**Experiment report for Learning-based**

**Design Tradeoff Selection**

Experiment 1 – Original Distribution

Total OM: 14

Train OM : 14

Test OM 1 : 1 (Ecommerce)

Dataset orientation: No space between words and symbols.

Type: Test set is seen by the model.

Training Set:

Total Dataset Instances: 30970

Pareto Optimal (P) instances: 399

Pareto Optimal (NP) instances: 30571

Testing Set:

Total Dataset Instances: 105

Pareto Optimal (P) instances: 25

Pareto Optimal (NP) instances: 80

Notebook URL: [Notebook-1](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/14-OM-seen/14_OM_Train_1_OM_Test(Ecommerce).ipynb)

Dataset URL: [Dataset-1](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/14-OM-seen/14-OM.xlsx)

Testset URL: [Testset-1](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/14-OM-seen/ecom.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.76 | 1.00 | 0.86 | 80 |
| 1 (P) | 0.00 | 0.00 | 0.00 | 25 |
| Accuracy |  | | 0.76 | 105 |
| Macro avg | 0.38 | 0.50 | 0.43 | 105 |
| Weighted avg | 0.58 | 0.76 | 0.66 | 105 |

Experiment 2 – Original Distribution

Total OM: 14

Train OM : 13

Test OM 1 : 1 (University)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 389 P

Pareto Optimal (NP) instances: 30551 NP

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-2](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/13-Om-train-1-OM(university)-test/13_OM_Train_1_OM_Test(University).ipynb)

Dataset URL: [Dataset-2](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/13-Om-train-1-OM(university)-test/University_OM_test.xlsx)

Testset URL: [Testset-2](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/13-Om-train-1-OM(university)-test/University_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| unknown or out-of-vocabulary (OOV) word in a text sequence.  All responses returned as UNK | | | | |

Experiment 3 – Original Distribution

Total OM: 14

Train OM : 12

Test OM 1 : 2 - University (Unseen) + Decider (Unseen)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30829

Pareto Optimal (P) instances: 377

Pareto Optimal (NP) instances: 30452

Testing Set:

Total instances - 111

Pareto Optimal (P) instances: 25

Pareto Optimal (NP) instances: 80

Notebook URL: [Notebook-3](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/12-OM-train-3-OM(decider)-test/12_OM_Train_2_OM_Test(University%2BDecider).ipynb)

Dataset URL: [Dataset-3](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/12-OM-train-3-OM(decider)-test/12-OM.xlsx)

Testset URL: [Testset-3](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/12-OM-train-3-OM(decider)-test/2-OM-test-set.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.80 | 1.00 | 0.89 | 89 |
| 1 (P) | 0.00 | 0.00 | 0.00 | 22 |
| Accuracy |  | | 0.80 | 111 |
| Macro avg | 0.40 | 0.50 | 0.45 | 111 |
| Weighted avg | 0.64 | 0.80 | 0.71 | 111 |

Experiment 4 – Original Distribution

Total OM: 14

Train OM : 10

Test OM : 4 - University (Unseen) + Decider (Unseen) + CO (Unseen) + OS (Unseen)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30701

Pareto Optimal (P) instances: 367

Pareto Optimal (NP) instances: 30334

Testing Set:

Total Dataset Instances: 127

Pareto Optimal (P) instances: 32

Pareto Optimal (NP) instances: 95

Notebook URL: [Notebook-4](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/10-OM-train-4-OM-test/10_OM_Train_4_OM_Test(University%2BDecider%2BCO%2BOS).ipynb)

Dataset URL: [Dataset-4](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/10-OM-train-4-OM-test/10-OM.xlsx)

Testset URL: [Testset-4](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/unseen/10-OM-train-4-OM-test/4_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.75 | 1.00 | 0.86 | 95 |
| 1 (P) | 0.00 | 0.00 | 0.00 | 32 |
| Accuracy |  | | 0.75 | 127 |
| Macro avg | 0.37 | 0.50 | 0.43 | 127 |
| Weighted avg | 0.56 | 0.75 | 0.64 | 127 |

Experiment 5 – Sampling Distribution

P (30%) – NP (70%)

Total OM: 14

Train OM : 13

Test OM : 1 (University OM)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 9282

Pareto Optimal (NP) instances: 21658

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-5](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/13_OM_30/13_OM_Train_1_OM_Test(University)_30%20(1).ipynb)

Dataset URL: [Dataset-5](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/13_OM_30/13_OM_30.xlsx)

