

# LAPORAN PROJECT PENERAPAN NEURAL NETWORKS UNTUK ANALISIS DAN PREDIKSI DATA PUBLIK

MATA KULIAH: MACHINE LEARNING (3 SKS)



**Nama Anggota Kelompok:**

1. Yuliana Olo
2. Angela Kaylee Sasha Rasheeda

# **BAB I**

## **PENDAHULUAN**

Perkembangan teknologi informasi dan ketersediaan data publik yang semakin melimpah mendorong kebutuhan akan metode analisis data yang mampu menangani kompleksitas tinggi. *Neural Networks* (NN) merupakan salah satu pendekatan *Machine Learning* yang terinspirasi dari cara kerja neuron biologis pada otak manusia, dan terbukti efektif dalam menyelesaikan berbagai permasalahan klasifikasi, regresi, serta prediksi *time-series*.

Dalam konteks Sistem Informasi, *Neural Networks* memiliki peran penting dalam mendukung pengambilan keputusan berbasis data (*data-driven decision making*). Dengan kemampuan untuk mempelajari pola non-linear dan hubungan kompleks antar variabel, model NN dapat memberikan insight yang lebih akurat dibandingkan metode statistik konvensional.

Tujuan dari project ini adalah menerapkan *Neural Networks* secara *end-to-end*, mulai dari pemahaman dataset, tahap preprocessing, perancangan arsitektur model, proses training dan evaluasi, hingga interpretasi hasil dari sudut pandang Sistem Informasi. Project ini menggunakan empat dataset publik dengan karakteristik berbeda untuk menunjukkan fleksibilitas *Neural Networks* dalam berbagai domain aplikasi.

# **BAB II**

## **DESKRIPSI DATASET**

### **2.1 Online Retail II (UCI)**

Dataset Online Retail II berisi data transaksi penjualan ritel online di Inggris. Atribut utama meliputi *InvoiceNo*, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *UnitPrice*, *CustomerID*, dan *Country*.

Dataset ini digunakan untuk menganalisis perilaku pelanggan dan memprediksi loyalitas atau segmentasi pelanggan. Tantangan utama pada dataset ini adalah adanya missing value, data transaksi negatif (retur), serta kebutuhan feature engineering untuk mengubah data transaksi menjadi representasi perilaku pelanggan.

### **2.2 Heart Failure Clinical Records**

Dataset ini berisi data klinis pasien gagal jantung, dengan target utama adalah variabel DEATH\_EVENT (0 atau 1). Atribut meliputi usia, tekanan darah, kadar serum kreatinin, ejection fraction, anemia, diabetes, dan faktor klinis lainnya.

Dataset ini digunakan untuk tugas klasifikasi biner guna memprediksi risiko kematian pasien. Dataset relatif kecil namun sangat sensitif terhadap kesalahan prediksi, sehingga evaluasi model menjadi aspek penting.

### **2.3 Air Quality Dataset (Beijing Multi-Site)**

Dataset ini merupakan data time-series kualitas udara yang dikumpulkan dari beberapa stasiun pemantauan di Beijing. Variabel utama meliputi PM2.5, PM10, SO2, NO2, CO, O3, serta data cuaca pendukung.

Dataset ini digunakan untuk tugas prediksi kualitas udara, khususnya prediksi nilai PM2.5 di masa depan. Tantangan utama terletak pada missing value, data musiman, serta ketergantungan temporal.

## **2.4 Hotel Booking Demand**

Dataset Hotel Booking Demand berisi data reservasi hotel, baik hotel kota maupun resort. Variabel target adalah is\_canceled yang menunjukkan apakah reservasi dibatalkan atau tidak.

Dataset ini digunakan untuk klasifikasi pembatalan reservasi. Informasi ini sangat penting bagi manajemen hotel dalam perencanaan kapasitas, pendapatan, dan strategi pemasaran.

