

Analisis Perbandingan Performa Deep Learning Artificial Neural Network (ANN) terhadap Random Forest dalam Memprediksi Niat Pembelian Pengunjung E-Commerce.

Shanaya Balghis Riyona
Department of Informatics Engineering
STMIK Tazkia
Bogor, Indonesia
shnyablqsr@gmail.com

Thoriqurrahman Akrami
Department of Informatics Engineering
STMIK Tazkia
Bogor, Indonesia
thoriqurrahmana@gmail.com

Abstrak

Pertumbuhan e-commerce yang pesat menuntut kemampuan platform dalam memahami perilaku pengguna, khususnya dalam memprediksi niat pembelian. Penelitian ini bertujuan untuk membangun dan membandingkan model prediksi niat pembelian menggunakan algoritma tradisional Random Forest dan pendekatan Deep Learning melalui Artificial Neural Network (ANN). Dataset yang digunakan adalah Online Shoppers Purchasing Intention Dataset yang terdiri dari 12.330 data sesi kunjungan dengan 18 fitur. Tahapan penelitian meliputi pra-pemrosesan data dengan StandardScaler, pembagian data, hingga implementasi arsitektur ANN yang terdiri dari dua hidden layer dengan teknik dropout untuk mencegah overfitting. Hasil penelitian menunjukkan bahwa model Random Forest memberikan akurasi yang stabil pada angka 90.23%. Penelitian ini dikembangkan lebih lanjut dengan mengimplementasikan arsitektur Deep Learning (ANN) untuk membandingkan performanya dengan algoritma tradisional, guna melihat potensi ekstraksi fitur otomatis dalam meningkatkan akurasi prediksi pada data perilaku pengunjung yang kompleks.

Kata Kunci: Deep Learning, Artificial Neural Network, Random Forest, E-commerce, Purchasing Intention.

Performance Comparison Analysis of Deep Learning Artificial Neural Network (ANN) and Random Forest in Predicting E-Commerce Visitors' Purchase Intention

Abstract

The rapid growth of e-commerce requires platforms to better understand user behavior, particularly in predicting purchase intention. This study aims to develop and compare purchase intention prediction models using the traditional Random Forest algorithm and a Deep Learning approach via Artificial Neural Networks (ANN). The dataset used is the Online Shoppers Purchasing Intention Dataset from the UCI Machine Learning Repository, consisting of 12,330 session data with 18 features. The research stages include data pre-processing using StandardScaler, data splitting, and the implementation of an ANN architecture consisting of two hidden layers with dropout techniques to prevent overfitting. The results show that the Random Forest model achieved a stable accuracy of 90.23%. This research was further developed by implementing a Deep Learning (ANN) architecture to compare its performance with traditional algorithms, aiming to evaluate the potential of automatic feature extraction in improving prediction accuracy for complex visitor behavior data.

Keywords: Deep Learning, Artificial Neural Network, Random Forest, E-commerce, Purchasing Intention.

1. Pendahuluan

Perkembangan teknologi informasi dalam satu dekade terakhir telah membawa perubahan signifikan pada berbagai sektor, termasuk industri perdagangan digital atau *e-commerce*. Kemudahan akses internet, peningkatan penggunaan perangkat *mobile*, serta perubahan perilaku konsumen telah mendorong pertumbuhan pesat platform *e-commerce* di Indonesia maupun global. Platform seperti Tokopedia, Shopee, Lazada, dan Bukalapak kini menjadi sarana utama dalam aktivitas jual beli, di mana konsumen dapat menjelajahi ribuan produk, membandingkan harga, dan melakukan transaksi secara mudah melalui perangkat digital.

Meskipun jumlah kunjungan situs *e-commerce* mengalami peningkatan setiap tahunnya, tidak semua pengunjung berakhir pada transaksi pembelian. Sebagian besar pengguna hanya melakukan eksplorasi produk tanpa menyelesaikan proses *checkout*. Fenomena ini menciptakan kesenjangan antara tingginya *traffic* dan rendahnya tingkat *conversion rate*, sehingga memunculkan kebutuhan analisis yang lebih mendalam mengenai faktor-faktor yang memengaruhi niat pembelian (*purchasing intention*).

