

Score vs Confidence Threshold: train, piqa

