7PAM2002-0509-2024 - Data Science Project

CONTEXT-AWARE SENTIMENT ANALYSIS USING LARGE LANGUAGE MODELS (LLMS)

SUPERVISOR : SUNINA SHARVY

SNEHLATA KUMARI
23080692
MSc Data Science
GitHub : [lata1207/Master_Project]

SCHOOL OF ENGINEERING AND COMPUTER SCIENCE
UNIVERSITY OF HERTFORDSHIRE

AGENDA

- Goal: Improve sentiment classification using dialogue context.
- Why it matters: Traditional models misclassify context-sensitive or multi-turn sentiment.
- Research Question: The traditional sentiment analysis often fails when context is subtle or multi turn like in conversations. Why this analysis failed to understand the deeper meaning of input texts. Can pre-trained language models improve sentiment classification by effectively incorporating conversational context?
- Objectives: 1.Use DailyDialog dataset (context-response pairs)

  2.Compare transformer models (RoBERTa and DistilBERT)

  3.Analyze impact of context on sentiment prediction

LITERATURE REVIEW

- ✓ Traditional models (SVM, LSTM) struggle with multi-turn dialogues.
- ✓ BERT (Devlin et al., 2019) revolutionized contextual NLP
- ✓ RoBERTa optimized transformer performance (Liu et al., 2019)
- ✓ CardiffNLP models fine-tuned for social media text (Barbieri et al., 2020)
- ✓ DailyDialog created for multi-turn emotion and dialogue act analysis (Li et al., 2017)

DATASET DESCRIPTION

Dataset: frankdarkluo/dailydialog (variant of DailyDialog)

Format:

73,554 samples with context and response fields

Why suitable:

Multi-turn structure with real-world dialogue style

Ideal for evaluating context-aware models

Dataset Ethics & GDPR Compliance:

License: CC BY-NC-SA 4.0

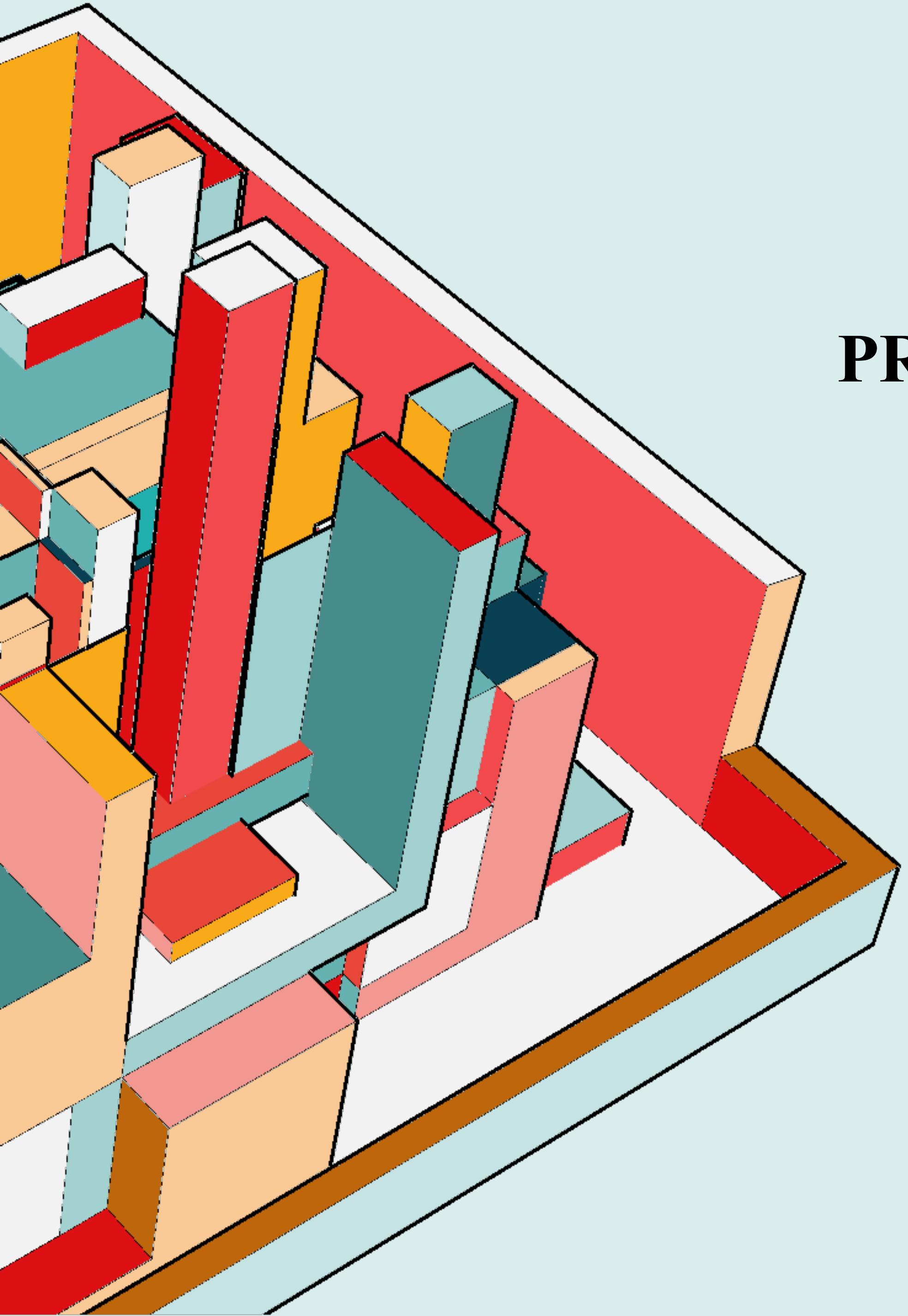
→Permits academic use with attribution (non-commercial)

