Counter-Speech	0	5	0	0
l Label Neutral	27	0	0	1
Human Label Refusal Neu	6	0	0	0
Hate Speech	0	0	0	0
	Counter-Speech	Refusal rediction	Hate Speech	