## **BAB III**

# **METODOLOGI PENELITIAN**

Metodologi yang digunakan dalam project ini mengikuti alur kerja *Machine Learning* secara sistematis, yaitu:

1. Data Understanding: Memahami struktur data, tipe variabel, distribusi data, serta potensi masalah seperti missing value dan outlier.
2. Data Preparation: Meliputi pembersihan data, penanganan missing value, encoding variabel kategorikal, normalisasi/standarisasi, serta pembagian data train dan test dengan rasio 70:30.
3. Desain Model: Menentukan arsitektur *Neural Networks* yang sesuai dengan karakteristik dataset dan jenis tugas (klasifikasi atau regresi).
4. Training Model: Melatih model menggunakan data training dengan parameter tertentu (*epoch, batch size, learning rate*).
5. Evaluasi Model: Mengukur performa model menggunakan metrik yang relevan.
6. Analisis Sistem: Menginterpretasikan hasil model dalam konteks Sistem Informasi.

## **BAB IV**

# **DATA PREPARATION**

Tahap data preparation dilakukan secara berbeda untuk setiap dataset, namun memiliki prinsip umum yang sama.

1. Untuk dataset Online Retail II, dilakukan penghapusan transaksi tanpa CustomerID, penghapusan nilai Quantity negatif, serta agregasi data transaksi per pelanggan.
2. Pada dataset Heart Failure, dilakukan normalisasi fitur numerik menggunakan StandardScaler untuk meningkatkan stabilitas training.
3. Dataset Air Quality memerlukan interpolasi missing value dan pembentukan fitur lag untuk menangkap ketergantungan temporal.
4. Dataset Hotel Booking Demand memerlukan encoding variabel kategorikal menggunakan one-hot encoding serta normalisasi fitur numerik.

# **BAB V**

## **DESAIN DAN ARSITEKTUR MODEL NEURAL NETWORKS**

Model Neural Networks yang digunakan merupakan Feed Forward Neural Network (Multilayer Perceptron) dengan struktur umum:

- Input Layer: Menyesuaikan jumlah fitur
- Hidden Layer 1–2: Menggunakan aktivasi ReLU
- Output Layer:
- Sigmoid untuk klasifikasi biner
- Linear untuk regresi

Optimizer yang digunakan adalah Ada karena kemampuannya dalam menyesuaikan learning rate secara adaptif.

# **BAB VI**

## **PROSES TRAINING DAN EVALUASI MODEL**

### **6.1 Training Model**

Training dilakukan dengan pembagian data 70% untuk training dan 30% untuk testing. Model dilatih selama 50–100 epoch dengan batch size 32.

Untuk mencegah overfitting, digunakan teknik early stopping dan dropout pada hidden layer.

### **6.2 Evaluasi Model**

Evaluasi dilakukan menggunakan metrik yang sesuai dengan jenis tugas:

- Klasifikasi: Accuracy, Precision, Recall, F1-Score, Confusion Matrix
- Regresi: Mean Squared Error (MSE) dan Mean Absolute Error (MAE)

Hasil evaluasi menunjukkan bahwa model Neural Networks mampu mencapai performa yang baik pada keempat dataset, dengan akurasi tinggi pada dataset Heart Failure dan Hotel Booking Demand.



## **BAB VII**

# **ANALISIS SISTEM INFORMASI DAN KETERBATASAN PENELITIAN**

### **7.1 Analisis Sistem Informasi**

Dari perspektif Sistem Informasi, hasil prediksi model dapat diintegrasikan ke dalam sistem pendukung keputusan.

1. dataset Hotel Booking Demand, sistem dapat memberikan peringatan dini terhadap potensi pembatalan reservasi.
2. dataset Heart Failure, sistem dapat membantu tenaga medis dalam mengidentifikasi pasien berisiko tinggi.
3. dataset Retail, sistem dapat digunakan untuk strategi loyalitas pelanggan.
4. dataset Air Quality, sistem dapat memberikan informasi prediksi kualitas udara kepada masyarakat.

### **7.2 Keterbatasan Penelitian**

Beberapa keterbatasan dalam project ini antara lain keterbatasan ukuran dataset, asumsi independensi data, serta penggunaan arsitektur NN yang relatif sederhana.

## **KESIMPULAN**

Berdasarkan hasil project, dapat disimpulkan bahwa Neural Networks merupakan metode yang efektif untuk analisis dan prediksi berbasis data publik. Mampu mempelajari pola kompleks dan memberikan hasil yang relevan untuk mendukung pengambilan keputusan dalam Sistem Informasi.

Pengembangan dapat dilakukan dengan menggunakan arsitektur yang lebih kompleks seperti LSTM untuk data time-series atau mengombinasikan Neural Networks dengan metode lain.