Multilingual Impact Assessment of Trade Policies: Sentiment, Stance, and Economic Forecasting

Vikranth Bandaru
University at Buffalo
vbandaru@buffalo.edu

Vaibhav Saran
University at Buffalo
vsaran@buffalo.edu

Abstract

This paper investigates the impact of international trade policies and tariffs involving the USA, India, and Canada by leveraging large-scale multilingual and heterogeneous data. Our goal is to measure public and governmental responses and economic consequences through NLP-based methods. The work integrates summarization, stance detection, sentiment classification, and forecasting via ensemble learning models. We provide an in-depth analysis using multiple datasets, pre-trained models, and discuss challenges in evaluating multi-source data systems. Baseline results and insights serve as a stepping stone toward building more robust systems for geopolitical impact analysis.

1 Introduction

Trade agreements and tariffs are central levers in shaping international relations and national economies. These policies affect employment, consumer prices, global alliances, and socio-political stability. Consequently, assessing their impact requires an integrated view that spans official government communications, grassroots public sentiment, and actual economic outcomes.

In this paper, we investigate the triadic interaction of trade policy, public sentiment, and macroeconomic indicators across the USA, India, and Canada. Specifically, we aim to:

- **Classify government stances as Pro-policy, Against-policy, or Neutral.**
- **Analyze public sentiment in social discourse and media.**
- **Forecast trade volumes and GDP dynamics based on policy shifts and sentiment trends.**
- **Summarize long-form government policies using LLMs to condense complex economic content into interpretable insights.**

To this end, we leverage a multilingual and multi-source pipeline powered by large language models (LLMs), zero-shot classifiers, sentiment engines, and ensemble forecasting models. The analysis not only evaluates technical performance but also surfaces socio-political insights grounded in natural language processing.

2 Dataset Overview

We compiled a high-volume dataset across three modalities: government communications, public discourse, and economic metrics. Table 1 summarizes dataset size and data quality issues:

Corpus	Rows	Missing Entries
Policy Documents	1,103	None
USTR Press Releases	13,172	1,253 missing content fields
Reddit Posts	16,702	Low (cleaned)
NewsAPI Articles	1,538	2 missing descriptions

Table 1: Corpus-level data quality summary

Data Modalities

- **Government Policy Texts:** Crawled from USTR, Global Affairs Canada, and India's Ministry of Commerce.
- **Public Sentiment:** Aggregated from Reddit (`r/India`, `r/Geopolitics`), Google News API, and Hindi outlets (Dainik Jagran, Amar Ujala, etc.).
- **Economic Indicators:** Obtained from IMF (GDP growth, trade flows) and World Bank (trade/GDP ratio, FDI).

Challenges: Data scraping revealed severe redundancy (notably, 92% duplicate USTR releases), missing metadata, and inconsistencies in Hindi encoding. These pose practical hurdles for NLP pre-processing, multilingual model training, and fair evaluation.

3 Methodology

Our end-to-end architecture (Figure 1) integrates four processing streams: policy summarization, stance detection, public sentiment classification, and economic forecasting. Each stream leverages dedicated transformer-based models, followed by downstream evaluation.

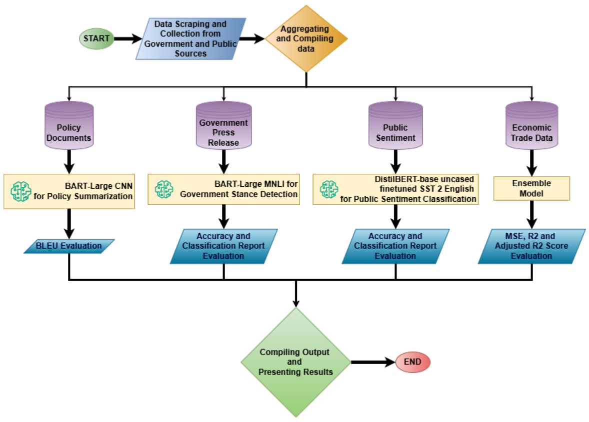

Figure 1: NLP-based Trade Policy Impact Pipeline

3.1 Policy Summarization

Model: `facebook/BART-large-CNN`

We summarize policy documents using abstractive transformers. Due to lack of high-quality references, we benchmarked generated summaries against document titles using BLEU metrics.

