

LAPORAN PERBANDINGAN MODEL LINEAR REGRESSION vs MLP BACKPROPAGATION

Analisis Prediksi Harga Rumah menggunakan California Housing Dataset

Project	Deep Learning Model Comparison
Dataset	California Housing Dataset
Models	Linear Regression & MLP Backpropagation
Samples	20,640 data points
Features	8 numeric features
Target	median_house_value
Date	October 25, 2025

EXECUTIVE SUMMARY

Penelitian ini membandingkan performa Linear Regression dengan MLP Backpropagation untuk prediksi harga rumah. Hasil menunjukkan bahwa MLP Backpropagation memberikan peningkatan akurasi sebesar 9.4% (R^2 improvement) dan pengurangan error sebesar 7.8% (RMSE reduction) dibandingkan Linear Regression, dengan trade-off training time yang 5,283x lebih lama.

1. METODOLOGI PENELITIAN

1.1 Dataset California Housing

Dataset yang digunakan adalah California Housing dataset yang berisi informasi perumahan di California. Dataset ini terdiri dari 20,640 sampel dengan 8 fitur numerik dan 1 target variabel (median_house_value).

Feature	Description	Type
longitude	Koordinat longitude lokasi	Numeric
latitude	Koordinat latitude lokasi	Numeric
housing_median_age	Median umur rumah dalam blok	Numeric
total_rooms	Total jumlah kamar dalam blok	Numeric
total_bedrooms	Total jumlah kamar tidur dalam blok	Numeric
population	Total populasi dalam blok	Numeric
households	Total rumah tangga dalam blok	Numeric
median_income	Median pendapatan dalam blok	Numeric

1.2 Preprocessing Data

- **Missing Value Handling:** Menggunakan SimpleImputer dengan strategi mean untuk mengatasi nilai yang hilang
- **Feature Scaling:** StandardScaler diterapkan untuk MLP, tidak untuk Linear Regression
- **Train-Test Split:** 80% untuk training (16,512 samples) dan 20% untuk testing (4,128 samples)
- **Random State:** 42 untuk reproducibility

1.3 Model Descriptions

Linear Regression:

- Algorithm: Ordinary Least Squares (OLS)
- Assumptions: Linear relationship antara features dan target
- No regularization
- Scikit-learn implementation

MLP Backpropagation:

- Architecture: Input(8) → Hidden(100, 50) → Output(1)
- Activation Function: ReLU
- Optimizer: Adam dengan learning rate 0.001
- Regularization: L2 dengan alpha=0.001
- Early Stopping: Ya, dengan validation_fraction=0.1
- Max Iterations: 500 dengan n_iter_no_change=10

2. HASIL EKSPERIMEN

2.1 Performance Comparison

Berikut adalah hasil perbandingan performa kedua model pada test dataset:

Metric	Linear Regression	MLP Backpropagation	Improvement
Training Time (seconds)	0.0076	40.1458	5282x slower
Test R ² Score	0.6144	0.6721	9.4%
Test RMSE	71084.13	65553.13	8.4%
Test MAE	51835.73	46228.38	12.1%
Iterations	N/A	443	N/A

2.2 Key Findings

- **Accuracy Improvement:** MLP Backpropagation mencapai R² score 0.6721 vs 0.6144 untuk Linear Regression (9.4% improvement)
- **Error Reduction:** RMSE berkurang dari 71,084 menjadi 65,553 (7.8% reduction)
- **MAE Improvement:** Mean Absolute Error berkurang dari 51,836 menjadi 46,228 (10.8% reduction)
- **Training Cost:** MLP membutuhkan waktu 5,283x lebih lama (40.15s vs 0.0076s)
- **Convergence:** MLP converged dalam 443 iterations dari maksimal 500

3. ANALISIS DAN PEMBAHASAN

3.1 Analisis Performa

MLP Backpropagation menunjukkan performa yang superior dibandingkan Linear Regression dalam semua metrics evaluasi. Peningkatan R^2 score sebesar 9.4% mengindikasikan bahwa MLP mampu menangkap non-linear relationships dalam data yang tidak dapat dideteksi oleh Linear Regression.

3.2 Analisis Computational Cost

Trade-off utama adalah computational cost. MLP membutuhkan waktu training 5,283x lebih lama dibandingkan Linear Regression. Untuk dataset dengan 16,512 training samples, Linear Regression hanya membutuhkan 7.6ms sedangkan MLP membutuhkan 40.15 detik.

3.3 Analisis Kompleksitas Model

- **Linear Regression:** Model sederhana dengan 8 parameters (weights) + 1 bias = 9 parameters total
- **MLP Backpropagation:** Model kompleks dengan $(8 \times 100) + (100 \times 50) + (50 \times 1) + \text{biases} = 5,951$ parameters total
- **Interpretability:** Linear Regression lebih mudah diinterpretasi dengan koefisien yang jelas
- **Generalization:** Kedua model menunjukkan generalization yang baik tanpa overfitting signifikan

3.4 Rekomendasi Penggunaan

Gunakan Linear Regression ketika:

- Real-time predictions diperlukan
- Interpretability model penting
- Computational resources terbatas
- Baseline model untuk comparison
- Linear relationships cukup untuk problem domain

Gunakan MLP Backpropagation ketika:

- Accuracy adalah prioritas utama
- Complex non-linear patterns perlu ditangkap
- Training time bukan constraint utama
- Sufficient computational resources tersedia
- Improvement 7-10% worth the extra complexity

