# Chapter 4

# Results

## 4.1 Performance

### 4.1.1 Credibility Classification

The performance of all 91 models were evaluated via the credibility classification experiment.

| Model | Criteria 1 | Criteria 2 | Criteria 3 | Criteria 4 | Criteria 5 | Criteria 6 | Criteria 7 | Average Performance |
|---|---|---|---|---|---|---|---|---|
| NB + BoW (all words) | 0.79 | 0.70 | 0.86 | 0.84 | 0.79 | 0.75 | 0.82 | 0.792 |
| NB + BoW (stopwords removed) | 0.85 | 0.79 | **0.90** | 0.84 | **0.80** | 0.74 | 0.80 | 0.820 |
| NB + TF-IDF (all words) | 0.79 | 0.74 | 0.87 | 0.80 | 0.79 | 0.62 | 0.80 | 0.773 |
| NB + TF-IDF (stopwords removed) | 0.79 | 0.80 | 0.89 | 0.81 | 0.77 | 0.69 | 0.81 | 0.794 |
| NB + GloVe | 0.82 | 0.78 | 0.89 | 0.82 | 0.78 | 0.68 | 0.83 | 0.800 |
| SVM + BoW (all words) | 0.82 | 0.70 | 0.86 | 0.80 | 0.68 | 0.75 | **0.86** | 0.799 |
| SVM + BoW (stopwords removed) | 0.78 | 0.80 | **0.90** | **0.87** | 0.78 | 0.72 | 0.81 | 0.809 |
| SVM + TF-IDF (all words) | 0.80 | 0.70 | 0.84 | 0.84 | **0.80** | 0.68 | 0.82 | 0.783 |
| SVM + TF-IDF (stopwords removed) | **0.86** | **0.85** | 0.89 | 0.86 | **0.80** | **0.77** | 0.82 | **0.836** |
| SVM + GloVe | 0.60 | 0.77 | 0.63 | 0.61 | 0.59 | 0.67 | 0.81 | 0.669 |
| QRNN + General LM | 0.38 | 0.36 | 0.41 | 0.37 | 0.36 | 0.38 | 0.39 | 0.379 |
| QRNN + Fine-tuned LM | 0.42 | 0.39 | 0.44 | 0.40 | 0.41 | 0.41 | 0.43 | 0.414 |
| QRNN + Domain Specific LM | 0.12 | 0.10 | 0.09 | 0.11 | 0.12 | 0.10 | 0.08 | 0.103 |

**Table 4.1:** Micro averaged f1-Scores of all models

## 4.1.2 Low Credibility Identification

### Single Model Approach

*−This model was chosen because it had the highest average performance value (average micro f1-score) from all of the models evaluated.−*

| Model | Micro Averaged f1-Score | Average Predicted Score Differential |
|---|---|---|
| SVM + TF-IDF (stopwords removed) | XX | XX |

**Table 4.2:** Performance of low credibility identification task via a single model approach.

*–Insert confusion matrix here–*

### Ensemble Approach

*–This will be composed of the classifiers that had the highest f1 score for each criteria–*

| Model | Micro Averaged f1-Score | Average Predicted Score Differential |
|---|---|---|
| Ensemble | XX | XX |

**Table 4.3:** Performance of low credibility identification task via an ensemble approach.

*–Insert confusion matrix here–*

# 4.2   Training Time

| Machine Learning Model | Training Time (seconds) |
|---|---|
| NB + BoW (all words) | 5.20 |
| NB + BoW (stopwords removed) | **4.86** |
| NB + TF-IDF (all words) | 5.05 |
| NB + TF-IDF (stopwords removed) | 5.05 |
| NB + GloVe | 3,628.82 |
| SVM + BoW (all words) | 5.91 |
| SVM + BoW (stopwords removed) | 4.87 |
| SVM + TF-IDF (all words) | 5.35 |
| SVM + TF-IDF (stopwords removed) | 5.58 |
| SVM + GloVe | 3,558.26 |

**Table 4.4:** Average training time of NB and SVM models.

| QRNN Model | Epoch Completion (minutes) |
|---|---|
| QRNN + General LM | 271 |
| QRNN + Fine-tuned LM | 292 |
| QRNN + Domain Specific LM | TBC |

**Table 4.5:** Average training time of all models.

## 4.3 Model Storage Requirements

| NB Model | Size (MB) |
|---|---|
| NB + BoW (all words) | 11.10 |
| NB + BoW (stopwords removed) | 11.00 |
| NB + TF-IDF (all words) | 13.20 |
| NB + TF-IDF (stopwords removed) | 13.20 |
| NB + GloVe | 25,410 |
| SVM + BoW (all words) | 7.91 |
| SVM + BoW (stopwords removed) | **7.87** |
| SVM + TF-IDF (all words) | 10.00 |
| SVM + TF-IDF (stopwords removed) | 7.88 |
| SVM + GloVe | 25,340 |
| QRNN + General LM | 3,612.70 |
| QRNN + Fine-tuned LM | 4,193.35 |
| QRNN + Domain Specific LM | TBC |

**Table 4.6:** Aggregated size of all models.