementasi dan tuning yang lebih besar (Sina et al., 2023).

Secara keseluruhan, literatur 2019–2025 mencerminkan diversifikasi metode forecasting: model klasik tetap relevan untuk kasus sederhana; ML menawarkan fleksibilitas tinggi dalam berbagai pola; DL unggul di dataset besar dan long-horizon forecasting; FTS relevan di data tidak pasti; hybrid dan generatif membuka jalan inovasi lanjut. Pendekatan paling sesuai sangat tergantung pada karakteristik data: ukuran, linearitas, musiman, tujuan interpretasi, dan sumber daya. Tulisan ini bertujuan menyajikan tinjauan sistematis dari metode-metode tersebut berdasarkan evaluasi teoretis dan empiris, untuk menyediakan panduan metodologis kepada peneliti dan praktisi dalam memilih teknik yang paling pas untuk kebutuhan forecasting.

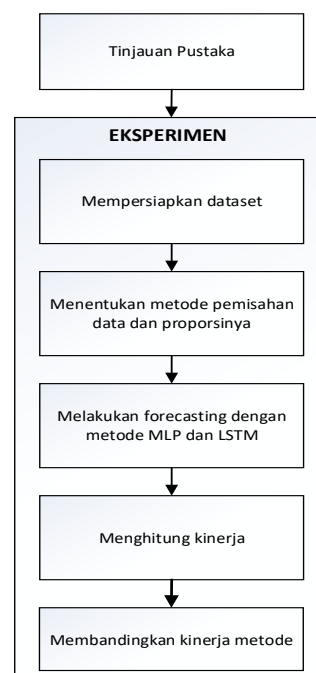
Pada penelitian ini dibahas penggunaan dua metode forecasting yang berbasis neural network yaitu Multi Layer Perceptron (MLP) dan Long Short-Term Memory (LSTM) yang diterapkan pada dataset tertentu yang bersifat time series. Penelitian ini bertujuan untuk mencari metode mana yang lebih cocok dari kedua metode tersebut dilihat dari kinerja metoda, yaitu dilihat dari nilai Root Mean Squared Error (RMSE) yang merupakan selisih antara nilai prediktif dan nilai aktual.

## 2. METODOLOGI PENELITIAN

Metodologi penelitian berupa tahapan atau langkah-langkah penelitian yang terdiri dari dua langkah utama yaitu tinjauan pustaka dan eksperimen. Implementasi untuk melakukan eksperimen dan perhitungan kinerja dilakukan menggunakan bahasa Python.

### 2.1 Tahapan Penelitian

Tahapan penelitian secara umum dapat dilihat pada gambar 1.



**Gambar 1.** Langkah-Langkah Penelitian



Pada Gambar 1 langkah pertama yang dilakukan adalah melakukan tinjauan pustaka membahas tentang MLP dan LSTM. Selanjutnya adalah tahapan eksperimen yang meliputi: mempersiapkan dataset, menentukan metode pemisahan data dan proporsinya, melakukan forecasting dataset, dan menghitung kinerja. Masing-masing langkah dijelaskan seperti berikut ini:

### 2.1.1 Tinjauan Pustaka

Multi-Layer Perceptron (MLP) merupakan salah satu jenis jaringan saraf tiruan yang digunakan secara luas dalam pembelajaran mesin untuk melakukan forecasting data deret waktu. MLP bekerja berdasarkan pendekatan supervised learning, di mana data historis (lag values) diubah menjadi format input-output untuk melatih model agar dapat memprediksi nilai masa depan. MLP terdiri dari lapisan input, satu atau lebih lapisan tersembunyi (hidden layers), dan satu lapisan output, dengan setiap neuron dihubungkan oleh bobot yang diperbarui melalui proses backpropagation untuk meminimalkan fungsi kerugian seperti Mean Squared Error (MSE). Salah satu keunggulan MLP adalah kemampuannya dalam memodelkan hubungan non-linear yang kompleks antar data, yang menjadikannya efektif untuk berbagai jenis data time series, baik univariat maupun multivariat (Egrioglu & Bas, 2024). Meski begitu, MLP tidak secara eksplisit mempertimbangkan urutan waktu seperti pada Recurrent Neural Network (RNN) atau Long Short-Term Memory (LSTM), sehingga pemilihan jumlah lag yang tepat menjadi sangat penting untuk menghasilkan prediksi yang akurat (Kong et al., 2025)(Lara-Benítez et al., 2021). Dalam praktiknya, MLP telah digunakan untuk memprediksi harga saham, konsumsi energi, dan data iklim dengan performa yang kompetitif dibandingkan metode statistik tradisional, terutama ketika pola data bersifat non-linear dan kompleks (Masini et al., 2023).