Testset URL: [Testset-5](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/13_OM_30/University_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| unknown or out-of-vocabulary (OOV) word in a text sequence.  All responses returned as UNK | | | | |

Experiment 6 – Sampling Distribution

P (30%) – NP (70%)

Total OM: 14

Train OM : 12

Test OM : 2 (University OM + Decider)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30829

Pareto Optimal (P) instances: 9249

Pareto Optimal (NP) instances: 21580

Testing Set:

Total instances - 111

Pareto Optimal (P) instances: 25

Pareto Optimal (NP) instances: 80

Notebook URL: [Notebook-6](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/12_OM_30/Copy_of_10_OM_Train_4_OM_Test(University%2BDecider%2BCO%2BOS).ipynb)

Dataset URL: [Dataset-6](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/12_OM_30/12_OM_30.xlsx)

Testset URL: [Testset-6](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/12_OM_30/2-OM-test-set.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.82 | 0.79 | 0.80 | 89 |
| 1 (P) | 0.27 | 0.32 | 0.29 | 22 |
| Accuracy |  | | 0.69 | 111 |
| Macro avg | 0.55 | 0.55 | 0.55 | 111 |
| Weighted avg | 0.71 | 0.69 | 0.70 | 111 |

Experiment 7 – Sampling Distribution

P (30%) – NP (70%)

Total OM: 14

Train OM : 10

Test OM : 4 (University OM + Decider + CO + OS)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30701

Pareto Optimal (P) instances: 9210

Pareto Optimal (NP) instances: 21491

Testing Set:

Total Dataset Instances: 127

Pareto Optimal (P) instances: 32

Pareto Optimal (NP) instances: 95

Notebook URL: [Notebook-7](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/10_OM_30/10_OM_Train_4_OM_Test(University%2BDecider%2BCO%2BOS).ipynb)

Dataset URL: [Dataset-7](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/10_OM_30/4_OM_test.xlsx)

Testset URL: [Testset-7](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_30%25_total_dataset/10_OM_30/4_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.74 | 0.77 | 0.76 | 95 |
| 1 (P) | 0.24 | 0.22 | 0.23 | 32 |
| Accuracy |  | | 0.63 | 127 |
| Macro avg | 0.49 | 0.49 | 0.49 | 127 |
| Weighted avg | 0.62 | 0.63 | 0.62 | 127 |

Experiment 8 – Sampling Distribution

P (50%) – NP (50%)

Total OM: 14

Train OM : 13

Test OM : 1 (University OM)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 15470

Pareto Optimal (NP) instances: 15470

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-8](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/13_OM_50/13_OM_Train_1_OM_Test(University)%20(1).ipynb)

Dataset URL: [Dataset-8](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/13_OM_50/13_OM_50.xlsx)

Testset URL: [Testset-8](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/13_OM_50/University_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.73 | 0.76 | 0.74 | 21 |
| 1 (P) | 0.44 | 0.40 | 0.42 | 10 |
| Accuracy |  | | 0.65 | 31 |
| Macro avg | 0.58 | 0.58 | 0.59 | 31 |
| Weighted avg | 0.65 | 0.64 | 0.65 | 31 |

Experiment 9 – Sampling Distribution

P (50%) – NP (50%)

Total OM: 14

Train OM : 12

Test OM : 2 (University OM + Decider)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30829

Pareto Optimal (P) instances: 15414

Pareto Optimal (NP) instances: 15414

Testing Set:

Total instances - 111

Pareto Optimal (P) instances: 25

Pareto Optimal (NP) instances: 80

Notebook URL: [Notebook-9](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/12_OM_50/12_OM_Train_2_OM_Test(University%2BDecider).ipynb)

Dataset URL: [Dataset-9](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/12_OM_50/12_OM_50.xlsx)

Testset URL: [Testset-9](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/12_OM_50/2-OM-test-set.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.82 | 0.81 | 0.82 | 89 |
| 1 (P) | 0.26 | 0.27 | 0.27 | 22 |
| Accuracy |  | | 0.70 | 111 |
| Macro avg | 0.54 | 0.54 | 0.54 | 111 |
| Weighted avg | 0.71 | 0.70 | 0.71 | 111 |

Experiment 10 – Sampling Distribution

P (50%) – NP (50%)

Total OM: 14

Train OM : 10

Test OM : 4 (University OM + Decider + CO + OS)

Dataset orientation: No space between words and symbols.

