Supplementary Table 1 HDI-MF-Gower

HDI-MF-Gower

Input: Dataset D, numerical features N, categorical features F, threshold ε, max iterations P

Output: Imputed dataset D^imp

Phase 1: Initialization

1. Identify complete samples C and incomplete samples I

2. Calculate ranges: R_k ← max(D[:,k]) - min(D[:,k]) for k ∈ N if R_k = 0 then R_k ← 1 end if

3. Calculate and normalize weights:

w_k^{raw} ← Var(D[:,k]) × (1-MR_k) for k ∈ N w_k^{raw} ← H(D[:,k]) × (1-MR_k) for k ∈ F

w_k ← normalize(log(1 + w_k^{raw})) to [0.1, 2.0] for k ∈ (N ∪ F)

4. Initial imputation using adaptive Gower distance:

for each i ∈ I do

j* ← argmin_{j∈C} (Σ_k w_k d_k) / (Σ_k w_k) where:

d_k = |x_i[k]-y_j[k]| / R_k for k ∈ N (if both observed)

d_k = I(x_i[k] ≠ y_j[k]) for k ∈ F (if both observed)

D^imp[i,k] ← D[j*,k] for missing values

end for

Phase 2: Iterative Refinement

1. Sort variables by missing rate in ascending order, p ← 0

2. while p < P do

D^imp_old ← D^imp

for each variable k with missing values do

Train RF_k on observed data, predict missing values

end for

Calculate: Δ_N and Δ_F (see formulas (3) and (4))

if Δ_N < ε and Δ_F < ε then break

p ← p + 1

end while

3. return D^imp