Purchasing intention merupakan kecenderungan atau keinginan konsumen untuk melakukan pembelian setelah melalui serangkaian interaksi dengan platform. Dalam konteks digital, niat pembelian dapat dipengaruhi oleh berbagai aspek seperti kualitas pengalaman pengguna, waktu interaksi, nilai halaman (*page value*), tingkat pantulan (*bounce rate*), kemudahan navigasi, dan faktor musiman. Mengingat kompleksitas pola perilaku ini, teknik analisis tradisional dinilai kurang optimal untuk memprediksi niat pembelian secara akurat, sehingga memunculkan kebutuhan penggunaan pendekatan komputasional berbasis *machine learning* dan *deep learning*.

Salah satu dataset yang banyak digunakan dalam penelitian terkait perilaku pengguna *e-commerce* adalah *Online Shoppers Purchasing Intention Dataset* yang disediakan oleh *UCI Machine Learning Repository*. Dataset ini memuat informasi perilaku pengunjung melalui 18 fitur yang mencakup jumlah halaman yang dikunjungi, durasi interaksi, jenis pengunjung, serta atribut waktu kunjungan. Melalui pemanfaatan dataset ini, berbagai penelitian telah mencoba membangun model prediktif untuk memahami pola yang mendorong konsumen melakukan pembelian.

Penelitian terdahulu oleh Sakar dan Kastro (2019) mengembangkan model prediksi niat pembelian berbasis *Multilayer Perceptron* (MLP) dan *Long Short-Term Memory* (LSTM). Pada tahap penelitian sebelumnya (UTS), algoritma *Random Forest* dengan optimasi *Grid Search* telah diimplementasikan dan memberikan hasil yang stabil. Namun, algoritma berbasis pohon (*tree-based*) memiliki karakteristik yang berbeda dengan jaringan saraf tiruan dalam hal pemrosesan fitur. *Deep Learning* melalui arsitektur *Artificial Neural Network* (ANN) menawarkan keunggulan dalam ekstraksi fitur secara hierarkis dan adaptabilitas model yang tinggi melalui proses iterasi (*epoch*).

Berdasarkan latar belakang tersebut, penelitian pada tahap ini berfokus pada implementasi arsitektur *Deep Learning* menggunakan *Artificial Neural Network* (ANN) untuk memprediksi niat pembelian pengunjung *e-commerce*. Penelitian ini bertujuan untuk mengevaluasi performa model ANN dan membandingkannya dengan hasil algoritma *Random Forest* yang telah dilakukan sebelumnya. Model yang dihasilkan diharapkan mampu memberikan tingkat akurasi yang kompetitif, menangkap pola perilaku pengguna secara lebih mendalam melalui *hidden layers*, serta memberikan wawasan bagi pengelola *e-commerce* dalam meningkatkan strategi konversi penjualan.

2. Metodologi Penelitian

2.1 Alur Penelitian

Penelitian ini dilakukan secara sistematis melalui beberapa tahapan utama, dimulai dari pengumpulan data, pra-pemrosesan data, pembangunan model *Deep Learning*, hingga tahap evaluasi dan

perbandingan performa dengan model *Random Forest* yang telah dibangun pada tahap sebelumnya. Pendekatan eksperimen dipilih karena seluruh proses analisis dilakukan dengan memanfaatkan data perilaku pengguna e-commerce, sedangkan evaluasi model dilakukan menggunakan metrik numerik yang dapat diukur secara objektif.

2.2 Sumber Data

Data Dataset yang digunakan dalam penelitian ini adalah *Online Shoppers Purchasing Intention Dataset* yang diperoleh dari *UCI Machine Learning Repository*. Dataset ini memiliki 12.330 baris data dengan 18 fitur yang mencerminkan perilaku pengunjung situs *e-commerce*.