Long Short-Term Memory (LSTM) merupakan salah satu arsitektur jaringan saraf tiruan yang dirancang khusus untuk mengatasi keterbatasan model jaringan saraf biasa (seperti MLP) dalam menangani data berurutan atau sekuensial, termasuk forecasting deret waktu (time series). LSTM termasuk dalam keluarga Recurrent Neural Network (RNN) namun memiliki struktur sel memori internal yang memungkinkan model ini untuk mempertahankan informasi dalam jangka waktu yang lebih panjang, sekaligus mengatasi masalah vanishing gradient yang umum terjadi pada RNN standar. Dalam konteks forecasting, LSTM mampu mempelajari pola temporal jangka pendek maupun jangka panjang, sehingga cocok untuk memprediksi nilai masa depan berdasarkan data historis yang kompleks dan bersifat non-linear (Kong et al., 2025). LSTM memiliki tiga gerbang utama—input gate, forget gate, dan output gate—yang secara dinamis mengatur aliran informasi, memungkinkan model untuk fokus pada bagian data yang relevan dan mengabaikan yang tidak penting (Lim & Zohren, 2021). Berbagai studi telah menunjukkan bahwa LSTM unggul dalam tugas-tugas forecasting seperti prediksi beban listrik, harga saham, suhu, hingga konsumsi energi, karena kemampuannya dalam menangkap dependensi waktu yang tidak bisa dijangkau oleh metode statistik klasik seperti ARIMA maupun model MLP (Casolaro et al., 2023) (Li & Law, 2024) (Kong et al., 2025). Karena itu, LSTM menjadi salah satu metode populer dalam pembelajaran mesin modern untuk forecasting deret waktu dengan tingkat akurasi yang tinggi.

### 2.1.2 Mempersiapkan Dataset

Dataset yang digunakan diambil dari Kaggle yang bersifat time series. Dataset yang digunakan yaitu monthly-beer-production-in-austr yang terdiri dari 476 data, dengan atribut waktunya adalah “bulan” dan atribut yang diforecast adalah jumlah produksi beer. Alamat dari dataset ini adalah: <https://www.kaggle.com/datasets/shenba/time-series-datasets>.

### 2.1.3 Menentukan Metode Split Data

Metode pemisahan data yang digunakan adalah splitting yaitu membagi data menjadi dua bagian, yaitu untuk pelatihan (training) dan untuk pengujian (testing) dengan proporsi tertentu. Proporsi pembagian yang optimal untuk pelatihan dan pengujian adalah masing-masing 70-80% untuk pelatihan dan 20-30% untuk pengujian. Proporsi yang digunakan dalam penelitian ini adalah 70% data untuk pelatihan dan 30% data untuk pengujian.

### 2.1.4 Melakukan Forecasting

Metode forecasting yang dipakai pada penelitian ini ada dua yaitu MLP dan LSTM. MLP (Multi-Layer Perceptron) dan LSTM (Long Short-Term Memory) merupakan dua jenis jaringan saraf buatan yang dapat digunakan untuk melakukan forecasting data time series. MLP bekerja dengan mengubah data deret waktu menjadi format supervised learning menggunakan teknik sliding window, di mana nilai-nilai historis dijadikan input untuk memprediksi nilai selanjutnya. LSTM merupakan arsitektur khusus dari Recurrent Neural Network (RNN) yang dirancang untuk memproses data sekuensial dan mempertahankan informasi dari waktu sebelumnya melalui mekanisme gerbang (gates). Proses forecasting diimplementasikan menggunakan bahasa Python dengan menggunakan pustaka tensorflow, khususnya Keras untuk membangun model MLP dan LSTM; scikit-learn untuk normalisasi data dan evaluasi model. Parameter yang digunakan dalam MLP adalah: dense layer dengan 50 neuron dan ReLU sebagai activation function, output layer: dense dengan 1 neuron, optimizer: adam, Loss function: mean\_squared\_error, epochs: 100, batch size: 1. Parameter yang digunakan dalam LSTM adalah: LSTM layer dengan 50 unit, output layer: dense dengan 1 neuron, optimizer: adam, loss function: mean\_squared\_error, epochs: 100, batch size: 1.

### 2.1.5 Menghitung Kinerja

Untuk mengevaluasi kinerja metode forecasting yang digunakan adalah dengan menghitung Root Mean Square Error (RMSE). RMSE (Root Mean Square Error) adalah salah satu metrik evaluasi yang digunakan untuk mengukur seberapa besar kesalahan antara nilai yang diprediksi oleh model dengan nilai aktual dalam sebuah dataset. RMSE dihitung dengan mengambil akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan nilai sebenarnya. Metrik ini memberikan penilaian yang sensitif terhadap kesalahan besar karena penggunaan kuadrat dalam perhitungannya, sehingga sangat berguna untuk menilai keakuratan model prediktif, terutama dalam kasus regresi atau peramalan (forecasting) data time series. Nilai RMSE yang lebih kecil menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang lebih rendah dan, dengan demikian, kinerjanya lebih baik. Namun, karena RMSE memiliki satuan yang sama dengan data aslinya, interpretasi hasilnya perlu disesuaikan dengan konteks dan skala dari data yang dianalisis. Rumus RMSE adalah sebagai berikut:

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

Root Mean Square Error (RMSE) merupakan salah satu ukuran yang digunakan untuk mengevaluasi tingkat kesalahan dalam suatu model prediksi. Secara matematis, RMSE dihitung dengan mengambil akar kuadrat dari rata-rata jumlah kuadrat selisih antara nilai aktual dan nilai prediksi. Pada persamaan (1), n menunjukkan jumlah data atau observasi,  $y_i$  merepresentasikan nilai aktual pada observasi ke-i, sedangkan  $\hat{y}_i$  merupakan nilai prediksi pada observasi ke-i. Selisih antara nilai aktual dan nilai prediksi kemudian dikuadratkan, yaitu  $(y_i - \hat{y}_i)^2$ , untuk memastikan bahwa semua nilai error bernilai positif. Nilai RMSE yang lebih kecil menunjukkan bahwa model prediksi memiliki tingkat kesalahan yang rendah dan semakin mendekati nilai aktual, sehingga dapat dikatakan bahwa model tersebut memiliki kinerja yang lebih baik.

## 3. HASIL DAN PEMBAHASAN

Bagian pembahasan berisi hasil dan analisis dari eksperimen forecasting dari dataset yang digunakan menggunakan metode MLP dan LSTM.

### 3.1 Hasil Eksperimen

Untuk masing-masing metode dilakukan pengulangan sebanyak 10 kali dan pada setiap pengulangan dilakukan pengukuran RMSE baik untuk data training maupun data testing. Setelah dilakukan pengulangan 10 kali untuk masing-masing metode kemudian dihitung nilai rerata dan simpangan baku atau standard deviation (SD)-nya dan hasilnya disajikan pada Tabel 1.

**Tabel 1.** Nilai RMSE

Pengulangan	LSTM Train	LSTM Test	MLP Train	MLP Test
1	10.274	12.992	8.295	11.225
2	9.834	11.869	8.686	11.470
3	10.085	11.284	8.895	11.610
4	9.686	11.937	9.501	11.397
5	9.851	12.272	8.666	12.287
6	10.030	11.091	9.319	10.939
7	9.919	12.504	9.099	11.942
8	9.653	11.987	9.056	12.307
9	9.935	11.246	8.972	12.317
10	11.411	14.910	9.154	11.818
Rerata	10.068	12.209	8.964	11.731
SD	0.506	1.119	0.349	0.485

Setelah dilakukan pengulangan 10 kali untuk masing-masing metode kemudian dihitung nilai rerata dan simpangan baku atau standard deviation (SD)-nya. Eksperimen yang dilakukan terhadap masing-masing metoda diulang sebanyak 10 kali, dan kemudian dicari nilai reratanya. Untuk mengevaluasi perbedaan kinerja model LSTM dan MLP, dilakukan uji beda berpasangan (paired t-test) terhadap nilai RMSE dari hasil pelatihan (train) dan pengujian (test) selama 10 kali pengulangan. Dari hasil uji statistik diperoleh hasil sebagai berikut:

**Tabel 2.** Hasil Uji Statistik

Perbandingan	Nilai t-statistik	p value	Interpretasi
LSTM vs MLP (Train)	5.62	0.0003	Signifikan ( $p < 0.05$ ) → Terdapat perbedaan bermakna
LSTM vs MLP (Test)	1.28	0.2324	Tidak signifikan ( $p > 0.05$ ) → Tidak ada perbedaan nyata

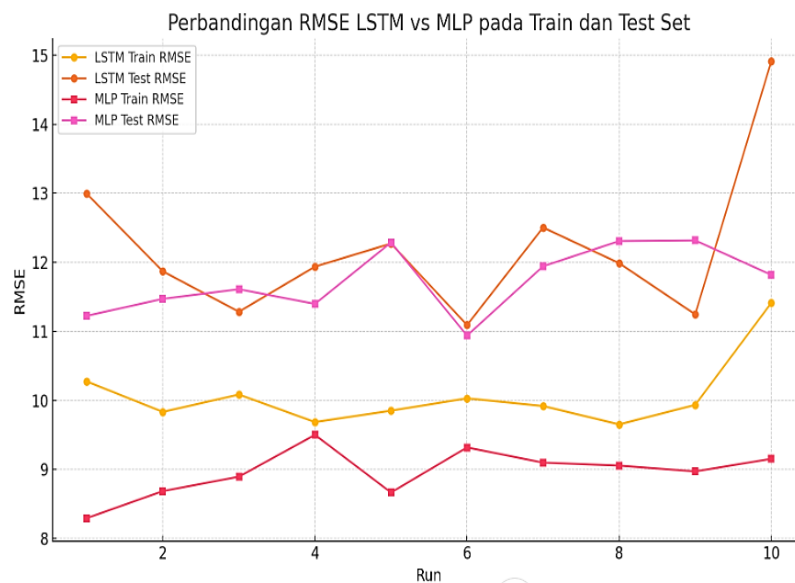
Berdasarkan hasil uji t di atas, terdapat perbedaan yang signifikan secara statistik pada nilai RMSE pelatihan antara LSTM dan MLP. Nilai p sebesar 0.0003 menunjukkan bahwa MLP secara konsisten memiliki performa pelatihan yang lebih baik (RMSE lebih rendah) dibandingkan LSTM. Hal ini terlihat juga dari visualisasi, di mana grafik MLP Train berada di bawah grafik LSTM Train untuk sebagian besar pengulangan.

