

✓ **Diabetes-Balanced.csv** — the final, cleaned, transformed, and balanced dataset.

Feature Name	Description	Type	Example	Notes
Pregnancies	Number of times pregnant	Numeric	3	Original
Glucose	Plasma glucose concentration	Numeric	120	Replaced 0 with NaN → imputed
BloodPressure	Diastolic blood pressure (mm Hg)	Numeric	70	Cleaned, imputed
SkinThickness	Triceps skin fold thickness (mm)	Numeric	25	Cleaned, imputed
Insulin	2-Hour serum insulin (mu U/ml)	Numeric	85	Cleaned, imputed
BMI	Body Mass Index	Numeric	27.5	Categorized: Underweight, Normal, Overweight, Obese
DiabetesPedigreeFunction	Genetic diabetes risk	Numeric	0.45	Kept as numeric
Age	Age in years	Numeric	33	Transformed into Age Groups
AgeGroup	Age category (young/middle/senior)	Categorical	Middle-aged	Engineered feature
BMI_Category	BMI-based category	Categorical	Normal	Engineered feature
Glucose_Level	Glucose-based category	Categorical	High	Engineered feature
Class	Target: 0 = Non-diabetic, 1 = Diabetic	Binary	0/1	Balanced using SMOTE+ENN