

Score vs Confidence Threshold: test, winogrande

