

SUMMARY OF FINDINGS REGARDING EXPERIMENTS WITH OM DATA AND RNN LANGUAGE MODEL.

SL	Experiment Name	Notebook & Dataset URL	Samples Used	Result	
01	Exp 1 Jan23 - Regular Dataset - Tokenization 1 6 OM Combined	Notebook: https://tinyurl.com/6om-exp1-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Unseen to the Model) Inconsistent Result	First 40 predicted as P and after that incomplete statements. No labels were appended after that	
		Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/dataset1-6OM-Reg-auth			
02	Exp 2 Jan23 - Regular Dataset - Tokenization 2 6 OM Combined	Notebook: https://tinyurl.com/6om-exp2-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Unseen to the Model)	Precision 100%	Recall 6.6%
		Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/dataset1-6OM-Reg-auth		F1 12%	Accuracy 86%

03	Exp 3 Jan23 - Regular Dataset - Tokenization 3 – 6 OM combined	Notebook: https://tinyurl.com/6om-exp3-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Unseen to the Model)	Precision 2.5%	Recall 13%
		Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/dataset1-6OM-Reg-auth		F1 2.09%	Accuracy 85%
04	Exp 4 Jan23 - Regular Dataset - Tokenization 4 6 OM combined	Notebook: https://tinyurl.com/6om-exp4-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Unseen to the Model)	Precision 13%	Recall 20%
		Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/dataset1-6OM-Reg-auth		F1 15%	Accuracy 68%

05	Exp 5 Jan23 - Regular Dataset - Tokenization 5	Notebook: https://tinyurl.com/6om-exp5-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Unseen to the Model)	Precision 10%	Recall 6.6%
	6 OM combined	Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/dataset1-6OM-Reg-auth		F1 7.9%	Accuracy 77%
7 OM combined dataset experiments					
06	Exp 6 Jan23 - Regular Dataset - Tokenization 1	Notebook: https://tinyurl.com/7-om-exp6-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Existing data / seen data)	Incomplete Statements	
	7 OM combined	Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/7-om-reg-auth			

07	Exp 7 Jan23 - Regular Dataset - Tokenization 2 –	Notebook: https://tinyurl.com/7-om-exp7-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Existing data / seen data)	Precision 27%	Recall 60%
	7 OM combined	Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/7-om-reg-auth		F1 19%	Accuracy 70%
08	Exp 8 Jan23 - Regular Dataset - Tokenization 3	Notebook: https://tinyurl.com/7-om-exp8-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 86 (Existing data / seen data)	Precision 32%	Recall 46%
	7 OM combined	Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/7-om-reg-auth		F1 38%	Accuracy 77%

09	Exp 9 Jan23 - Regular Dataset - Tokenization 4 7 OM combined	Notebook: https://tinyurl.com/7-om-exp9-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 31 (Existing data / seen data)	Precision 0%	Recall 0%
		Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/7-om-reg-auth		F1 0%	Accuracy 85%
10	Exp 6 Jan23 - Regular Dataset - Tokenization 5 7 OM combined	Notebook: https://tinyurl.com/7-om-exp10-reg-auth	Total: 101 P Labeled : 15 NP Labeled: 31 (Existing data / seen data)	Precision 100%	Recall 20%
		Sample set: https://tinyurl.com/sampleset1-reg Dataset: https://tinyurl.com/7-om-reg-auth		F1 33%	Accuracy 88%

RESULTS COMPARISON BETWEEN 6 OM AND 7 OM DATASETS BASED EXPERIMENTS.

6 OM Combined (Unseen data)	6 OM based results		7 OM Combined (Trained on Existing)	7 OM based results	
Exp 1 Jan23 - Regular Dataset - Tokenization 1	Incomplete Statements		Exp 6 Jan23 - Regular Dataset - Tokenization 1	Incomplete Statements	
Exp 2 Jan23 - Regular Dataset - Tokenization 2	Precision 100%	Recall 6.6%	Exp 7 Jan23 - Regular Dataset - Tokenization 2	Precision 27% (Reduction)	Recall 60% (Increase)
	F1 12%	Accuracy 86%		F1 19% (Increase)	Accuracy 70% (Increase)
Exp 3 Jan23 - Regular Dataset - Tokenization 3	Precision 2.5%	Recall 13%	Exp 8 Jan23 - Regular Dataset - Tokenization 3	Precision 32% (Increase)	Recall 46% (Increase)
	F1 2.09%	Accuracy 85%		F1 38% (Increase)	Accuracy 77% (Reduction)
Exp 4 Jan23 - Regular Dataset - Tokenization 4	Precision 13%	Recall 20%	Exp 9 Jan23 - Regular Dataset - Tokenization 4	Precision 0% (Reduction)	Recall 0% (Reduction)
	F1 15%	Accuracy 68%		F1 0% (Reduction)	Accuracy 85% (Increase)
Exp 5 Jan23 - Regular Dataset - Tokenization 5	Precision 10%	Recall 6.6%	Exp 10 Jan23 - Regular Dataset - Tokenization 5	Precision 100% (Increase)	Recall 20% (Increase)
	F1 7.9%	Accuracy 77%		F1 33% (Increase)	Accuracy 88% (Increase)