Privacy & Anonymity:

→ No personal data or identifiers included

GDPR Alignment:

→ No personally identifiable information (PII)



PROBLEM WITH TRADITIONAL MODELS

- ✓ Sentence-level models ignore context
- ✓ Fail to detect sarcasm, emotion buildup, or indirect cues
- ✓ Real-world misclassification examples show model limitations

INPUT DESIGN & PREPROCESSING

- Input format: context + [SEP] + response
- Goal: Analyze how context shifts sentiment prediction
- Tokenization using respective model tokenizer
- Zero-shot inference only, pipeline-based inference
- Sentiment labels generated using Hugging Face pipelines
- Label normalization: mapped model labels to common Positive/Neutral/Negative classes (3-class standardization)

METHODOLOGY OVERVIEW (PIPELINE)

Dataset → Preprocessing → Context/Response Formatting → Model Inference → Sentiment Analysis → Visualization/Evaluation

(Fine Tuning in Progress)

Step 1: Load Dataset frankdarkluo/dailydialog → Context–Response dialogues Multi-turn → ideal for context-based sentiment

Step 2: Choose Models RoBERTa (3-class, trained on tweets) DistilBERT (2-class, trained on reviews)

Step 3: Format Inputs Without context → response only With context → context [SEP] response

Step 4: Predict Sentiment (Zero-shot), Use Hugging Face pipelines (no training), Label each sample twice (with vs. without context)

Step 5: Normalize Labels Standardize: Positive, Neutral, Negative. Enable fair model comparison

Step 6: Compare Predictions%, changed sentiment with context, which model was more robust?

Step 7: Visualize Results, Horizontal bar plots, Pie chart (context impact), Confusion matrix