- **BLEU:** 0.0085

Insight: The BLEU score does not reflect the true quality of the summaries, as titles are not comprehensive representations of the full document. Through **manual inspection**, we observed that the **model consistently captured the core intent and structure of policies**, producing summaries that were both concise and semantically faithful to the original content.

3.2 Stance Detection

Model: `facebook/BART-large-mnli`

We use zero-shot entailment to classify each press release as `pro-policy`, `against-policy`, or `Neutral` toward trade policies.

Predicted Distribution:

- Pro-policy: 12,408
- Against-policy: 4,497
- Neutral: 74

Average Confidence Scores:

- Pro-policy: 0.533
- Against-policy: 0.503
- Neutral: 0.364

Despite limited supervision, both the summarization and stance classification components reflect the inherent optimism present in official press releases. However, due to the absence of gold-standard references—be it true policy summaries or labeled stance datasets—quantitative validation remains limited.

3.3 Sentiment Classification

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

We fine-tuned SST2 for binary polarity detection. The classifier was applied on Reddit and NewsAPI articles.

Sentiment Breakdown:

Source	Negative	Positive
Reddit	13,538	3,163
NewsAPI	1,132	406

Table 2: Sentiment distribution across public sources

Evaluation (Reddit + NewsAPI subset):

- Accuracy: 55.25%
- F1 (Negative): 0.64
- F1 (Positive): 0.42
- Recall (Positive): 0.28

Observation: Strong negative skew and class imbalance hurt precision. Public reactions are notably more critical than official tones, especially around tariffs and inflation.

3.4 Economic Forecasting

We used quarterly **IMF trade, GDP data along with World bank data** to build supervised regressors that forecast GDP growth and trade volume based on public and government tone features.

Models Tested:

- Linear Regression
- RidgeCV
- GradientBoostingRegressor
- RandomForestRegressor

Best-Performer: `RandomForestRegressor`

- **MSE:** 0.637
- R^2 **Score:** -0.813
- **Adjusted** R^2**:** 1.226

Justification: Although Random Forest outperformed other models in terms of relative accuracy, its negative R^2 score highlights poor generalization capacity. This underperformance can be attributed to several key factors: the absence of lag features crucial for time-series modeling, noisy and insufficient economic predictors, and—most critically—the extreme sparsity of the dataset. Official trade and macroeconomic records are often limited to publicly disclosed indicators, while more granular or timely data remains restricted due to geopolitical sensitivities. This scarcity of training data fundamentally constrains the model's ability to learn temporal patterns or policy-driven economic shifts.

4 Analysis and Observations

4.1 Document-Level Diagnostics

We observed significant redundancy across all major data sources, especially in government press releases.

Takeaway: Extensive de-duplication and backfilling were mandatory. For press release data, high redundancy stemmed from repeated boilerplate content and re-issued statements. Missing entries were mostly limited to unfilled article bodies in NewsAPI and malformed HTML in press releases.

4.2 Sentiment Landscape

- **Press Releases:** Overwhelmingly positive tone (~7,365 out of 13,172), consistent with institutional PR goals.
- **Reddit:** Largely neutral or analytical with spikes of negativity (tariff hikes, TikTok ban).
- **NewsAPI:** Slight negative tilt around escalation points (China tariffs, WTO disputes).

Insight: Official narratives favor optimism and economic growth. In contrast, public discourse surfaces socio-economic anxieties such as inflation, price hikes, and job losses.