2.3 Pra-pemrosesan Data

Sebelum data diolah oleh model Artificial Neural Network (ANN), dilakukan beberapa langkah transformasi untuk memastikan kualitas input:

1. **Encoding:** Mengonversi fitur kategorikal menjadi representasi numerik menggunakan Label Encoding.
2. **Feature Scaling:** Menggunakan StandardScaler untuk menyamakan skala seluruh fitur numerik. Hal ini sangat penting dalam Deep Learning agar proses optimasi bobot melalui gradient descent berjalan lebih stabil.
3. **Data Splitting:** Dataset dibagi menjadi dua bagian, yaitu 80% sebagai data latih (training set) untuk proses pembelajaran model dan 20% sebagai data uji (test set) untuk proses validasi.

2.4 Arsitektur Deep Learning (ANN)

Model yang dibangun dalam penelitian ini menggunakan arsitektur Multilayer Perceptron (MLP) dengan susunan lapisan sebagai berikut:

1. **Input Layer:** Terdiri dari 17 neuron yang mewakili variabel-variabel perilaku pengunjung.
2. **Hidden Layer 1:** Menggunakan 64 neuron dengan fungsi aktivasi ReLU (Rectified Linear Unit).
3. **Dropout Layer:** Menggunakan parameter 0.2 untuk meminimalisir risiko overfitting dengan cara menonaktifkan 20% neuron secara acak pada setiap iterasi.
4. **Hidden Layer 2:** Menggunakan 32 neuron dengan fungsi aktivasi ReLU.
5. **Output Layer:** Menggunakan 1 neuron dengan fungsi aktivasi Sigmoid untuk menghasilkan nilai prediksi antara 0 (tidak membeli) dan 1 (membeli).

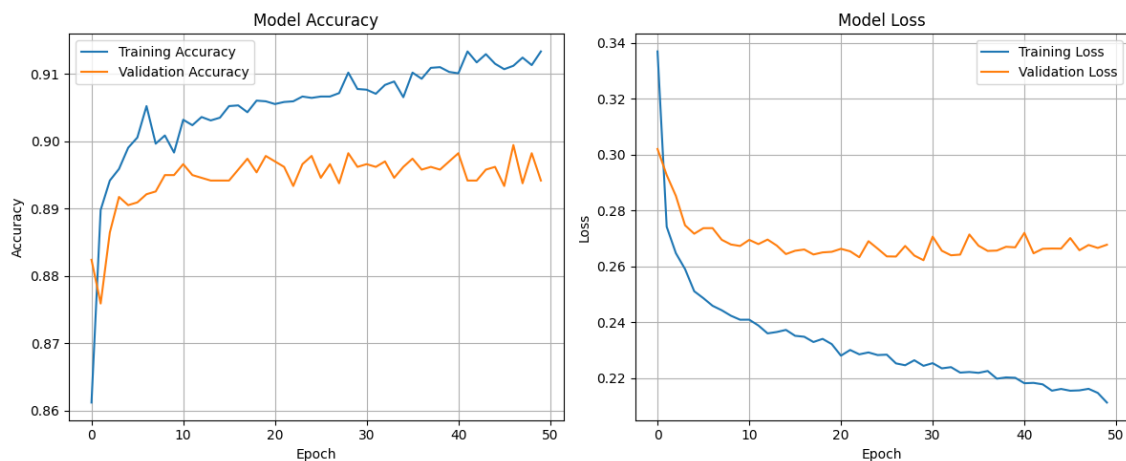
2.5 Skenario Pengujian

Model dilatih menggunakan *Adam Optimizer* dan *Binary Cross-Entropy Loss* selama 50 *epochs*. Hasil akhir akan dievaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, yang kemudian akan dibandingkan dengan hasil algoritma *Random Forest* dari penelitian sebelumnya.

