**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: 100  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: 100  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: 100  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  6 OM combined | Notebook:  <https://tinyurl.com/6om-exp5-reg-auth> | Total: 100  P Labeled : 15  NP Labeled: 86  (Unseen to the Model) | Precision  10% | Recall  6.6% |
| 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  7 OM combined | Notebook:  <https://tinyurl.com/7-om-exp6-reg-auth> | Total: 100  P Labeled : 15  NP Labeled: 86  (Existing data / seen data) | Incomplete Statements | |
| Sample set:  <https://tinyurl.com/sampleset1-reg>  Dataset:  <https://tinyurl.com/7-om-reg-auth> |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 07 | Exp 7 Jan23 - Regular Dataset - Tokenization 2 –  7 OM combined | Notebook: | Total: 100  P Labeled : 15  NP Labeled: 86  (Existing data / seen data) | Precision  27% | Recall  60% |
| 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  7 OM combined | Notebook:  <https://tinyurl.com/expr8-reg-jan23> | Total: 100  P Labeled : 15  NP Labeled: 31  (Existing data / seen data) | Precision  87% | Recall  67% |
| Sample set:  <https://tinyurl.com/expr1-reg-jan23> | F1  76% | Accuracy  71% |
| 09 | Exp 9 Jan23 - Regular Dataset - Tokenization 4  7 OM combined | Notebook:  <https://tinyurl.com/expr9-reg-jan23> | Total: 100  P Labeled : 15  NP Labeled: 31  (Existing data / seen data) | Precision  75% | Recall  71% |
| Sample set:  <https://tinyurl.com/expr1-reg-jan23> | F1  73% | Accuracy  64% |
| 10 | Exp 6 Jan23 - Regular Dataset - Tokenization 5  7 OM combined | Notebook:  https://tinyurl.com/expr10-reg-jan23 | Total: 100  P Labeled : 15  NP Labeled: 31  (Existing data / seen data) | Precision  70% | Recall  87% |
| Sample set:  Sample set:  <https://tinyurl.com/expr1-reg-jan23> | F1  77% | Accuracy  66% |

**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** | | **Results comparison** | |
| Exp 1 Jan23 - Regular Dataset - Tokenization 1 | Precision  65% | Recall  66% | Exp 6 Jan23 - Regular Dataset - Tokenization 1 | Incomplete Statements | | Incomplete Statements | |
| F1  65% | Accuracy  52% |
| Exp 2 Jan23 - Regular Dataset - Tokenization 2 | Precision  69% | Recall  63% | Exp 7 Jan23 - Regular Dataset - Tokenization 2 | Precision  76% | Recall  93% | Precision  76 – 69 = 7%  (increase) | Recall  93 – 63 = 30%  (increase) |
| F1  64% | Accuracy  55% | F1  83% | Accuracy  74% | F1(increase)  83 – 64 = 19% | Accuracy  74 – 55 = 19%  (increase) |
| Exp 3 Jan23 - Regular Dataset - Tokenization 3 | Precision  70% | Recall  78% | Exp 8 Jan23 - Regular Dataset - Tokenization 3 | Precision  87% | Recall  67% | Precision  87 – 70 = 17%  (increase) | Recall  93 – 63 = 30%  (increase) |
| F1  74% | Accuracy  72% | F1  76% | Accuracy  71% | F1  76 – 74 = 2%  (increase) | **Accuracy**  **72 – 71 = 1%**  (decrease) |
| Exp 4 Jan23 - Regular Dataset - Tokenization 4 | Precision  0% | Recall  0% | Exp 9 Jan23 - Regular Dataset - Tokenization 4 | Precision  75% | Recall  71% | Precision  75 – 0 = 75%  (increase) | Recall  71 – 0 = 71%  (increase) |
| F1  0% | Accuracy  30% | F1  73% | Accuracy  64% | F1  73 – 0 = 73%  (increase) | Accuracy  64 – 30 = 34%  (increase) |
| Exp 5 Jan23 - Regular Dataset - Tokenization 5 | Precision  80% | Recall  80% | Exp 10 Jan23 - Regular Dataset - Tokenization 5 | Precision  70% | Recall  87% | **Precision**  **80 – 70 = 10**  (decrease) | Recall  87 – 80 = 7%  (increase) |
| F1  78% | Accuracy  72% | F1  77% | Accuracy  66% | **F1**  **77 – 78 = 1%**  (decrease) | Accuracy  66 – 72 = 6%  (increase) |