

Result Replication Guide

Machine Learning Text Classification Project

March 27, 2025

1 Replication Environment Setup

1.1 System Requirements

- Operating System: Linux/macOS/Windows
- Python 3.8+ recommended
- Anaconda/Miniconda (optional but recommended)

2 Experimental Results Across Datasets

2.1 Performance Metrics Summary

Table 1: Classification Performance Metrics Across Datasets

Dataset	Model	Accuracy	Precision	Recall	F1 Score	AUC
PyTorch	Baseline	0.6278	0.6077	0.7446	0.5557	0.7446
	SVM	0.9206	0.7130	0.5563	0.6170	0.9099
	BERT	0.9121	0.6000	0.8182	0.6923	0.9545
TensorFlow	Baseline	0.5611	0.6353	0.7214	0.5390	0.7214
	SVM	0.8931	0.7376	0.7419	0.7376	0.9385
	BERT	0.9265	0.8140	0.8333	0.8235	0.9687
Keras	Baseline	0.5590	0.6290	0.6974	0.5388	0.6974
	SVM	0.8800	0.7035	0.5733	0.6193	0.8879
	BERT	0.8947	0.6667	0.8000	0.7273	0.9511
MXNet	Baseline	0.6087	0.6140	0.7511	0.5486	0.7511
	SVM	0.9083	0.7183	0.3375	0.4391	0.8199
	BERT	0.8958	0.6667	0.3333	0.4444	0.6349

3 Replication Experimental Conditions

3.1 BERT Classification Script

- Random seed: 42
- Test split: 20% of data
- Epochs: 15
- Batch size: 32
- Learning rate: 2e-5
- Weight decay: 0.004
- Prediction threshold: 0.32

3.2 Naive Bayes Classification Script

- Runs: 10 experiments
- Test split: 30% of data
- TF-IDF Configuration:
 - Ngram range: (1,1)
 - Max document frequency: 0.9
 - Min document frequency: 3
- Oversampling: SMOTETomek with sampling strategy 0.7
- Classifier: Linear SVM with balanced class weights

4 Key Observations

- Across all datasets, BERT and SVM significantly outperform baseline models
- TensorFlow dataset shows the most consistent and high performance across metrics
- Accuracy ranges from 0.8800 to 0.9265 for advanced models
- AUC consistently above 0.8 for SVM and BERT models
- Recall varies more widely compared to other metrics

5 Replication Steps

1. Clone the repository
2. Install dependencies (refer to requirements.pdf)
3. Download dataset: `datasets/pytorch.csv`
4. Run script: `python bert_classification.py`

6 Potential Variations

Slight metric variations may occur due to:

- Hardware differences
- Specific dataset version
- Random initialization

7 Troubleshooting

If results significantly differ:

- Verify exact dependency versions
- Check dataset integrity
- Ensure identical random seeds