Namun, pada data pengujian, perbedaan antara LSTM dan MLP tidak signifikan secara statistik ( $p = 0.2324$ ). Artinya, meskipun MLP menunjukkan rata-rata RMSE yang sedikit lebih rendah pada data test, perbedaan ini kemungkinan besar terjadi secara acak dan tidak cukup kuat untuk menyimpulkan bahwa salah satu model secara konsisten lebih unggul dalam hal generalisasi terhadap data baru.

Dengan demikian, meskipun MLP unggul pada pelatihan, kinerja prediktif pada data uji antara kedua model secara statistik tidak berbeda signifikan, yang berarti baik LSTM maupun MLP sama-sama layak dipertimbangkan dalam konteks forecasting data time series, tergantung pada kompleksitas data dan kebutuhan model.

### 3.2 Pembahasan

Hasil eksperimen yang sudah ada pada Tabel 1 dapat disajikan dalam bentuk gambar seperti dapat dilihat pada Gambar 2.



**Gambar 2.** Grafik Nilai RMSE

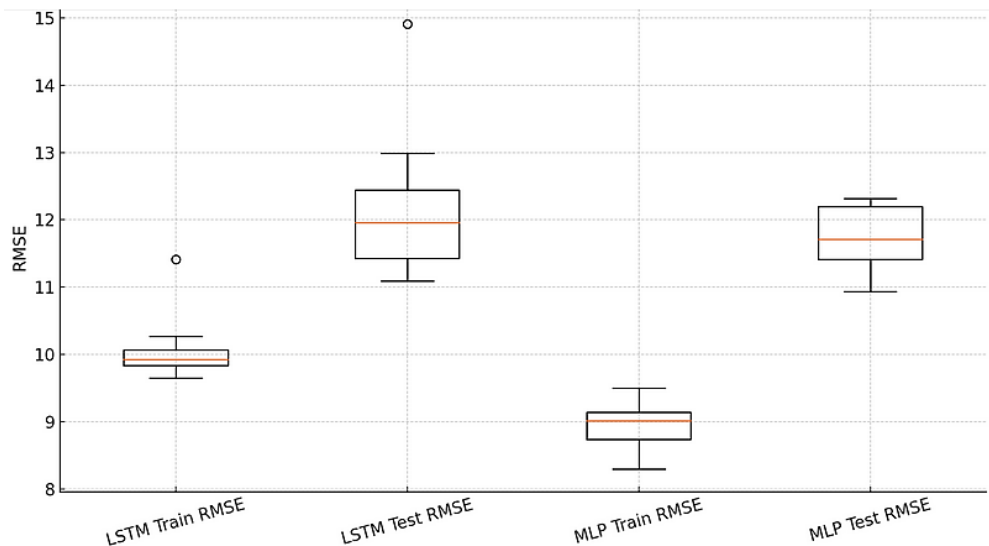
Peramalan data deret waktu (time series forecasting) merupakan salah satu pendekatan penting dalam berbagai bidang seperti keuangan, kesehatan, energi, dan meteorologi. Dua metode deep learning yang populer dalam pendekatan ini adalah LSTM dan MLP, masing-masing dengan karakteristik dan kekuatan yang berbeda dalam memodelkan pola data.

