E-ISSN: 2962-6110
P-ISSN: 2964-7843
JSRET
jsret.knpub.com
Journal of Scientific Research, Education, and Technology

Cross-Lingual Sentiment Analysis for Indonesian Monetary Policy

Akbar Ramadhan, Umar Zaky
University Technology of Yogyakarta, Yogyakarta, Indonesia

ABSTRACT

This research develops a cross-lingual sentiment analysis system (RoBERTa-IndoBERT) to monitor public opinion on Bank Indonesia’s 2025 monetary policy from X (Twitter), addressing the scarcity of Indonesian labels and noisy social media text. We introduce a "translate-then-classify" pipeline: Indonesian posts are translated into English, auto-labeled by a mature English RoBERTa model, and these labels are used to fine-tune IndoBERT on the original texts. We compare this cross-lingual (CL) approach, with and without back-translation (BT) augmentation, against a baseline Indo-only model. Performance measured by Accuracy and Macro-F1 indicates the CL pipeline is substantially better than the baseline. The complete model (IndoBERT + CL + BT) yields a Macro-F1 of 98.1%, a 2.8 percentage point (pp) improvement over the baseline (95.3%). Qualitative error analysis corroborates the CL model is more stable, less prone to extreme polarity flips, and better at detecting implicit sentiment. This research demonstrates that a CL auto-labeling pipeline is an efficient and resilient solution for Indonesian sentiment analysis in low-resource scenarios.

Keywords: *Cross-Lingual Sentiment Analysis, Auto-Labeling, Weak Supervision, RoBERTa, IndoBERT, Back-Translation, Monetary Policy, Indonesia.*

Corresponding author
Name: Akbar Ramadhan
Email: akbarrmdhn1304@gmail.com

INTRODUCTION

Three external headwinds significantly challenged the rupiah in 2024, including U.S. dollar appreciation, heightened global policy uncertainty, and volatile capital flows, resulting in higher exchange-rate volatility and sustained depreciation vis-à-vis the USD. Consistent with its mandate for macroeconomic stability, Bank Indonesia has consistently stated that it is ready to intervene in the FX market to smooth out undue fluctuations and maintain price stability. The rupiah has a history of hitting multi-year lows in mid-2024, followed by wobbles into the end of the year. These at least seem consistent with market thinking and Bank Indonesia's official communication at the time (Pramono, 2025;Reuters, 2024).

Uncertainty about the exchange rate is not only the result of changes in the underlying macroeconomic fundamentals but also of the formation of expectations and public sentiment. Social media is a kind of real-time channel for opinion formation and diffusion within Indonesia’s digital ecosystem. Although X (formerly Twitter) is not the market leader by audience share, its user base is still quite large. DataReportal estimates that the potential advertising reach of the platform in Indonesia was around 24.7 million accounts at the beginning of 2024, making it a reliable indicator of policy attitudes. Worldwide, user attention is more evenly distributed among the competing platforms, which means that X is still an influential but not the leading mainstream channel (Kemp, 2024).

Modern monetary economics progressively relies on text and sentiment analyses to understand how the market perceives the central bank's policy. Several studies provide a compelling picture that sentiment derived from the conventional media and social platforms is in line with the reactions of the financial market and can be helpful in predicting the policy path, thus pointing out that NLP-driven methods are not just a replacement but a complementary way of work and visualization (Picault et al., 2022)).

Recent literature provides more evidence for this, with research pointing to the fact that macroeconomic issues like inflation and monetary policy are the ones that resist the highest public engagement on X in a systematic manner (Gorodnichenko et al., 2025). This draws attention to the worth of NLP-based sentiment gauges as decision-support tools in the monetary policy environment (Czudaj & Nguyen, 2025).

Large-scale text analysis is very useful; however, analyzing tweets is quite difficult due to factors such as the use of colloquial language, abbreviations, sarcasm, and implicit context. The difficulty is compounded by the scarcity of high-quality, gold-standard datasets in the Indonesian language that specifically focus on policy and monetary discourse. The absence of in-domain supervision typically results in poorer model generalization, especially when models are presented with linguistically unusual inputs. The Indonesian NLP research also highlights the lack of annotated resources as a long-standing issue (Koto et al., 2020).

This paper introduces a cross-lingual pipeline to address the problem of insufficient labeled data for Indonesian. The first step involves translating Indonesian tweets into English. Next, an English RoBERTa model is used to assign sentiment labels automatically. Then, a post-label back-translation augmentation that goes from Indonesian to English and back to Indonesian is employed, which is essentially aimed at minority classes to relieve class imbalance while semantic fidelity (e.g., similarity thresholds) is preserved to prevent label drift. Finally, the labels are mapped to the Indonesian text to fine-tune IndoBERT as the final classifier. Basically, the approach employs the advancements of the English NLP ecosystem (large corpora and robust training recipes) for the advantage of local nuance through IndoBERT fine-tuning on Indonesian data. The arrangement corresponds to cross-lingual sentiment analysis and distant/weak supervision literature that suggests that cross-lingual transfer, automatic labeling, and back-translation augmentation can achieve higher performance when there are few gold labels (Xu et al., 2022).

