

سوالات داخل نوتبوک جواب داده شده‌اند و در اینجا صرفاً نتایج آزمایشات پارت دوم و سوم آورده شده‌است.

Part 2:

	EM(%)	F1(%)
ORG	16.9	36.4
Answer Absence	0.5	2.4
Answer Absence (gt='Not enough info.')	76.2	79.8
Entity Substitution(original answer)	7.0	20.6
Entity Substitution (Modified answer)	9.0	26.1
Nonsense Word Substitution	7.9	30

Part 3:

	Accuracy	F1(Negative) - F1(Positive)	Macro avg f1
Zero shot	50	67 - 2	34
Few-shot	50	66 - 7	37
Few-shot (different example)	51	67 - 7	36
few-shot(p1p2n)	78	80 - 76	78
few-shot(p1np2)	76	72 - 79	76
few-shot(p2p1n)	82	81 - 83	82
few-shot(p2np1)	82	83 - 81	82
few-shot(np1p2)	75	79 - 70	75
few-shot(np2p1)	63	73 - 43	58
CC	69	76 - 58	67
DC	76	79 - 72	75

	ECE
CC	0.0539
DC	0.0935