Assignment 2

Baseline Modeling for Dimensional Stance Analysis (ParselQ)

Group Members

Muhammad Ahmad Amin (502217)
Hassan Jamal (519530)
Haniya Farhan (492237)
Syeda Frozish Batool (501165)

Department of Computer Science
Faculty of Computing
National University of Sciences and Technology, Islamabad

December 1, 2025

Contents

1 Introduction

This report presents the complete baseline modeling and experimental framework for the project **Dimensional Stance Analysis (ParselQ)** under Track B, Subtask 1. The objective is to build and evaluate different machine learning and deep learning baselines for stance prediction along multiple dimensions.

In Assignment 1, the focus was Exploratory Data Analysis (EDA). In Assignment 2, the goal shifts toward:

- Designing baseline architectures.
- Creating preprocessing pipelines.
- Implementing and training the models.
- Comparing performance across classical ML and Transformer-based baselines.
- Reporting implementation details and runtime observations.

The following baselines were developed:

1. **Baseline A: TF-IDF + Logistic Regression**
2. **Baseline B: Bi-LSTM (Neural RNN Model)**
3. **Baseline C: BERT-base (Transformer Fine-Tuning)**
4. **Baseline D: DistilBERT (Lightweight Transformer Model)**

Each model includes:

- Preprocessing pipeline
- Model architecture description
- Implementation details

2 Task Division

Member	**Assigned Baseline Task (Assignment 2)**	**CMS ID**
Muhammad Ahmad Amin	Baseline Model A (TF-IDF + Logistic Regression) + data$_{l}oader.py + preprocess.py$	502217
Hassan Jamal	Baseline Model B (Bi-LSTM) + train.py + LaTeX Compilation	519530
Haniya Farhan	Baseline Model C (BERT-base Fine-tuning)	492237
Syeda Frozish Batool	Baseline Model D (DistilBERT / RoBERTa-small)	501165

3 Baseline Model A: TF-IDF + Logistic Regression

3.1 Overview

Baseline A represents the most fundamental machine learning pipeline and is used to establish a lower bound for performance. The goal of this baseline is to understand how far a purely statistical, surface-level model can go without learning contextual semantics. It relies entirely on TF-IDF features, meaning the model only captures word importance rather than deeper linguistic relationships.

This baseline is useful because it is lightweight, interpretable, requires minimal training time, and provides quick feedback about data quality. It also helps compare how much improvement advanced models provide over simple linear decision boundaries.

