

Score vs Confidence Threshold: MSP, mmlu