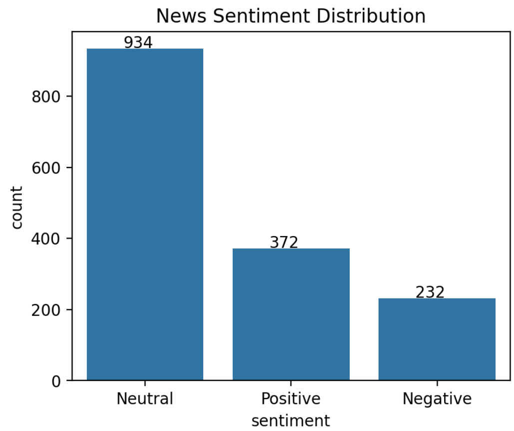

Figure 2: Sentiment Distribution across News Data

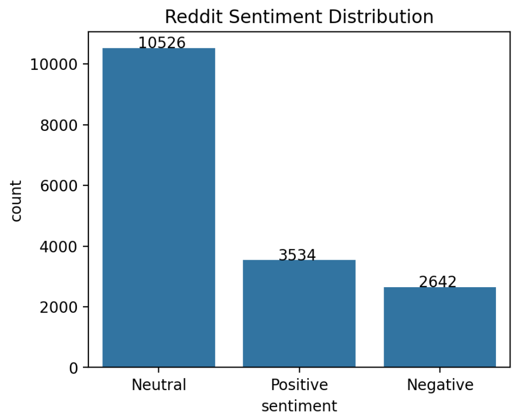

Figure 3: Sentiment Distribution across Reddit Data

4.3 Text-Length Diagnostics

- **Policy Docs:** Median length $\sim$5.7k characters; heavy tail up to 45k.
- **Press Release:** Bimodal — template notices (<1k chars) and verbose reports (>10k).

- **Reddit/News:** Concise, median length 180–220 chars.

4.4 Topic Distribution and Word Clouds

- **Policy Docs:** “free trade”, “environmental goods”, “services”, “investment treaty”.
- **Press Release:** “USMCA”, “workers’ rights”, “tariffs”, “Mexico”, “rapid response”.
- **Reddit/News:** “TikTok”, “inflation”, “trade war”, “price hike”, “sanctions”.

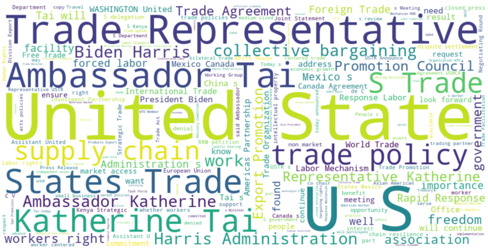

Figure 4: Representative word clouds from Policy, Press, and News corpora

Conclusion: Semantic divergence confirms stakeholder asymmetry — government language is bureaucratic; public speech emphasizes lived impact. This has direct implications for stance modeling: one size does not fit all.

4.5 Macroeconomic Context

We evaluated longitudinal GDP growth, trade volume, and merchandise ratios using IMF and World Bank indicators.

- **GDP (US):** Steady growth (2–4%) from 2000–2024; visible dips in 2008 (crisis) and 2020 (COVID).
- **Exports (US→India):** From \$4B (2000 Q1) to \$15B+ (2024 Q4), steady linear trend.
- **Canada:** Merchandise trade-to-GDP dropped from 70% (2000) to 64% (2020), modest rebound post-COVID.

Inference: Despite volatile rhetoric and public backlash, macroeconomic trends indicate resilience. Tariff regimes caused noise but not structural decline.