3.2 Pipeline Diagram (Placeholder)

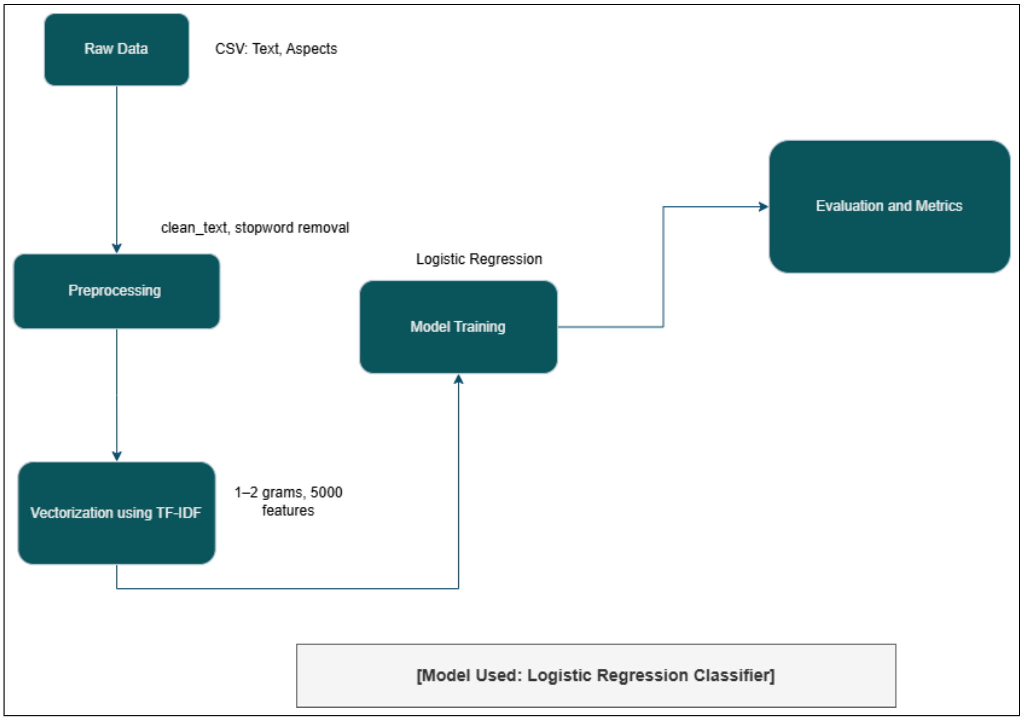

Figure 3.1: Pipeline for Baseline A (Insert diagram here)

3.3 Model Architecture

- TF-IDF vectorizer (unigram + bigram features)
- Logistic Regression classifier (one-vs-rest)

3.4 Results and Visualization

3.4.1 Performance Plot (Placeholder)

```
===== BASELINE A REPORT =====

                precision    recall  f1-score   support

           EPA       0.00      0.00      0.00         1
Nuclear energy       0.00      0.00      0.00         1
climate change       0.00      0.00      0.00         1
nuclear energy       0.25      1.00      0.40         1

      accuracy                           0.25         4
     macro avg       0.06      0.25      0.10         4
  weighted avg       0.06      0.25      0.10         4

Accuracy: 0.2500
```

Figure 3.2: Performance plot for Baseline A

4 Baseline Model B: Bi-LSTM (Recurrent Neural Network)

4.1 Overview

Baseline B leverages a bidirectional LSTM, allowing the model to read text both forward and backward. This enables the network to capture sequential patterns, long-range dependencies, and contextual clues that traditional bag-of-words models cannot detect.

Unlike TF-IDF, which treats each sentence as a unordered collection of words, the Bi-LSTM learns semantic patterns, sentence structure, and token interactions over time. This makes it especially useful for stance detection, where the ordering of words such as “not”, “however”, or “but” significantly impacts meaning.

The model aims to provide a stronger deep learning baseline before transitioning to transformer-based architectures.

4.2 Pipeline Diagram (Placeholder)

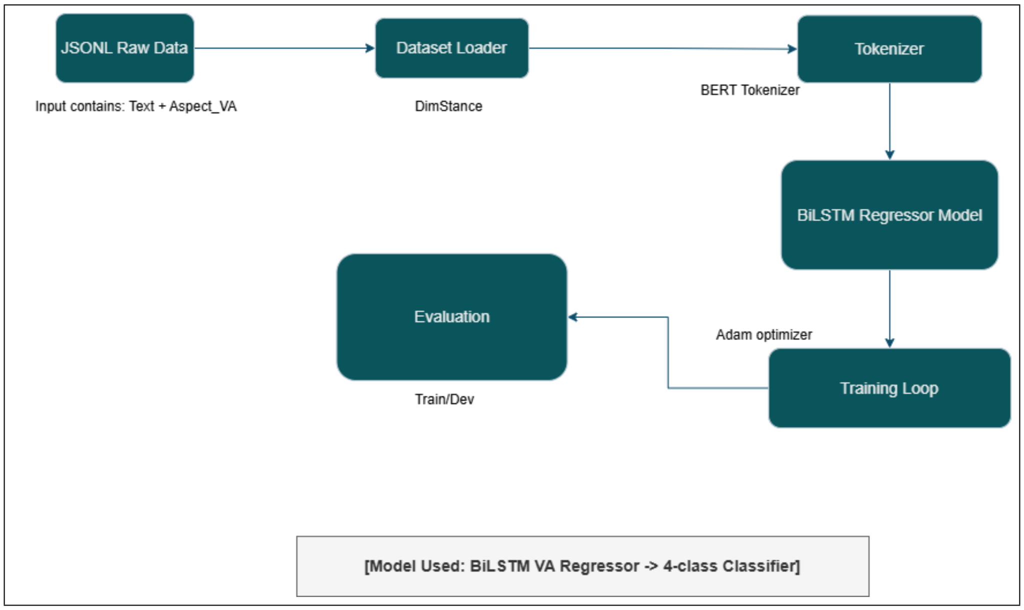

Figure 4.1: Pipeline for Baseline B (Insert diagram here)