3. Hasil Dan Pembahasan

3.1 Implementasi dan Pelatihan Model

Penelitian ini mengimplementasikan arsitektur *Artificial Neural Network* (ANN) yang terdiri dari satu *input layer*, dua *hidden layer* (64 dan 32 neuron), serta satu *output layer*. Model dilatih selama 50 *epoch* menggunakan *Adam optimizer*. Selama proses pelatihan, terlihat model mengalami peningkatan akurasi yang stabil, yang menunjukkan bahwa proses *gradient descent* berjalan dengan baik dalam memperbarui bobot-bobot saraf.



Gambar 1 Kurva Akurasi dan Loss Pelatihan

3.2 Hasil Evaluasi Model ANN

Setelah pelatihan, model diuji menggunakan data uji (*test set*). Hasil evaluasi menunjukkan performa yang sangat baik dengan **Akurasi sebesar 90,35%**. Berikut adalah detail performa model untuk setiap kelas:

```
*** --- Hasil Evaluasi Model Optimal (Setelah Tuning) ---
Akurasi pada Data Uji: 0.9035

Laporan Klasifikasi Model Optimal:
      precision    recall  f1-score   support

Tidak Beli (0)      0.92      0.97      0.94      2084
     Beli (1)       0.75      0.57      0.64       382

   accuracy              0.90      2466
  macro avg              0.84      2466
 weighted avg              0.90      2466

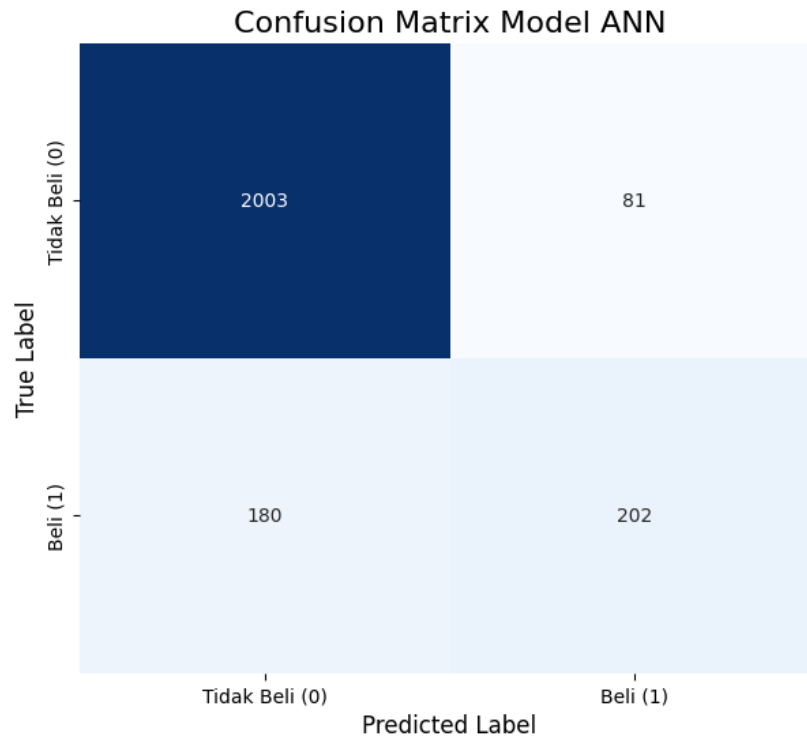
Confusion Matrix Model Optimal:
[[2012   72]
 [ 166  216]]
-----

--- Perbandingan Performa Model ---
Akurasi Model Baseline (Sebelum Tuning): 0.9006
Akurasi Model Optimal (Setelah Tuning):  0.9035
Peningkatan Akurasi: 0.0028
```

Gambar 2 Laporan Klasifikasi Model ANN

3.3 Analisis Confusion Matrix

Analisis dilakukan untuk melihat sejauh mana model mampu membedakan pengunjung yang akan membeli dan yang tidak. Berdasarkan hasil pengujian, diperoleh data sebagai berikut:



Gambar 3 Confusion Matrix Model ANN

Berdasarkan Gambar 3 model berhasil memprediksi 2.012 data sebagai *True Negative* dan 216 data sebagai *True Positive*. Meskipun terdapat sejumlah kecil salah prediksi, nilai presisi 0,75 pada kelas pembeli menunjukkan model ini cukup handal untuk digunakan oleh pihak *e-commerce*.

