

Results for submission **qnli_ensemble_2**

Score: **60.9**

Task	Metric	Score
The Corpus of Linguistic Acceptability	Matthew's Corr	0.0
The Stanford Sentiment Treebank	Accuracy	80.0
Microsoft Research Paraphrase Corpus	F1 / Accuracy	81.5/73.4
Semantic Textual Similarity Benchmark	Pearson-Spearman Corr	61.2/58.7
Quora Question Pairs	F1 / Accuracy	51.4/79.1
MultiNLI Matched	Accuracy	56.0
MultiNLI Mismatched	Accuracy	56.4
Question NLI	Accuracy	92.8
Recognizing Textual Entailment	Accuracy	54.1
Winograd NLI	Accuracy	62.3
Diagnostics Main	Matthew's Corr	9.2