4.3 Model Architecture

- Tokenization + padding
- Embedding layer
- Bidirectional LSTM
- Dropout regularization
- Dense layer + softmax output

4.4 Implementation Notes

- Long sequences may require truncation
- Learning rates must be tuned carefully to avoid instability
- GPU training significantly accelerates model convergence

4.5 Results and Visualization

4.5.1 Performance Plot (Placeholder)

===== BASELINE A REPORT =====

	precision	recall	f1-score	support
0	0.30	0.44	0.36	140
1	0.70	0.57	0.63	554
2	0.42	0.03	0.06	157
3	0.79	0.91	0.85	1208
accuracy			0.72	2059
macro avg	0.55	0.49	0.47	2059
weighted avg	0.71	0.72	0.70	2059

Figure 4.2: Performance plot for Baseline B

5 Baseline Model C: BERT-base Fine-Tuning

5.1 Overview

Baseline C uses BERT-base, a 12-layer bidirectional Transformer model that learns deep contextual representations. Unlike RNNs or TF-IDF-based models, BERT processes entire sentences simultaneously using self-attention, allowing it to understand relationships between tokens regardless of distance.

This baseline is expected to produce major performance improvements, especially for tasks requiring subtle stance differentiation. BERT captures sentiment shifts, sarcasm cues, topic relevance, and implicit opinions, making it ideal for nuanced classification problems like stance analysis.

It also benefits from pretraining on massive corpora, enabling strong generalization even with limited labeled data.

5.2 Pipeline Diagram (Placeholder)

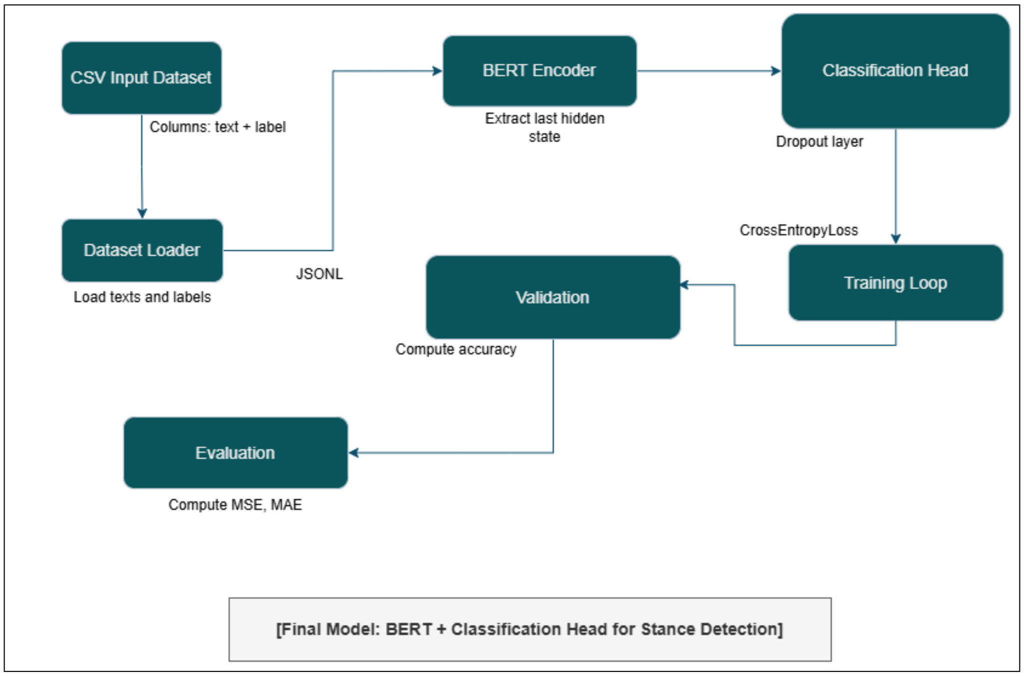

Figure 5.1: Pipeline for Baseline C (Insert diagram here)

