

EXPLORATORY DATA ANALYSIS UNTUK MACHINE LEARNING

Final Assignment UAS

Nama : _____

NIM : _____

Course : Exploratory Data Analysis for Machine Learning (IBM – Coursera)

1. Pendahuluan

Exploratory Data Analysis (EDA) merupakan tahapan penting dalam proses data science untuk memahami karakteristik data sebelum digunakan dalam pemodelan machine learning. Pada tugas akhir ini, dilakukan analisis eksploratif terhadap dataset customer churn untuk mengidentifikasi pola, hubungan antar variabel, serta faktor-faktor yang memengaruhi pelanggan berhenti berlangganan.

2. Dataset Summary

Dataset yang digunakan adalah Customer Churn Dataset yang berisi data pelanggan sebuah perusahaan layanan. Dataset ini terdiri dari 300 baris data dan 5 variabel utama, yaitu tenure, MonthlyCharges, TotalCharges, Contract, dan Churn. Variabel target dalam analisis ini adalah Churn.

3. Exploratory Data Analysis

Hasil analisis statistik deskriptif menunjukkan bahwa rata-rata masa berlangganan pelanggan berkisar pada nilai menengah. Visualisasi histogram menunjukkan bahwa distribusi biaya bulanan cenderung menyebar dengan beberapa nilai ekstrem. Boxplot antara tenure dan Churn menunjukkan bahwa pelanggan yang melakukan churn cenderung memiliki masa berlangganan yang lebih pendek.

4. Data Cleaning dan Feature Engineering

Tahapan data cleaning dilakukan untuk memastikan kualitas data. Dataset tidak memiliki missing value yang signifikan. Selain itu, dilakukan pemeriksaan outlier menggunakan metode Interquartile Range (IQR). Feature engineering dilakukan dengan membuat variabel tambahan yang dapat membantu analisis lanjutan.

5. Hipotesis dan Pengujian

Hipotesis utama dalam penelitian ini adalah masa berlangganan pelanggan berpengaruh signifikan terhadap churn. Pengujian dilakukan dengan membandingkan rata-rata tenure antara pelanggan churn dan non-churn. Hasil pengujian menunjukkan perbedaan yang signifikan, sehingga hipotesis diterima.

6. Kesimpulan dan Rencana Lanjutan

Berdasarkan hasil Exploratory Data Analysis, dapat disimpulkan bahwa tenure dan jenis kontrak merupakan faktor penting yang memengaruhi churn pelanggan. Dataset telah dibersihkan dan siap digunakan untuk tahap pemodelan machine learning. Langkah selanjutnya adalah membangun model klasifikasi untuk memprediksi churn pelanggan.