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30701

Pareto Optimal (P) instances: 15350

Pareto Optimal (NP) instances: 15350

Testing Set:

Total Dataset Instances: 127

Pareto Optimal (P) instances: 32

Pareto Optimal (NP) instances: 95

Notebook URL: [Notebook-10](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/10_OM_50/10_OM_Train_4_OM_Test(University%2BDecider%2BCO%2BOS).ipynb)

Dataset URL: [Dataset-10](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/10_OM_50/10_OM_50.xlsx)

Testset URL: [Testset-10](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_50%25_total_dataset%20/10_OM_50/4_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.84 | 0.89 | 0.87 | 95 |
| 1 (P) | 0.62 | 0.50 | 0.55 | 32 |
| Accuracy |  | | 0.80 | 127 |
| Macro avg | 0.73 | 0.70 | 0.71 | 127 |
| Weighted avg | 0.78 | 0.80 | 0.79 | 127 |

Experiment 11 – Sampling Distribution

P (60%) – NP (40%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (University OM)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 18564 P

Pareto Optimal (NP) instances: 12376 NP

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-11](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/13_OM_60/13_OM_Train_1_OM_Test(University)_60.ipynb)

Dataset URL: [Dataset-11](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/13_OM_60/13_OM_60.xlsx)

Testset URL: [Testset-11](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/13_OM_60/University_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.76 | 0.90 | 0.83 | 10 |
| 1 (P) | 0.67 | 0.40 | 0.50 | 21 |
| Accuracy |  | | 0.74 | 31 |
| Macro avg | 0.71 | 0.65 | 0.66 | 31 |
| Weighted avg | 0.73 | 0.74 | 0.72 | 31 |

Experiment 12 – Sampling Distribution

P (60%) – NP (40%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 12

Test OM : 2 (University OM + Decider)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30829

Pareto Optimal (P) instances: 18497 P

Pareto Optimal (NP) instances: 12332 NP

Testing Set:

Total instances - 111

Pareto Optimal (P) instances: 25

Pareto Optimal (NP) instances: 80

Notebook URL: [Notebook-12](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/12_OM_60/12_OM_Train_2_OM_Test(University%2BDecider)_60.ipynb)

Dataset URL: [Dataset-12](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/12_OM_60/12_OM_60.xlsx)

Testset URL: [Testset-12](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/12_OM_60/2-OM-test-set.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.87 | 0.90 | 0.88 | 89 |
| 1 (P) | 0.53 | 0.45 | 0.49 | 22 |
| Accuracy |  | | 0.81 | 111 |
| Macro avg | 0.70 | 0.68 | 0.69 | 111 |
| Weighted avg | 0.80 | 0.81 | 0.81 | 111 |

Experiment 13 – Sampling Distribution

P (60%) – NP (40%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 10

Test OM : 4 (University OM + Decider + CO + OS)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30701

Pareto Optimal (P) instances: 18420

Pareto Optimal (NP) instances: 12281

Testing Set:

Total Dataset Instances: 127

Pareto Optimal (P) instances: 32

Pareto Optimal (NP) instances: 95

Notebook URL: [Notebook-13](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/10_OM_60/10_OM_Train_4_OM_Test(University%2BDecider%2BCO%2BOS)_60.ipynb)

Dataset URL: [Dataset-13](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/10_OM_60/4_OM_test.xlsx)

Testset URL: [Testset-13](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_60%25_total_dataset/10_OM_60/4_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.81 | 0.22 | 0.35 | 95 |
| 1 (P) | 0.27 | 0.84 | 0.41 | 32 |
| Accuracy |  | | 0.38 | 127 |
| Macro avg | 0.54 | 0.53 | 0.38 | 127 |
| Weighted avg | 0.67 | 0.38 | 0.36 | 127 |

Experiment 14 – Sampling Distribution

P (70%) – NP (30%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (University OM )

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 21658P

Pareto Optimal (NP) instances: 9282 NP

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-14](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/13_OM_70/13_OM_Train_1_OM_Test(University)_70.ipynb)

Dataset URL: [Dataset-14](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/13_OM_70/13_OM_70.xlsx)

Testset URL: [Testset-14](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/13_OM_70/University_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.83 | 0.71 | 0.77 | 21 |
| 1 (P) | 0.54 | 0.70 | 0.61 | 10 |
| Accuracy |  | | 0.71 | 31 |
| Macro avg | 0.69 | 0.71 | 0.69 | 31 |
| Weighted avg | 0.74 | 0.71 | 0.72 | 31 |

