

std. dev. of F1 macro values (avg. over datasets)