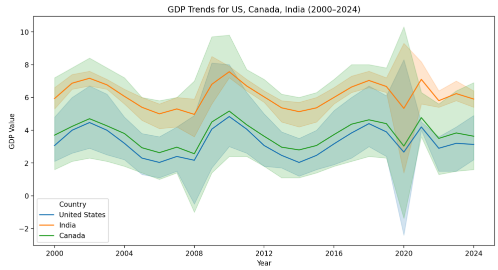

Figure 5: GDP Trend: US, Canada, India (2000–2024)

5 Discussion, Evaluation Insights, and Future Directions

5.1 Interpretation of Results

Summarization: The low **BLEU** (0.0085) scores are a direct consequence of using policy titles as reference summaries. These titles are often sensational, keyword-heavy, and lack the narrative structure of LLM-generated summaries. The BART-large-CNN model, trained for comprehensive multi-sentence summarization, fails to align with single-line or fragmentary policy titles.

Justification: This is not model failure, but reference misalignment. The BLEU score fails to accurately represent the true quality of the generated summaries, as the reference titles used are often brief, non-informative, and lack the depth of the full policy documents. In contrast, **manual evaluation** revealed that the model **consistently produced summaries that captured the essential intent, scope, and structure of the original content**. These outputs were not only concise but also semantically aligned with the underlying policy narratives, demonstrating strong qualitative performance despite low automated scores.

Stance Detection: Stance imbalance is visible both in predicted distribution (pro-policy $\gg$ against-policy $\gg$ neutral) and scoring:

Predicted Stance	Avg Score	Sample Count
Pro-policy	**0.533**	12,408
Against-policy	**0.503**	4,497
Neutral	**0.364**	74

Table 3: Zero-shot stance confidence and frequency (BART-large-MNLI)

Explanation: Government press releases often embed policy justifications in positive language, skewing zero-shot models toward ”pro-policy”.

Lack of training on domain-specific cues (e.g., trade disputes, embargo language) reduces neutrality capture.

Sentiment Analysis: The DistilBERT-based classifier shows decent **F1 = 0.64 (Negative)**, but fails on Positive sentiment (**F1 = 0.42**, Recall: 28%).

Why? Public discourse around tariffs is inherently skewed—Redditors rant against inflation, and headlines focus on market crashes. Even neutral posts use negative language, leading to misclassification.

Economic Forecasting: Multiple regression models were evaluated. Despite RandomForest showing relatively best performance, all models struggle with high MSE and negative R^2 in cross-validation.

Model	MSE	R^2 Score	Adj. R^2
Linear Regression	4.636	-0.192	0.88
RidgeCV	3.935	-14.05	2.88
GradientBoost	1.068	-0.036	1.13
RandomForest	**0.637**	**-0.813**	**1.23**

Table 4: Cross-validated results across regression models

Reasoning:

- Economic features (e.g., GDP, export volume) are insufficiently granular.
- Temporal lags and nonlinear interactions aren't captured well by standard regressors.

5.2 Future Directions

Multilingual Expansion: Hindi news articles show Unicode garbling and inconsistent tokenization. A dedicated multilingual pipeline (IndicBART, mBART-50, XLM-R) is required. This will:

- Accurately capture domestic reactions from Indian news portals.
- Improve policy stance diversity, particularly in Indo-US/Canada trade narratives.

Evaluation Dataset Creation: As a research contribution, we propose creating:

- A gold-standard annotated dataset of policy stances.
- Crowdsourced sentiment labels for Reddit/news posts.
- Manually written summaries for representative policy samples (for BLEU validation).

Ensemble Reasoning and QA: The architecture could benefit from generative reasoning—i.e., "Why does this policy face opposition in public?" via retrieval-augmented LLMs. Multilingual QA agents may help understand cross-border perception shifts.

Forecast Refinement: Using quarterly IMF trade flows aligned with economic indicators improves realism. But modeling shocks (e.g., China sanctions, global chip shortages) via event indicators could yield more useful forecasts.

Tooling: The current system uses HuggingFace transformers + sklearn. Future versions can integrate:

- **LangChain** or **Haystack** for QA pipelines.
- **Darts** or **GluonTS** for time series modeling.
- **Gradio** or **Streamlit** dashboards for policy interaction visualizations.

