

Overconfidence vs. Macro-F1 Before and After Re-Evaluation