Berdasarkan data eksperimen yang dianalisis, MLP menunjukkan rata-rata RMSE pada data pelatihan sebesar 8.96, sedangkan LSTM sebesar 10.07. Ini menunjukkan bahwa MLP lebih baik dalam menyesuaikan diri terhadap data pelatihan. Rendahnya RMSE pada pelatihan menandakan bahwa MLP mampu menangkap pola data secara efektif tanpa terlalu banyak error residual. Hasil ini sejalan dengan penelitian oleh (Sineglazov & Horbatiuk, 2025)(Bandara et al., 2020)(Ping et al., 2024)(Chomiak & Hu, 2024) yang menyatakan bahwa MLP, dengan struktur yang lebih sederhana, sering kali lebih efisien dalam mempelajari pola linier dan jangka pendek dibandingkan LSTM, terutama pada dataset dengan kompleksitas rendah. Saat mengevaluasi performa pada data pengujian, LSTM menghasilkan rata-rata RMSE sebesar 12.21, sedangkan MLP hanya sebesar 11.73. Meskipun perbedaan ini tidak terlalu signifikan, perlu dicatat bahwa standar deviasi RMSE pengujian LSTM cukup tinggi (~1.12), sementara MLP jauh lebih stabil dengan deviasi hanya sekitar 0.48. Hal ini mengindikasikan bahwa performa LSTM lebih rentan terhadap fluktuasi pada data testing atau pengaturan awal pelatihan (initialization, hyperparameter), yang juga ditekankan dalam studi oleh (Kong et al., 2025) bahwa LSTM cenderung overfit pada data pelatihan jika tidak distabilkan melalui regularisasi atau dropout.

LSTM, sebagai bagian dari keluarga Recurrent Neural Network (RNN), dirancang untuk menangkap hubungan jangka panjang dalam data deret waktu. Namun, struktur kompleks ini seringkali membutuhkan tuning parameter lebih intensif dan pelatihan yang lebih lama (Kong et al., 2025). Sementara itu, MLP memiliki arsitektur feedforward yang lebih sederhana, namun tetap mampu memberikan hasil kompetitif pada data dengan tren musiman atau fluktuasi ringan. Penelitian oleh (Li & Law, 2024) juga menunjukkan bahwa dalam beberapa aplikasi energi dan iklim, model MLP dapat memberikan hasil akurat yang stabil tanpa memerlukan perhatian khusus terhadap urutan data temporal seperti halnya LSTM. Stabilitas performa model adalah salah satu indikator penting untuk implementasi sistem prediktif dalam praktik. Dalam analisis ini, deviasi standar MLP untuk Train RMSE hanya 0.35, dibandingkan 0.51 pada LSTM. Demikian pula, pada Test RMSE, MLP tetap lebih stabil dengan deviasi 0.48 dibanding LSTM 1.12. Hal ini

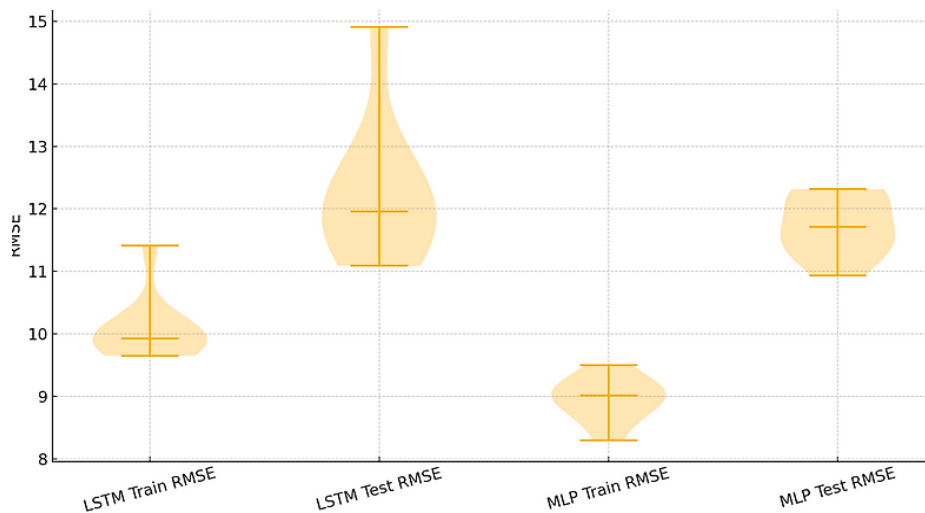
menunjukkan bahwa MLP lebih konsisten dalam performa lintas beberapa run eksperimental, sesuai dengan temuan dari (Kong et al., 2025) (Li & Law, 2024) yang menyatakan bahwa pada dataset dengan ukuran kecil–menengah, model sederhana cenderung lebih stabil dibanding model kompleks. Perbedaan RMSE antara train dan test yang tinggi pada LSTM (10.07 ke 12.21) dibandingkan dengan MLP (8.96 ke 11.73) dapat mengindikasikan overfitting ringan pada LSTM. Hal ini dikarenakan LSTM cenderung terlalu menyesuaikan dengan data pelatihan dan tidak berhasil mentransfer pola secara optimal ke data pengujian. Dalam studi oleh (Mojtahedi et al., 2025), ditemukan bahwa model LSTM sering kali membutuhkan mekanisme regulasi tambahan seperti dropout, early stopping, dan data augmentation untuk mencapai generalisasi yang lebih baik.