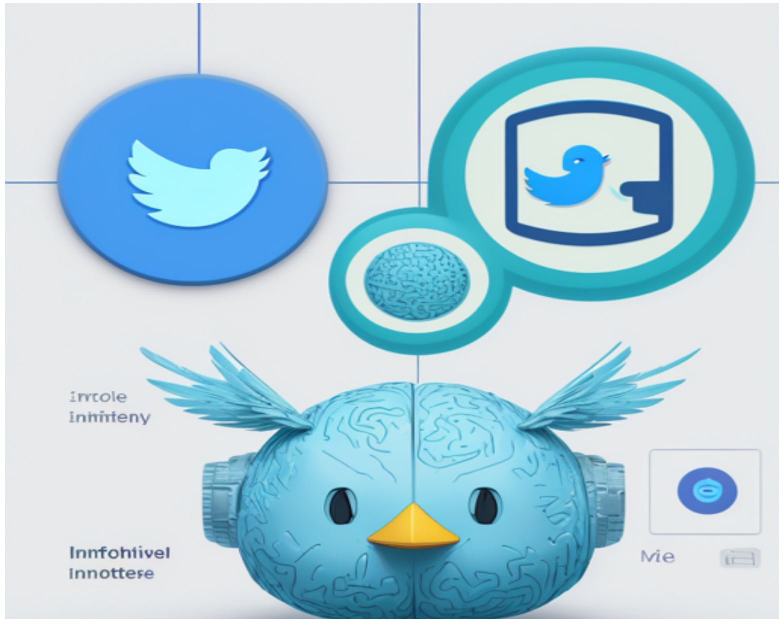
MODEL 1 – ROBERTA TWITTER SENTIMENT

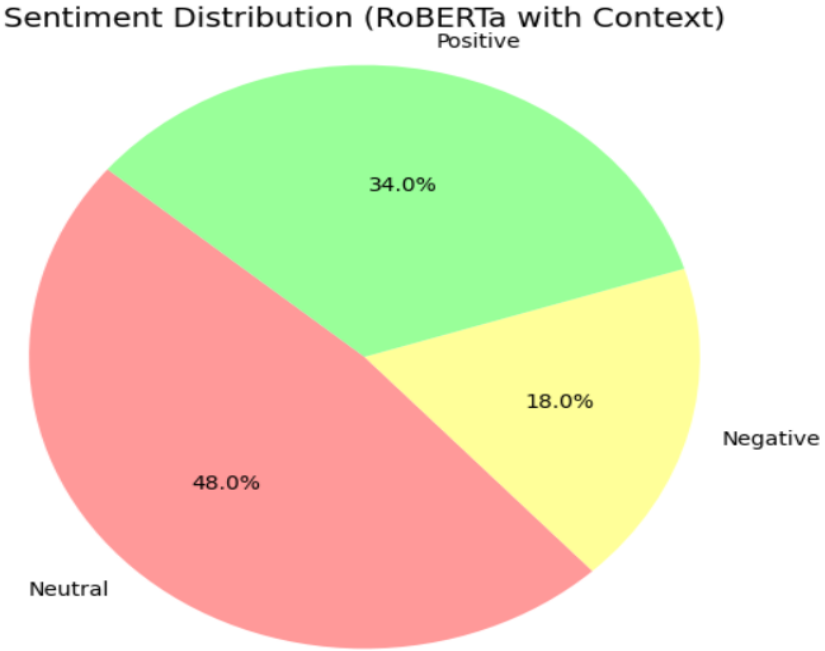
Model: cardiffnlp/twitter-roberta-base-sentiment

Pretrained on Twitter data

Outputs: Positive, Neutral, Negative

Strength: Good with informal, social-language data





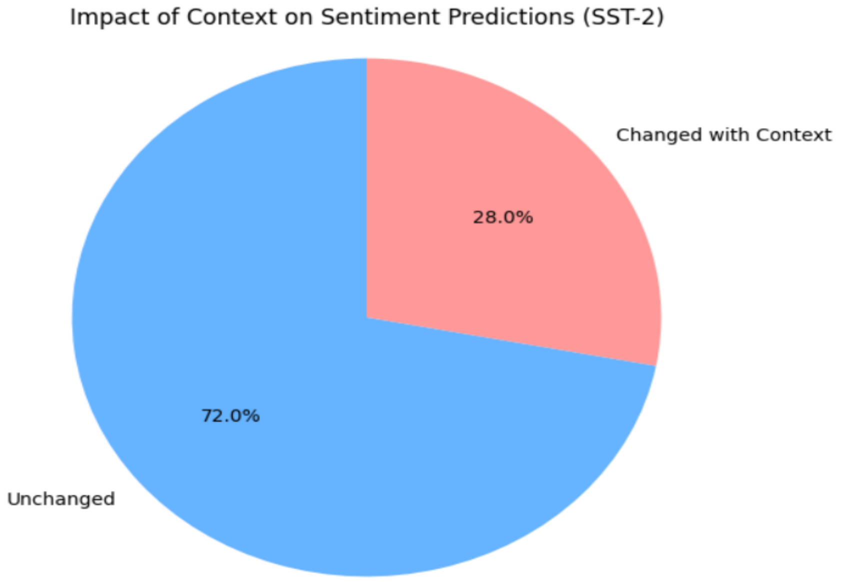
MODEL 2 – DISTILBERT SST-2

Model: distilbert-base-uncased-finetuned-sst-2-english

Pretrained on Stanford SST-2 movie review data (Stanford Sentiment Treebank (movie reviews))

