PROJECT ABSTRACT

Stance Detection using BERT

bert-base-uncased fine-tuned on the SemEval-2016 Task 6 stance benchmark to classify tweets as FAVOR, AGAINST or NONE toward debate targets, with target-conditioned inputs and token-evidence highlighting - delivered as a built-to-order student project with code, notebook, report, PPT and viva kit.

01. Title

Stance Detection using BERT

02. Introduction / Background

Knowing whether a text supports or opposes a topic is different from knowing whether it is positive - an angry attack on a politician's critics is negative in tone but FAVOR in stance. Stance detection captures this: given a target topic and a text, decide FAVOR, AGAINST or NONE. It underpins debate analysis, policy monitoring and misinformation triage, and it is harder than sentiment analysis because stance lives in references, sarcasm and implication. This project fine-tunes bert-base-uncased on the SemEval-2016 benchmark with target-conditioned inputs ('target [SEP] tweet'), so the classifier learns the relationship between topic and text rather than memorizing keywords.

03. Problem Statement

Analysts, moderators and researchers need to know where texts stand on contested topics, and sentiment tools answer the wrong question. The project addresses this by building a stance classifier: given a debate target and a tweet, output FAVOR / AGAINST / NONE with confidence and highlight the words carrying the stance signal, with per-target evaluation that shows honestly where the model is strong and where it leans on shallow cues.

04. Objectives

- Fine-tune bert-base-uncased on SemEval-2016 Task 6 (4,870 tweets, 5 debate targets) for 3-way stance classification with target-conditioned inputs.
- Reach a design target of macro-averaged F1 approx. 0.60-0.68 across targets, with the training notebook logging per-target and macro F1 on the official test split.
- Build a demo app where the same text can be re-scored against different targets live, with token-evidence highlighting, dataset explorer and per-target evaluation views.
- Deliver the complete student kit: source code, notebook, report, PPT and viva Q&A.

05. Proposed System / Working

The system has three stages. (1) Data pipeline: SemEval-2016 tweets are formatted as target [SEP] tweet pairs and WordPiece-tokenized (max 128 tokens); FAVOR / AGAINST / NONE labels are kept from the official annotation. (2) Model: bert-base-uncased encodes the pair and the CLS embedding feeds a dropout + linear 3-way head, fine-tuned for about 4 epochs with AdamW; every epoch logs per-target F1 and macro-averaged F1. (3) Demo application: the chosen target and pasted tweet are formatted, run through the fine-tuned model, rendered as a stance verdict with confidence, and each token is colored by its attention-based evidence for the predicted stance.

06. Technologies / Components Used

- Python 3.10, PyTorch, Hugging Face transformers (bert-base-uncased)
- scikit-learn (metrics, confusion matrix), NumPy, pandas
- Matplotlib (per-target F1 bars, training curves)
- Single-file HTML/CSS/JS demo app (analyzer, dataset explorer, evaluation views)
- SemEval-2016 Task 6 stance benchmark (4,870 tweets, 5 targets)
- Trained weights exported from the included fine-tuning run

07. Key Features

- Stance analyzer: FAVOR / AGAINST / NONE verdict with confidence bar per target
- Target-conditioned input - re-score the same text against different targets live
- Token-evidence highlighting showing which words carried the stance signal
- Dataset explorer with per-target label distribution and annotated examples
- Per-target F1 bars and confusion matrices reproduced by the notebook
- Batch mode scoring (target, tweet) CSV pairs into a stance report
- Full per-target and macro evaluation logging - never pre-claimed

08. Applications / Use Cases

- Debate and policy monitoring: where does discussion stand on a contested topic
- Teaching material for pair-sequence modeling and stance-vs-sentiment reasoning
- First-pass triage for misinformation and moderation workflows (with human review)
- Reference build for other target-conditioned classification tasks

09. Results & Scope

The deliverable is a design-target-driven build, not a pre-measured product. The training notebook computes per-target and macro-averaged F1 on the official SemEval-2016 test split; the report presents these as the build's measured results. The design

target is macro-F1 approx. 0.60-0.68. Scope is the five SemEval targets; unseen topics, implicit sarcastic stances and the hard NONE class are documented limitations.

10. Conclusion

Stance Detection using BERT delivers an end-to-end applied NLP project on a subtle problem: target-conditioned modeling, per-target evaluation and an explainable demo that makes the stance-vs-sentiment distinction tangible. The honest per-target analysis gives the student credible, examiner-respecting material for the report and viva.

11. References

- Mohammad, S. et al. - SemEval-2016 Task 6: Detecting Stance in Tweets.
- Devlin, J. et al. - BERT: Pre-training of Deep Bidirectional Transformers (NAACL 2019).
- ALDayel, A. and Magdy, W. - Stance Detection on Social Media: State of the Art and Trends (Information Processing & Management, 2021).
- Hugging Face course - Fine-tuning a pretrained model (huggingface.co/course).