

Testing Multiple Iterations - Summary Stats

Zero shot learning

Table 1: Performance of model in zero-shot learning across 100 classifications of each comment at a temperature of 0.3

	Metric	Mean (%)	Standard Error (%)
Racism	Accuracy	58.0366	6.455005
	Precision	58.0318	6.680448
	Recall	59.4751	9.159564
	F1	58.4805	6.730550
Sexism	Accuracy	54.5494	5.220326
	Precision	52.7399	3.665168
	Recall	78.8363	6.896958
	F1	63.1200	4.335575
Overall	Accuracy	56.3002	4.271000
	Precision	54.8434	3.498226
	Recall	69.0706	5.873358
	F1	69.7670	5.688525

One shot learning

Table 2: Performance of model in one-shot learning across 100 classifications of each comment at a temperature of 0.3

	Metric	Mean (%)	Standard Error (%)
Racism	Accuracy	61.5518	1.0659158
	Precision	58.4770	0.8275515
	Recall	79.7336	1.2232912
	F1	67.4684	0.8614607
Sexism	Accuracy	60.1363	1.3521002
	Precision	57.2348	1.1099736
	Recall	74.8251	1.7333488
	F1	64.8506	1.1535565
Overall	Accuracy	60.8446	0.8622480
	Precision	57.8762	0.7008967
	Recall	77.3242	1.0172990
	F1	64.0590	1.2788169