Experiment 15 – Sampling Distribution

P (70%) – NP (30%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 12

Test OM : 2 (University OM + Decider )

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30829

Pareto Optimal (P) instances: 21580

Pareto Optimal (NP) instances: 9248

Testing Set:

Total instances - 105

Pareto Optimal (P) instances: 25

Pareto Optimal (NP) instances: 80

Notebook URL: [Notebook-15](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/12_OM_70/12_OM_Train_2_OM_Test(University%2BDecider)_70.ipynb)

Dataset URL: [Dataset-15](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/12_OM_70/12_OM_70.xlsx)

Testset URL: [Testset-15](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/12_OM_70/2-OM-test-set.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.94 | 0.17 | 0.29 | 89 |
| 1 (P) | 0.22 | 0.95 | 0.36 | 22 |
| Accuracy |  | | 0.32 | 111 |
| Macro avg | 0.58 | 0.56 | 0.32 | 111 |
| Weighted avg | 0.80 | 0.32 | 0.30 | 111 |

Experiment 16 – Sampling Distribution

P (70%) – NP (30%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 10

Test OM : 4 (University OM + Decider + CO + OS)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30701

Pareto Optimal (P) instances: 21491 P

Pareto Optimal (NP) instances: 9210 NP

Testing Set:

Total Dataset Instances: 127

Pareto Optimal (P) instances: 32

Pareto Optimal (NP) instances: 95

Notebook URL: [Notebook-16](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/10_OM_70/10_OM_Train_4_OM_Test(University%2Bdecider_co%2Bos)_70.ipynb)

Dataset URL: [Dataset-16](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/10_OM_70/10_OM_70.xlsx)

Testset URL: [Testset-16](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_70%25_total_dataset/10_OM_70/4_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.75 | 0.16 | 0.26 | 95 |
| 1 (P) | 0.25 | 0.84 | 0.39 | 32 |
| Accuracy |  | | 0.33 | 127 |
| Macro avg | 0.50 | 0.50 | 0.32 | 127 |
| Weighted avg | 0.62 | 0.33 | 0.29 | 127 |

Experiment 17 – Sampling Distribution

P (80%) – NP (20%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (University OM)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 24752

Pareto Optimal (NP) instances: 6188

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-17](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/13_OM_80/13_OM_Train_1_OM_Test(University)_80.ipynb)

Dataset URL: [Dataset-17](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/13_OM_80/13_OM_80.xlsx)

Test set URL: [Testset-18](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/13_OM_80/University_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.68 | 0.71 | 0.70 | 21 |
| 1 (P) | 0.33 | 0.30 | 0.32 | 10 |
| Accuracy |  | | 0.58 | 31 |
| Macro avg | 0.51 | 0.51 | 0.51 | 31 |
| Weighted avg | 0.57 | 0.58 | 0.57 | 31 |

Experiment 18 – Sampling Distribution

P (80%) – NP (20%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (University OM)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30940

Pareto Optimal (P) instances: 24752

Pareto Optimal (NP) instances: 6188

Testing Set:

Total Dataset Instances: 31

Pareto Optimal (P) instances: 10

Pareto Optimal (NP) instances: 21

Notebook URL: [Notebook-18](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/13_OM_80/13_OM_Train_1_OM_Test(University)_80.ipynb)

Dataset URL: [Dataset-18](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/13_OM_80/13_OM_80.xlsx)

Test set URL: [Testset-18](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/13_OM_80/University_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.68 | 0.71 | 0.70 | 21 |
| 1 (P) | 0.33 | 0.30 | 0.32 | 10 |
| Accuracy |  | | 0.58 | 31 |
| Macro avg | 0.51 | 0.51 | 0.51 | 31 |
| Weighted avg | 0.57 | 0.58 | 0.57 | 31 |

Experiment 19 – Sampling Distribution

P (80%) – NP (20%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 12

Test OM : 2 (University OM + Decider)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30829

Pareto Optimal (P) instances: 24663

Pareto Optimal (NP) instances: 6165

Testing Set:

Total instances : 111

Pareto Optimal (P) instances: 22

Pareto Optimal (NP) instances: 89

Notebook URL: [Notebook-19](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/12_OM_80/12_OM_Train_2_OM_Test_(University_%2B_Decider)_80.ipynb)

Dataset URL: [Dataset-19](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/12_OM_80/12_OM_80.xlsx)