The different pieces of research put together in this paper show that RoBERTa is a very strong and stable model that can be used as a single backbone to carry out various types of classification tasks on English texts, amongst which is sentiment analysis (Liu et al., 2019; Semary et al., 2023). Besides, Back-Translation (BT) has been hailed as an effective augmenting technique that can be used to progressively create more diverse data without additional human annotation, especially in lightly labeled situations (Xie et al., 2020; Li et al., 2022). This method is one in which a sentence is converted into a pivot language and then back to its original language for the purpose of generating paraphrases.

This study mainly centers on developing and evaluating a sentiment-analysis system that detects people's reactions to X concerning Bank Indonesia's 2025 monetary policy. We employ a cross-lingual pipeline that transfers labels from an English RoBERTa model to IndoBERT in Indonesian, thereby cutting down the label count and still keeping the language-specific nuances. The ultimate aim is to have such sentiment indicators that are robust, quick in terms of reaction time, and can be easily scaled up to allow decision-makers to track public sentiments almost in real-time.

METHOD

This study utilizes a quantitative experimental design to create and assess a cross-lingual sentiment analysis system, RoBERTa-IndoBERT, to track public sentiment toward the 2025 Bank Indonesia monetary policy based on X (formerly Twitter) data. The method used two different pipelines for comparison, as shown in **Figure 1**: (A) a baseline Indo-only pipeline and (B) a cross-lingual pipeline. The cross-lingual approach (Pipeline B) uses machine translation (Indonesian to English) to allow the automatic sentiment labeling by a mature English RoBERTa model. After that, these induced labels are matched with the original Indonesian texts to adjust the final IndoBERT classifier that would be able to handle noisy Twitter texts even if there were few labels (Přibáň et al., 2024).

