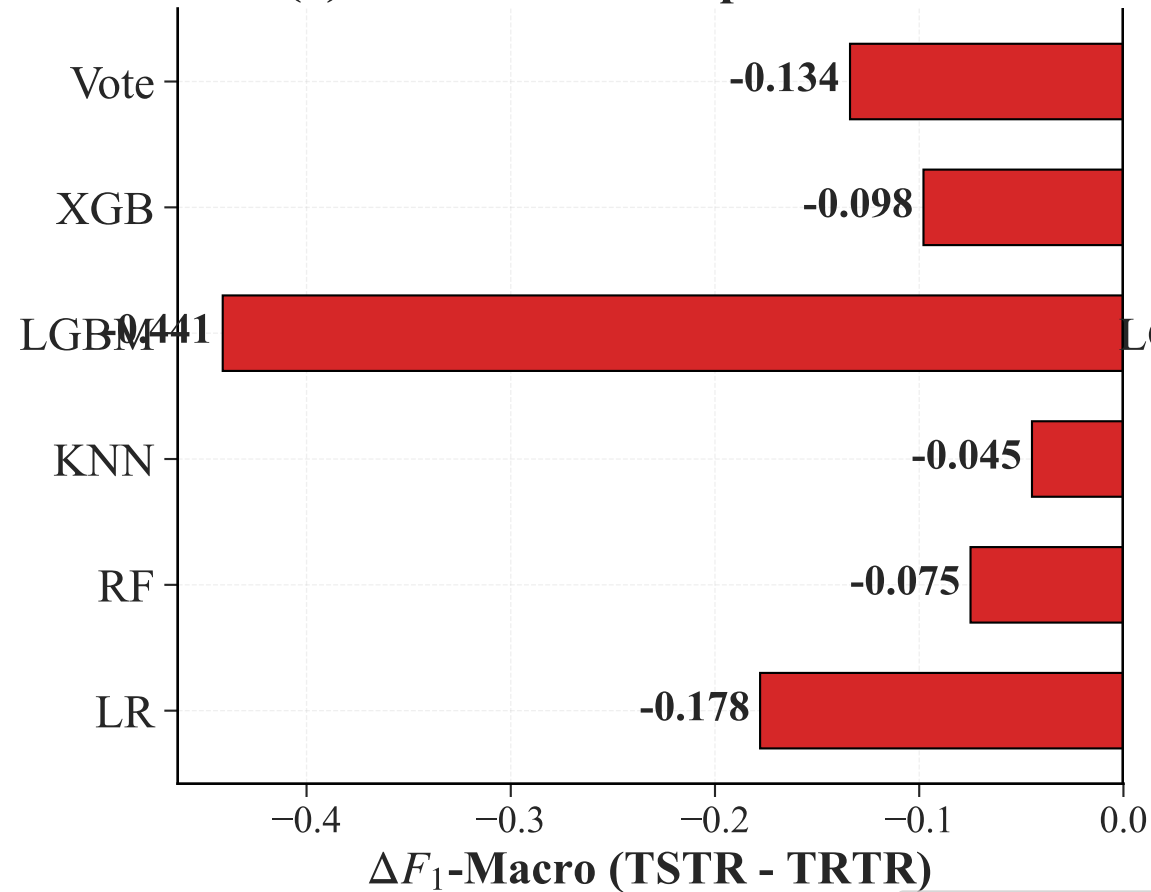**(a) Performance Gap - Multiclass**

(b) Performance Gap - Binary

Vote: −0.134 / −0.001
XGB: −0.098 / −0.004
LGBM: −0.441 / −0.043
KNN: −0.045 / +0.001
RF: −0.075 / −0.002
LR: −0.178 / +0.030

ΔF_1-Macro (TSTR - TRTR)

Legend: TSTR better (green) / TRTR better (red)