Dari Tabel 1 dapat disajikan pula dalam bentuk grafik lain untuk analisis yang lebih mendalam, misalnya dalam grafik box plot, violin plot, dan bar chart seperti dalam Gambar 3 berikut:



**Gambar 3.** Distribusi RMSE per Model (Box Plot)

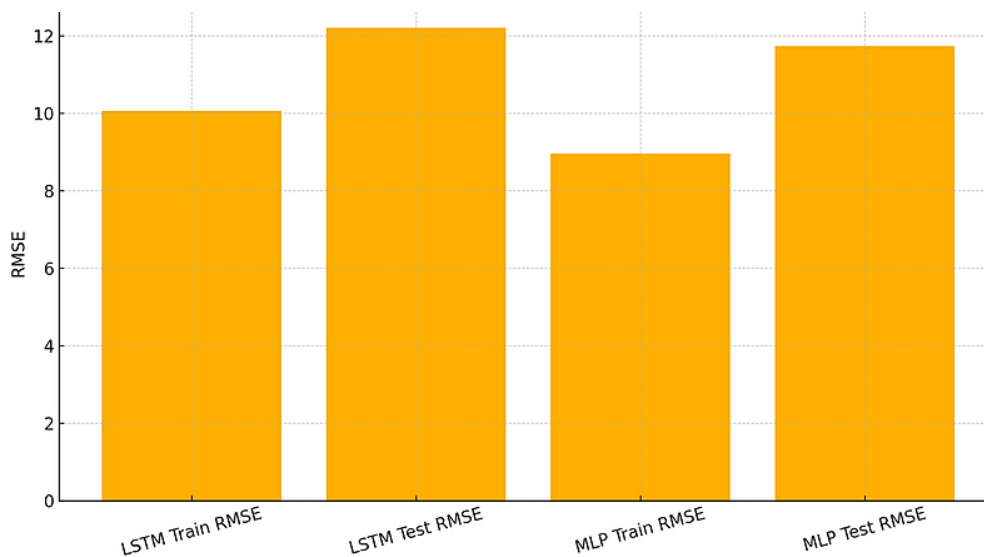
Box plot, atau diagram kotak pada Gambar 3, adalah jenis visualisasi statistik yang digunakan untuk menunjukkan sebaran dan penyimpangan data numerik. Grafik ini menampilkan lima elemen utama: nilai minimum, kuartil pertama (Q1), median (Q2), kuartil ketiga (Q3), dan nilai maksimum. Bagian dalam kotak menggambarkan rentang interkuartil (IQR), yaitu antara Q1 dan Q3, sementara garis di tengah kotak menunjukkan median sebagai nilai tengah data. Garis (whisker) yang menjulur dari kotak menunjukkan rentang data di luar IQR, dan titik-titik individual di luar whisker menandakan outlier atau nilai yang menyimpang jauh dari distribusi utama. Box plot sangat berguna untuk membandingkan distribusi antar kelompok data serta mengidentifikasi simetri, variabilitas, dan potensi outlier dengan cepat dan jelas.



**Gambar 4.** Distribusi RMSE per Model (Violin Plot)

Violin plot seperti disajikan pada gambar 4, adalah jenis visualisasi data yang menggabungkan elemen dari box plot dan grafik kerapatan distribusi (density plot). Grafik ini menunjukkan distribusi probabilitas dari data numerik melalui bentuk simetris menyerupai biola, di mana lebar grafik pada setiap titik mencerminkan kepadatan data di nilai

tersebut. Seperti box plot, violin plot juga menampilkan informasi seperti median dan rentang kuartil, tetapi dengan tambahan bentuk yang menggambarkan penyebaran data secara lebih mendetail. Violin plot sangat berguna untuk membandingkan distribusi antar kelompok data, terutama ketika ingin memahami bentuk distribusi yang mungkin tidak simetris atau mengandung beberapa puncak (multimodal).



**Gambar 5.** Rerata RMSE per Model

Bar chart, atau diagram batang yang ditunjukkan pada Gambar 5, adalah jenis visualisasi data yang digunakan untuk menampilkan dan membandingkan nilai-nilai numerik dari berbagai kategori. Grafik ini terdiri dari batang-batang vertikal atau horizontal, di mana panjang masing-masing batang mewakili besar nilai dari suatu kategori. Bar chart sangat efektif untuk menyoroti perbedaan antar kelompok, mengenali pola, dan mempermudah interpretasi data secara visual. Karena tampilannya yang sederhana dan mudah dibaca, bar chart sering digunakan dalam laporan statistik, presentasi, dan analisis data untuk memperjelas informasi kuantitatif secara cepat dan intuitif.