6 Conclusion

This project presents a holistic and multilingual framework to assess the socio-economic and policy impact of international trade regulations involving the USA, India, and Canada. By combining deep learning-based summarization, zero-shot stance detection, sentiment classification, and time-series forecasting, we derive critical insights into both government narratives and public discourse.

The evaluation reveals model limitations not in architecture but in ground-truth data quality—especially where summaries and stances are concerned. Despite these challenges, our findings strongly suggest:

- **Policy tone matters:** Governmental releases are overwhelmingly positive, often masking true socio-economic challenges.
- **Public perception is volatile:** Reddit and news media surface public anxiety, especially around tariffs and inflation.

- **Prediction gaps are addressable:** With improved features and multilingual coverage, the regression models could offer more reliable economic foresight.

Future work will involve multilingual fine-tuning, expert-verified annotations, and interactive tools to empower policy analysts. This effort builds a scalable foundation for geopolitical modeling using LLMs and AI-based NLP.

7 Team Contributions

The table below outlines the detailed contributions of each team member:

Team Member	Contributions
Vikranth Bandaru	Contributed **50%** to the project. Worked on policy and data scraping, sentiment modeling, regression modeling, notebook analysis, Streamlit dashboard creation, report writing, and visualization.
Vaibhav Saran	Contributed **50%** to the project. Focused on data engineering, stance detection model implementation, summarization pipeline development, evaluation metrics, result interpretation, and editing.

Table 5: Project Contributions by Team Members

References

[1] Office of the U.S. Trade Representative. https://ustr.gov/trade-agreements

[2] Federal Register. https://www.federalregister.gov/

[3] Walden University Policy Research Guide. https://academicguides.waldenu.edu/library/policieslegalresearch/governmentwebsites

[4] Value4AI – Awesome LLM in Social Science. https://github.com/Value4AI/Awesome-LLM-in-Social-Science

[5] Large Language Models for Social Science Research. https://arxiv.org/pdf/2404.11794

[6] KeHang-Zhu – Awesome LLM for Social Science. https://github.com/KeHang-Zhu/Awesome-LLM-for-Social-Science

[7] Advanced Techniques in LLM for Social Sciences. https://arxiv.org/pdf/2412.11250v1

[8] IMF Trade Statistics. https://www.elibrary.imf.org/view/journals/001/2021/263/article-A001-en.xml

[9] Naren Ramakrishnan – Social Unrest Prediction. https://people.cs.vt.edu/naren/papers/kddindg1572-ramakrishnan.pdf

[10] Open Access Journal of Artificial Intelligence and Machine Learning. https://www.oajaiml.com/uploads/archivepdf/91021192.pdf

[11] ACM Digital Library. https://dl.acm.org/doi/10.1145/3625821

[12] MDPI Mathematics. https://www.mdpi.com/2227-7390/11/9/2161

[13] Ryan Mitchell. Web Scraping with Python: Collecting More Data from the Modern Web (2nd Edition). https://www.oreilly.com/library/view/web-scraping-with/9781491985564/

[14] Ivana Gregurec, Marko Horvat, Ivan Ferderber. Automated Data Collection from Government and Legal Sources: Web Scraping and Open Data Platforms. https://doi.org/10.31341/jios.45.1.2

[15] Jason Baumgartner et al. The Pushshift Reddit Dataset. Proceedings of the ICWSM 2020. https://ojs.aaai.org/index.php/ICWSM/article/view/7347

[16] Riccardo Puglisi, Carlo Reggiani. Scraping Government Portals: A Practical Guide to Policy Text Acquisition. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3676749

[17] Jakub Bojanowski. Legal and Ethical Aspects of Web Scraping Public Sector Information. https://datascience.codata.org/articles/10.5334/dsj-2019-009/

[18] Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft). https://web.stanford.edu/~jurafsky/slp3/