Binary sentiment classification: Positive or Negative

Lightweight & fast inference

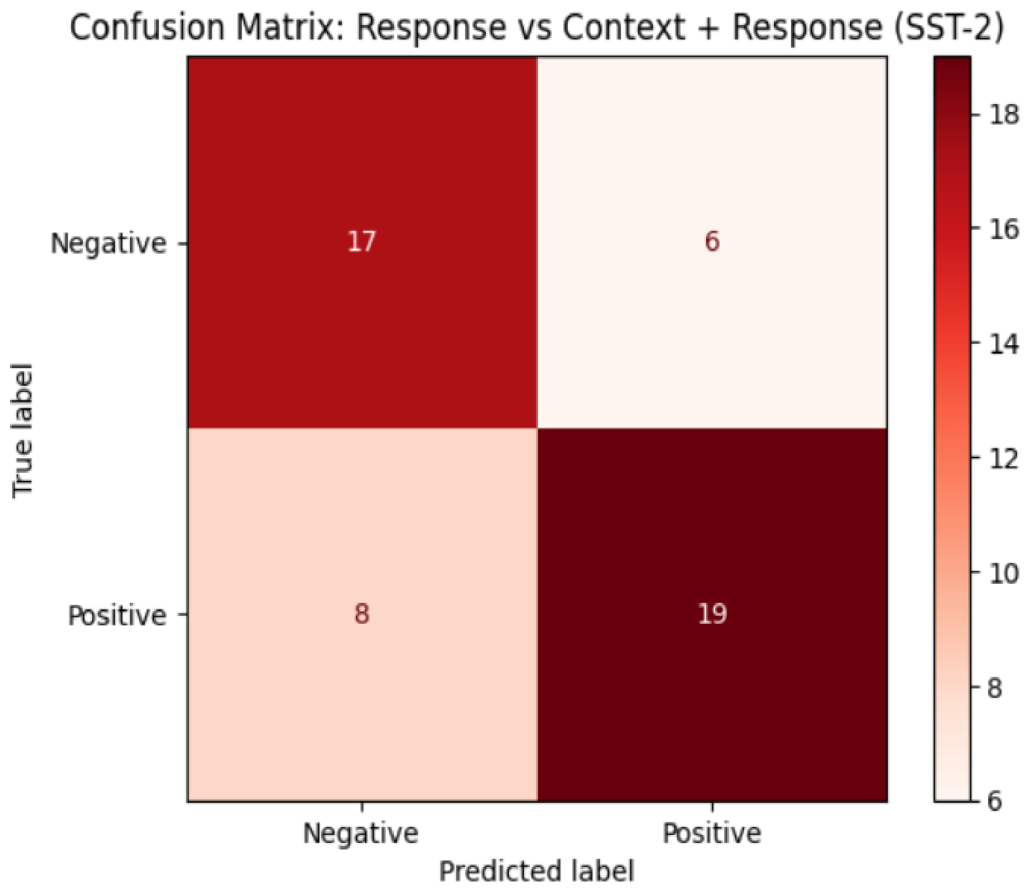


RESULTS – SENTIMENT DISTRIBUTION

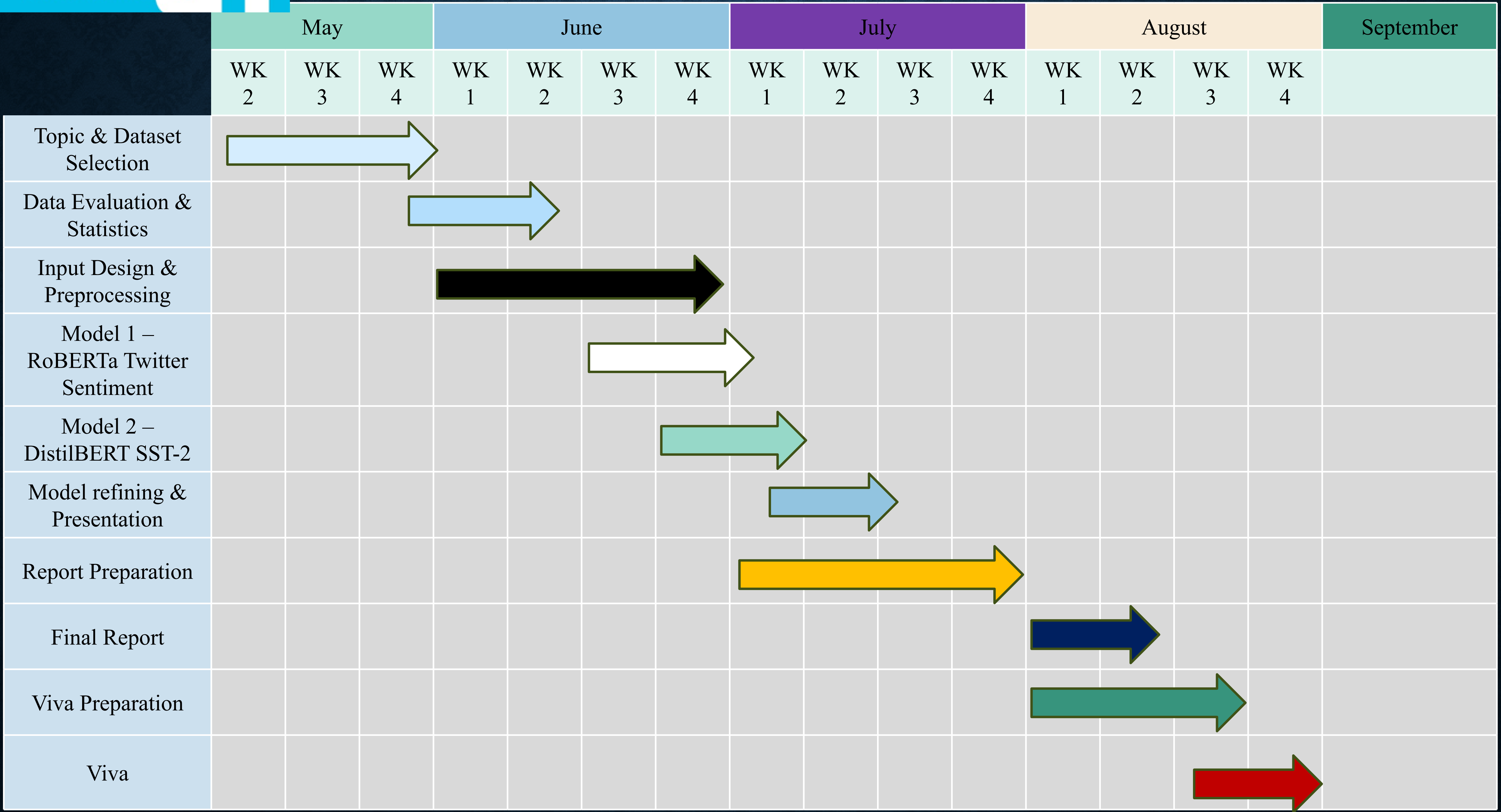
- Sentiment changed in ~22% of samples when context was included.
- RoBERTa captured nuanced sentiment better than DistilBERT
- Example: “Fine, I guess” → Positive alone, Neutral with context
- Context helps distinguish subtle emotional shifts
- RoBERTa: higher sensitivity to polite disagreement or sarcasm
- DistilBERT: strong but less adaptable in multi-turn scenarios
- Zero-shot generalization used for both models

ANALYSIS & INTERPRETATION

- Context improves accuracy in multi-turn dialogue
- DistilBERT misses emotional nuance without prior utterances
- Highlights need for dialogue-aware sentiment modeling



Timeline


May
June
July
August
September
WK 2
WK 3
WK 4
WK 1
WK 2
WK 3
WK 4
WK 1
WK 2
WK 3
WK 4
WK 1
WK 2
WK 3
WK 4
Topic & Dataset Selection
Data Evaluation & Statistics
Input Design & Preprocessing
Model 1 – RoBERTa Twitter Sentiment
Model 2 – DistilBERT SST-2
Model refining & Presentation
Report Preparation
Final Report
Viva Preparation
Viva

REFERENCES

- Barbieri, F., Camacho-Collados, J., Espinosa Anke, L. and Neves, L., 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421*.
- Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*.
- Li, Y., Su, H., Shen, X., Li, W., Cao, Z. and Niu, S., 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of IJCNLP*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... and Stoyanov, V., 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

THANK YOU !