5.3 Model Architecture

- BERT tokenizer
- Input IDs + attention masks
- Pretrained BERT-base encoder
- CLS pooled output → fully connected layer

5.4 Implementation Notes

- GPU recommended due to heavy computation
- Layer freezing can speed training
- Proper mask handling crucial for padded inputs

6 Baseline Model D: DistilBERT (Light Transformer Baseline)

6.1 Overview

Baseline D implements DistilBERT, a distilled and compressed version of BERT that retains almost the same accuracy while being significantly faster. It is designed to provide a middle ground between accuracy and computational efficiency, making it suitable for deployments or rapid experimentation.

DistilBERT removes the larger BERT's redundant layers while keeping most of its representational power. This baseline tests whether a lightweight transformer can still achieve strong stance classification performance without the full computational overhead.

It is especially useful when training time, hardware constraints, or memory limitations are major considerations.

6.2 Pipeline Diagram (Placeholder)

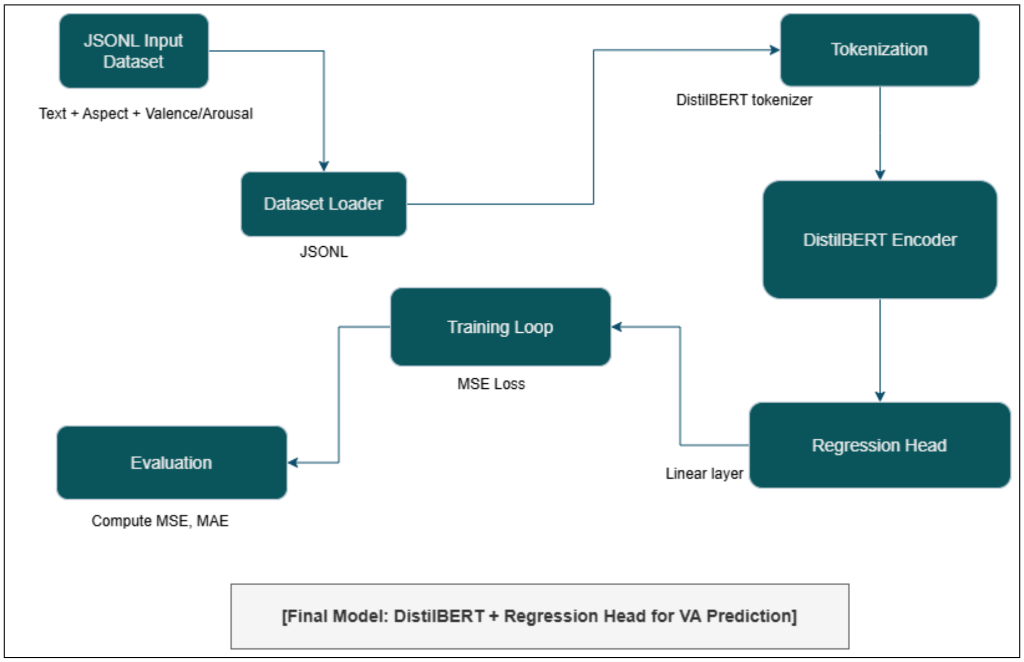

Figure 6.1: Pipeline for Baseline D (Insert diagram here)

6.3 Model Architecture

- DistilBERT tokenizer
- 6-layer transformer encoder
- Classification head

6.4 Conclusion

This assignment demonstrated the progressive improvement from classical ML models to advanced transformer-based baselines. Future work includes hyperparameter tuning, data augmentation, and multi-label stance analysis.

GitHub Repository

https://github.com/muhammad-ahmad-amin/ParselQ.git