Berikut adalah analisis lanjutan dari hasil eksperimen yang disajikan pada tabel 1 jika dilihat dari tiga visualisasi menggunakan box-plot, violin plot, dan bar chart. Pertama, box-plot memaparkan gambaran kuartil dan pencilan secara ringkas. Untuk LSTM Train, interkuartil (IQR) hanya  $\pm 0,23$  poin ( $Q1 \approx 9,84$ ;  $Q3 \approx 10,07$ ), menandakan pelatihan LSTM relatif stabil. Namun begitu grafik beralih ke LSTM Test, IQR melonjak lebih dari empat kali lipat ( $\pm 1,02$ ), garis whisker melebar, dan satu pencilan ekstrem muncul di atas 14 RMSE. Selisih ini langsung mengisyaratkan dua hal: varian uji LSTM tinggi dan ada risiko run “buruk” yang menodai konsistensi. Di sisi lain, MLP menunjukkan IQR yang masih rapat pada train ( $\pm 0,40$ ) dan walau melebar saat test ( $\pm 0,79$ ) tetap lebih sempit ketimbang LSTM. Hal ini mencerminkan karakter algoritme feed-forward yang lebih deterministik pada konfigurasi hiperparameter tetap. Violin plot memvalidasi bacaan box-plot dengan menambahkan dimensi kepadatan. Bentuk “seruling” LSTM Test tampak gemuk di tengah (banyak run bertumpuk di RMSE 11,5–12,5) tetapi bertaji panjang ke atas, visualisasi skew-kanan akibat outlier run ke-10 (14,9 RMSE). MLP Test sebaliknya membentuk siluet yang hampir simetris dengan kepadatan puncak di kisaran 11,3–11,8 RMSE dan ekor pendek; ini menegaskan bahwa performa tipikal MLP lebih dapat diprediksi. Selain itu, tinggi median kedua model (garis putih pada violin) tampak persis di simpul bar chart rerata: MLP di kisaran 11,7, LSTM di 12,2. Terakhir, bar chart rerata menyaring detail ke dalam satu perbandingan absolut. Rerata RMSE uji menempatkan urutan performa: MLP Test (11,73) < LSTM Test (12,21), sedangkan rerata latih menegaskan efisiensi MLP yang memang lebih rendah (8,96 vs 10,07). Tetapi angka rerata sekaligus menyingkap gap over-fitting: LSTM 2,14 poin, MLP 2,77 poin. Dengan kata lain, MLP menang akurasi berkat baseline train yang sangat rendah, namun membayar harga berupa jarak train–test yang sedikit lebih lebar. Gabungan tiga grafik ini memperjelas trade-off: MLP unggul pada akurasi rata-rata dan stabilitas distribusi, sedangkan LSTM—meski kurang presisi—menawarkan gap over-fitting yang relatif lebih kecil tetapi rentan fluktuasi besar akibat pencilan. Dalam konteks produksi, risiko sebuah run “terpeleset” 3 RMSE di atas median bisa lebih merugikan daripada selisih rerata 0,5 RMSE; karenanya MLP muncul sebagai pilihan lebih aman, dengan catatan regulasi ekstra (dropout, early stopping) layak ditambahkan untuk merapatkan jarak train–test-nya.

## 4. KESIMPULAN

Penelitian ini membandingkan dua metode deep learning populer, yaitu Multi-Layer Perceptron (MLP) dan Long Short-Term Memory (LSTM), dalam melakukan forecasting data deret waktu pada dataset produksi bir bulanan di Australia. Hasil eksperimen menunjukkan bahwa MLP memiliki performa yang lebih stabil dan konsisten dibandingkan LSTM, baik pada data pelatihan maupun pengujian, ditinjau dari nilai Root Mean Square Error (RMSE) dan standar deviasi yang lebih rendah. Berdasarkan data eksperimen yang dianalisis, MLP menunjukkan rata-rata RMSE pada rendah



dibandingkan LSTM, yang menunjukkan bahwa MLP lebih baik dalam menyesuaikan diri terhadap data. Meskipun LSTM dirancang untuk menangkap pola jangka panjang dalam data sekuensial, hasil yang diperoleh memperlihatkan bahwa model ini cenderung mengalami overfitting ringan dan memiliki variabilitas performa yang lebih tinggi dibandingkan MLP. Di sisi lain, MLP dengan arsitektur feedforward yang lebih sederhana justru menunjukkan ketahanan dan keakuratan yang lebih baik pada dataset dengan karakteristik yang tidak terlalu kompleks. Oleh karena itu, dalam konteks data time series dengan kompleksitas rendah hingga sedang, MLP dapat menjadi alternatif yang lebih efisien dan stabil dibandingkan LSTM, terutama ketika sumber daya komputasi terbatas dan stabilitas hasil menjadi prioritas utama. Keterbatasan dari penelitian ini adalah hanya membandingkan dua metode forecasting dan tidak dilakukan optimisasi parameter pada kedua metode digunakan.