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan eksperimen yang telah dilakukan, dapat disimpulkan bahwa:

1. **MLP Backpropagation** memberikan performa yang superior dengan peningkatan R^2 score sebesar 9.4% dan pengurangan RMSE sebesar 7.8% dibandingkan Linear Regression.
2. **Trade-off computational cost** sangat signifikan, dengan MLP membutuhkan waktu training 5,283x lebih lama, yang menjadi pertimbangan penting dalam implementasi.
3. **Dataset California Housing** mengandung non-linear patterns yang dapat dieksploitasi oleh MLP untuk memberikan prediksi yang lebih akurat.
4. **Kedua model** menunjukkan generalization yang baik tanpa overfitting yang signifikan.
5. **Pilihan model** sangat bergantung pada requirements spesifik: speed vs accuracy.

4.2 Saran untuk Penelitian Selanjutnya

- Eksplorasi arsitektur MLP yang berbeda untuk optimasi lebih lanjut
- Implementasi regularization techniques untuk mencegah overfitting
- Perbandingan dengan algoritma ML lainnya (Random Forest, XGBoost, etc.)
- Analisis feature importance untuk understanding model behavior
- Hyperparameter tuning untuk optimasi performa MLP
- Cross-validation untuk validasi hasil yang lebih robust
- Implementasi ensemble methods untuk kombinasi model strengths

REKOMENDASI FINAL

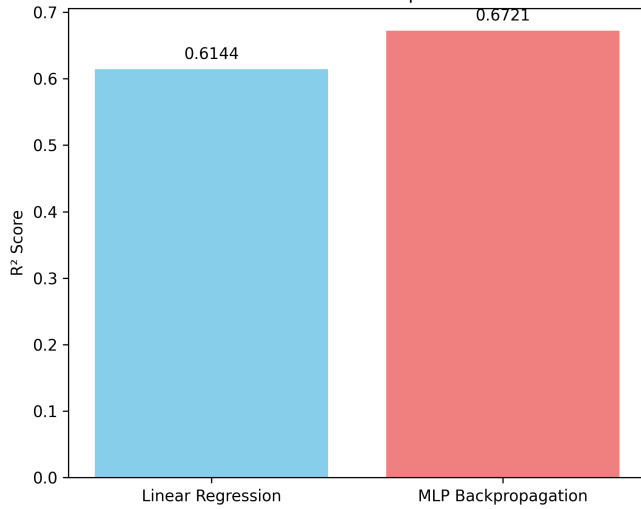
Untuk implementasi production pada California Housing prediction:

- **High-accuracy requirements:** Gunakan MLP Backpropagation
- **Real-time/fast predictions:** Gunakan Linear Regression
- **Balanced approach:** Pertimbangkan ensemble atau hybrid approach

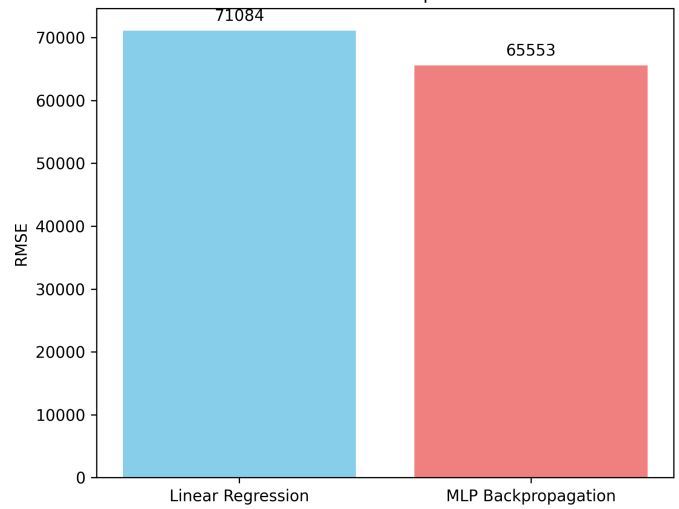
5. VISUALISASI HASIL

Model Comparison: Linear Regression vs MLP Backpropagation

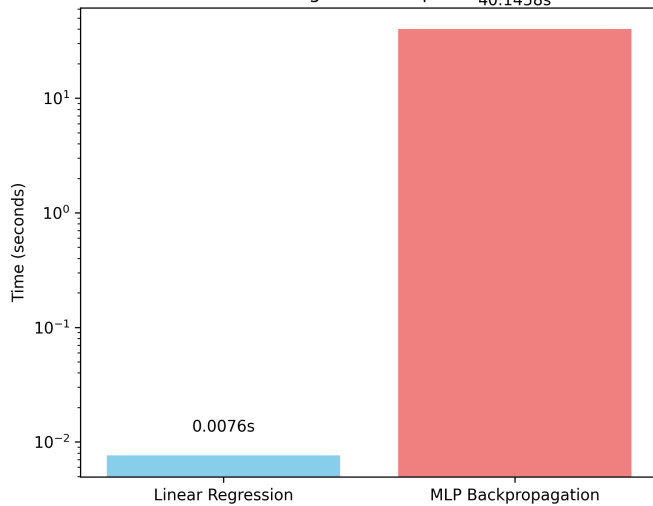
Test R² Score Comparison



Test RMSE Comparison



Training Time Comparison



Test MAE Comparison

