

NAMA : PUTRI ARFINA

KELAS : 5 AI-A

NIM : 105841116123

1. Pendahuluan

Laporan ini disusun sebagai kelanjutan dari tahap Data Preprocessing dataset Pinjaman Nasabah. Fokus utama dari laporan ini adalah melaksanakan tahap Feature Engineering (Rekayasa Fitur) untuk meningkatkan daya prediksi dataset. Tahapan yang dilakukan meliputi penciptaan fitur baru (rasio pinjaman/pendapatan), penanganan nilai ekstrem (outlier), serta transformasi dan penskalaan akhir yang menghasilkan dataset yang sepenuhnya siap untuk Pemodelan Risiko Kredit.

2. Metodologi inti

Berikut adalah rangkuman langkah-langkah utama yang telah berhasil dilaksanakan:

- Penanganan Nilai Hilang: Imputasi menggunakan Modus (kategorikal) dan Rata-rata (numerik).
- Feature Engineering: Penciptaan fitur Rasio_Pinjaman_Pendapatan.
- Penanganan Outlier: Koreksi nilai ekstrem pada JumlahPinjaman menggunakan Winsorization ($1.5 \times 1QR$).
- Transformasi Fitur: Konversi kategorikal dengan One-Hot Encoding.
- Penskalaan: Normalisasi fitur numerik menggunakan Min-Max Scaler ke rentang $[0, 1]$.
- Pembagian Data: Dataset dibagi menjadi data latih dan data uji (80:20).

3. Bukti tahapan dan penjelasan

a. Feature creation :

```
35 print("\n--- Melakukan Feature Engineering ---")
36 df['Total_Income'] = df['IncomeNasabah'] + df['IncomePasangan']
37 df['Total_Income'] = df['Total_Income'].replace(0, np.nan)
38 df['Total_Income'] = df['Total_Income'].fillna(df['Total_Income'].mean())
39 df['Rasio_Pinjaman_Pendapatan'] = df['JumlahPinjaman'] / df['Total_Income']
40 print("Fitur Rasio_Pinjaman_Pendapatan telah dibuat.")
```

Pembuatan fitur Rasio_Pinjaman_Pendapatan dari total pendapatan rumah tangga dan jumlah pinjaman.

b. Outlier Handling

```
print("\n--- Menangani Outlier (Winsorization) ---")
outlier_cols = ['JumlahPinjaman']

for col in outlier_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_limit = Q1 - 1.5 * IQR
    upper_limit = Q3 + 1.5 * IQR
    df[col] = df[col].clip(lower=lower_limit, upper=upper_limit)
    print(f"Outlier pada kolom {col} dikoreksi.")
```

Implementasi kode Winsorization (Clipping) berdasarkan batas $1.5 \times \text{IQR}$ pada kolom JumlahPinjaman.

c. Transformation & Scaling

```
53
54 print("\n--- Melakukan One-Hot Encoding ---")
55 cols_to_encode = ['JenisKelamin', 'StatusPernikahan', 'Wiraswasta',
56                  'WilayahTempatTinggal', 'JumTanggungan', 'Pendidikan']
57
58 df = df.drop(['ID_Nasabah', 'IncomeNasabah', 'IncomePasangan'], axis=1, errors='ignore')
59 df = pd.get_dummies(df, columns=cols_to_encode, drop_first=True)
60 print("One-Hot Encoding selesai. Jumlah kolom saat ini:", df.shape[1])
61
62 print("\n--- Melakukan Min-Max Scaling ---")
63 scaler = MinMaxScaler()
64 scale_cols = df.drop(columns=['StatusPinjaman'], errors='ignore').select_dtypes(include=np.number).columns.tolist()
65 df[scale_cols] = scaler.fit_transform(df[scale_cols])
66 print("Min-Max Scaling selesai.")
```

Implementasi One-Hot Encoding untuk data kategori dan Min-Max Scaler untuk menormalisasi semua fitur numerik akhir.