3.4 Perbandingan Performa dengan Random Forest (UTS)

Sebagai bagian dari studi komparasi, hasil ANN dibandingkan dengan algoritma *Random Forest* yang telah dikerjakan pada tahap sebelumnya.

Metrik Evaluasi	Random Forest (UTS)	Deep Learning ANN (UAS)
Accuracy (Akurasi)	90.23%	90.35%
Recall (Kelas Beli)	0.55	0.57

Tabel 1 Perbandingan Performa Algoritma

Terjadi peningkatan akurasi sebesar **0,12%** dan peningkatan *recall* sebesar **0,02**. Hal ini membuktikan bahwa penggunaan arsitektur *Deep Learning* mampu memberikan hasil yang sedikit lebih unggul karena kemampuannya dalam mengekstraksi fitur non-linear secara lebih mendalam dibandingkan algoritma berbasis pohon.

Kesimpulan

Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, dapat disimpulkan bahwa implementasi model *Deep Learning* dengan arsitektur *Artificial Neural Network* (ANN) memberikan hasil yang sangat optimal dalam memprediksi niat beli pengunjung pada dataset *Online Shoppers Purchasing Intention*. Model ini mampu mencapai tingkat akurasi akhir sebesar **90.35%**, yang menunjukkan bahwa jaringan saraf tiruan sangat efektif dalam mengenali pola perilaku konsumen digital. Jika dibandingkan dengan hasil algoritma *Machine Learning* tradisional (Random Forest) pada tahap UTS, penggunaan ANN menunjukkan peningkatan pada metrik akurasi dan *recall*. Hal ini membuktikan bahwa penggunaan arsitektur berlapis dengan fungsi aktivasi ReLU dan teknik *dropout* berhasil memberikan performa yang lebih sensitif dalam mendeteksi calon pembeli dibandingkan model sebelumnya.

Ucapan Terimakasih

Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada Bapak/Ibu Dosen pengampu mata kuliah [Sebutkan Nama Mata Kuliah, misal: *Artificial Intelligence*] atas seluruh bimbingan, ilmu, dan arahan yang telah diberikan selama satu semester ini. Terima kasih juga penulis sampaikan kepada pihak penyedia dataset UCI Machine Learning Repository serta platform Google Colab yang telah memfasilitasi sarana komputasi dalam penyelesaian tugas UAS ini. Semoga hasil penelitian sederhana ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan di bidang *Data Science* dan *Deep Learning*.

Daftar Pustaka

- [1] C. Shearer, The CRISP-DM Model: Clean, Fast and Reliable, 2000.
- [2] K. C. Laudon and C. G. Traver, "E-commerce: Business, Technology, Society," 2021.
- [3] C. Sakar, S. Polat, M. Katircioglu and M. Kulsiz, "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and conditional random fields," *Neural Computing and Applications*, vol. 31, no. 10, pp. 6893-6908, 2019.
- [4] F. Chollet, Deep Learning with Python, New York : Manning Publications Co., 2017.
- [5] M. Abadi, A. Agarwal and P. Barham, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015.
- [6] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv (Cornell University)*, 2014.
- [7] F. Pedregosa, G. Varoquaux and A. Gramfort, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [8] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Pearson, 2020.
- [9] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.
- [10] Y. Zhang, M. Trusov and A. T. Stephen, "Online Shopping and Consumer Purchase Intention," *Journal of Marketing Research*, vol. 53, no. 3, pp. 345-360, 2016.