Test set URL: [Testset-19](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/12_OM_80/2-OM-test-set.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.00 | 0.00 | 0.00 | 89 |
| 1 (P) | 0.20 | 1.00 | 0.33 | 22 |
| Accuracy |  | | 0.20 | 111 |
| Macro avg | 0.10 | 0.50 | 0.17 | 111 |
| Weighted avg | 0.04 | 0.20 | 0.07 | 111 |

Experiment 20 – Sampling Distribution

P (80%) – NP (20%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 10

Test OM : 4 (University OM + Decider + CO + OS)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 30701

Pareto Optimal (P) instances: 24560

Pareto Optimal (NP) instances: 6140

Testing Set:

Total Dataset Instances: 127

Pareto Optimal (P) instances: 32

Pareto Optimal (NP) instances: 95

Notebook URL: [Notebook-20](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/10_OM_80/10_OM_Train_4_OM_Test_(University_%2B_Decider_%2B_CO_%2B_OS)_80.ipynb)

Dataset URL: [Dataset-20](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/10_OM_80/10_OM_80.xlsx)

Test set URL: [Testset-20](https://github.com/rashedhasan090/OM-ML_Research/blob/main/Combined_datasets/14-OM-datasets/P_labels_oversample_80%25_total_dataset/10_OM_80/4_OM_test.xlsx)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.67 | 0.02 | 0.04 | 95 |
| 1 (P) | 0.25 | 0.97 | 0.40 | 32 |
| Accuracy |  | | 0.26 | 127 |
| Macro avg | 0.46 | 0.49 | 0.22 | 127 |
| Weighted avg | 0.56 | 0.26 | 0.13 | 127 |

**Experiments with separate Test set on 13 OM based sampling distributions.**

Experiment 21 – Sampling Distribution

P (50%) – NP (50%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (Traffic Controller)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 27379

Pareto Optimal (P) instances: 13689

Pareto Optimal (NP) instances: 13690

Testing Set:

Total Dataset Instances: 2739

Pareto Optimal (P) instances: 51

Pareto Optimal (NP) instances: 2688

Notebook URL:

Dataset URL:

Test set URL:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) | 0.98 | 0.99 | 0.99 | 2688 |
| 1 (P) | 0.17 | 0.08 | 0.11 | 51 |
| Accuracy |  | | 0.98 | 2739 |
| Macro avg | 0.58 | 0.54 | 0.55 | 2739 |
| Weighted avg | 0.97 | 0.98 | 0.97 | 2739 |

Experiment 22 – Sampling Distribution

P (60%) – NP (40%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (Traffic Controller)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 27379

Pareto Optimal (P) instances: 16427

Pareto Optimal (NP) instances: 10952

Testing Set:

Total Dataset Instances: 2739

Pareto Optimal (P) instances: 51

Pareto Optimal (NP) instances: 2688

Notebook URL:

Dataset URL:

Test set URL:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) |  |  |  | 2688 |
| 1 (P) |  |  |  | 51 |
| Accuracy |  | |  | 2739 |
| Macro avg |  |  |  | 2739 |
| Weighted avg |  |  |  | 2739 |

Experiment 23 – Sampling Distribution

P (70%) – NP (30%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (Traffic Controller)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 27379

Pareto Optimal (P) instances: 19165

Pareto Optimal (NP) instances: 8214

Testing Set:

Total Dataset Instances: 2739

Pareto Optimal (P) instances: 51

Pareto Optimal (NP) instances: 2688

Notebook URL:

Dataset URL:

Test set URL:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) |  |  |  | 2688 |
| 1 (P) |  |  |  | 51 |
| Accuracy |  | |  | 2739 |
| Macro avg |  |  |  | 2739 |
| Weighted avg |  |  |  | 2739 |

Experiment 24 – Sampling Distribution

P (80%) – NP (20%)

Dataset orientation: Training set contained space between words and symbols.

Total OM: 14

Train OM : 13

Test OM : 1 (Traffic Controller)

Type: Test set is unseen by the model.

Training Set:

Total Dataset Instances: 27379

Pareto Optimal (P) instances: 21903

Pareto Optimal (NP) instances: 5476

Testing Set:

Total Dataset Instances: 2739

Pareto Optimal (P) instances: 51

Pareto Optimal (NP) instances: 2688

Notebook URL:

Dataset URL:

Test set URL:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1 | Support |
| 0 (NP) |  |  |  | 2688 |
| 1 (P) |  |  |  | 51 |
| Accuracy |  | |  | 2739 |
| Macro avg |  |  |  | 2739 |
| Weighted avg |  |  |  | 2739 |