## REFERENCES

- Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, 140. <https://doi.org/10.1016/j.eswa.2019.112896>
- Casolaro, A., Capone, V., Iannuzzo, G., & Camastra, F. (2023). Deep Learning for Time Series Forecasting: Advances and Open Problems. *Information (Switzerland)*, 14(11). <https://doi.org/10.3390/info14110598>
- Chomiak, T., & Hu, B. (2024). Time-series forecasting through recurrent topology. *Communications Engineering*, 3(1), 1–10. <https://doi.org/10.1038/s44172-023-00142-8>
- Damayanti, S., Yosmar, S., & Afandi, N. (2023). Implementation of Fuzzy Time Series Chen for Forecasting Indonesian Oil and Gas Imports Value. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 0685–0694. <https://doi.org/10.30598/barekengvol17iss2pp0685-0694>
- Egrioglu, E., & Bas, E. (2024). A new deep neural network for forecasting: Deep dendritic artificial neural network. *Artificial Intelligence Review*, 57(7), 1–25. <https://doi.org/10.1007/s10462-024-10790-7>
- Hariyanto, S., Sumanto, Y. D., Khabibah, S., & Zaenurrohman. (2023). Average-Based Fuzzy Time Series Markov Chain Based on Frequency Density Partitioning. *Journal of Applied Mathematics*, 2023. <https://doi.org/10.1155/2023/9319883>
- Kim, J., Kim, H., Kim, H., Lee, D., & Yoon, S. (2024). *A Comprehensive Survey of Time Series Forecasting: Architectural Diversity and Open Challenges*. <http://arxiv.org/abs/2411.05793>
- Kong, X., Chen, Z., Liu, W., Ning, K., Zhang, L., Muhammad Marier, S., Liu, Y., Chen, Y., & Xia, F. (2025). Deep learning for time series forecasting: a survey. In *International Journal of Machine Learning and Cybernetics* (Issue 0123456789). Springer Berlin Heidelberg. <https://doi.org/10.1007/s13042-025-02560-w>
- Li, W., & Law, K. L. E. (2024). Deep Learning Models for Time Series Forecasting: A Review. *IEEE Access*, 12(May), 92306–92327. <https://doi.org/10.1109/ACCESS.2024.3422528>
- Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194). <https://doi.org/10.1098/rsta.2020.0209>
- Lin, L., Li, Z., Li, R., Li, X., & Gao, J. (2024). Diffusion models for time-series applications: a survey. *Frontiers of Information Technology and Electronic Engineering*, 25(1), 19–41. <https://doi.org/10.1631/FITEE.2300310>
- Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1), 76–111. <https://doi.org/10.1111/joes.12429>
- Mojtahedi, F. F., Yousefpour, N., Chow, S. H., & Cassidy, M. (2025). Deep Learning for Time Series Forecasting: Review and Applications in Geotechnics and Geosciences. *Archives of Computational Methods in Engineering*. <https://doi.org/10.1007/s11831-025-10244-5>
- Montet, F., Pasquier, B., Wolf, B., & Hennebert, J. (2024). Enabling Diffusion Model for Conditioned Time Series Generation †. *Engineering Proceedings*, 68(1). <https://doi.org/10.3390/engproc2024068025>
- Okia Dinda Kelana, Atus Amadi Putra, Nonong Amalita, & Admi Salma. (2023). Comparison of the Chen and Singh's Fuzzy Time Series Methods in Forecasting Farmer Exchange Rates in Indonesia. *UNP Journal of Statistics and Data Science*, 1(4), 264–270. <https://doi.org/10.24036/ujsds/vol1-iss4/36>
- Oliveira, J. M., & Ramos, P. (2024). Evaluating the Effectiveness of Time Series Transformers for Demand Forecasting in Retail. *Mathematics*, 12(17), 1–28. <https://doi.org/10.3390/math12172728>
- Ping, W., Hu, Y., & Luo, L. (2024). Price Forecast of Treasury Bond Market Yield: Optimize Method Based on Deep Learning Model. *IEEE Access*, 12(November), 194521–194539. <https://doi.org/10.1109/ACCESS.2024.3519438>
- Rahmad Revi Fadillah, Dony Permana, Yenni Kurniawati, & Admi Salma. (2023). Application of the Fuzzy Time Series-markov Chain Method to the Rupiah Exchange Rate Against the US Dollar (USD). *UNP Journal of Statistics and Data Science*, 1(4), 369–376. <https://doi.org/10.24036/ujsds/vol1-iss4/91>
- Sina, L. B., Secco, C. A., Blazevic, M., & Nazemi, K. (2023). Hybrid Forecasting Methods—A Systematic Review. *Electronics (Switzerland)*, 12(9). <https://doi.org/10.3390/electronics12092019>
- Sineglazov, V., & Horbatiuk, V. (2025). Time Series Forecasting Using Recurrent Neural Networks Based on Recurrent Sigmoid Piecewise Linear Neurons. *Applied Artificial Intelligence*, 39(1). <https://doi.org/10.1080/08839514.2025.2490057>