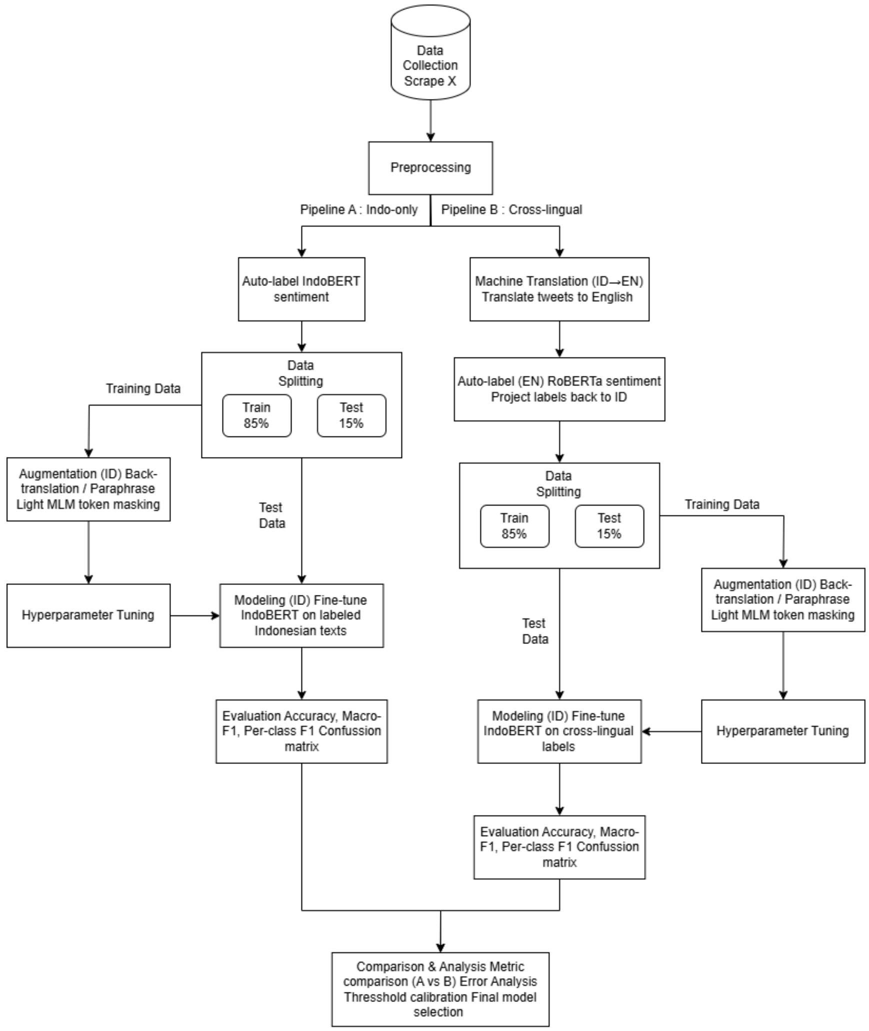

Figure 1: Research Methodology

The design method is in line with the latest CLSA surveys, which show that machine translation-based (MT-based) pipelines (translate-then-classify) are very competitive when there are no gold labels in the target language but a strong source-language model is available. Locally, the source-side labels are projected back to the target-language texts, and the target model is fine-tuned for effective domain adaptation. Besides that, the progress in weak/distant supervision reveals that emoji- and hashtag-driven noisy labels can be used to scale the training when principled denoising and curation (e.g., confidence filtering) are involved, therefore, giving more strength to the auto-labeling plus selective filtering before supervised fine-tuning (Gaurav et al., 2024).

Automated labeling by AI components, along with downstream classification to make the system more efficient, is quite central to the system's architecture. Recent papers keep presenting RoBERTa as the main backbone for sentiment tasks, which is in line with its use as a trusted English-side labeler. Meanwhile, resources for the Indonesian language are becoming more and more numerous; projects like NusaCrowd (standardized, open data across Indonesian and local languages) and IndoBERTweet (an X-domain Indonesian PLM) are opening the way for strong downstream adaptation in the target environment (Semary et al., 2023).

Besides being accurate, the system architecture is basically designed to be efficient and scalable for real-world scenarios. To minimize the computational cost, we have put in place resource-aware measures (batched translation/inference, sequence truncation, early stopping on validation macro-F1, and targeted back-translation augmentation) that do not cause a drop in the performance level. Recent comprehensive studies on data augmentation confirm that one of the best augmentation methods is translation-based augmentation (especially back-translation) in the case of label scarcity, whereas the developments in weak/distant supervision show that the noisy labels generated from emojis and hashtags can be used for training large datasets if denoising/curation is done properly (Pellicer et al., 2023).

Corpus Construction & Data Collection

We got the info from the X API by running the query "(Kebijakan BI 2025 OR #BankIndonesia OR #KebijakanMoneter) lang:id -is:retweet" that restricts the search to Indonesian-language posts and excludes retweets. The gathering was in place from March to October 2025 (UTC+7) and was in accordance with the platform's rate limits and Terms of Service. Only public posts were accessed. The first crawl fetched 2,588 posts; after deduplication/retweet removal and low-signal filtering, 2,587 posts were left.

For reproducibility purposes, we keep the post IDs, timestamps, and cleaned text, and we do not include the nonessential personally identifiable information. The collection was done in weekly installments to be able to follow the topical drift and have a balanced temporal coverage.

Text Processing & Cross-Lingual Auto-Labeling

Firstly, we perform a concise standardizing process for the text data, which includes lowering the case, removing punctuation, normalizing the whitespace, filtering stopwords, and using a split-based tokenizer. In addition to that, informal language is regularized with the help of a slang lexicon. For modeling, the canonical text field is the cleaned_tweet (if not available, cleaned_text is used). Then, the Indonesian posts are translated into English with the help of the Helsinki-NLP/opus-mt-id-en to alleviate the truncation artifacts (batch size = 64; max sequence length = 256 subword tokens). After that, sentiment to the English text is assigned by the CardiffNLP/twitter-roberta-base-sentiment-latest model, which returns {negative, neutral, positive} labels along with confidence scores. These labels are then mapped to the original Indonesian texts for forming supervised training pairs.

The label distribution after the mapping is quite imbalanced, i.e., negative 504 (19.48%), neutral 1,457 (56.32%), and positive 626 (24.20%), which is the reason that class-weighted losses are preferred during training instead of aggressive oversampling in order to keep linguistic realism intact.

Robustness Measures & Data Augmentation

In order to promote out-of-distribution generalization on noisy, user-generated social-media text, we implement a small augmentation-stability loop that is specially aimed at the most underrepresented and most error-prone half of the cases (in particular, the negative and positive classes, IDs 0 and 2). In agreement with the recent reviews, paraphrasing and back-translation give stable improvements when there are few labels, which is the case when semantic-preserving constraints are used to limit the change of meaning (Feng et al., 2021).

Therefore, we generate one augmented variant for each chosen instance (AUG_PER_SAMPLE = 1) under three protections. Firstly, to be more robust against local deletions/typos, we introduce about 12% of MLM-style token masking. Secondly, in order to have controlled lexical variation, we pick the top-5 paraphrases with a cosine similarity $\geq$ 0.80 from which one is selected. Lastly, we have a post-augmentation length restriction (max_length = 256). Besides, we implement label consistency as a requirement: augmented samples are kept only if the English-side labeler gives the same sentiment to the original instance. This cautious augmentation method increases the lexical/syntactic variation without changing the meaning, it helps minority-class signal more since errors are concentrated there, and it avoids an aggressive oversampling, which would result in class priors getting skewed.

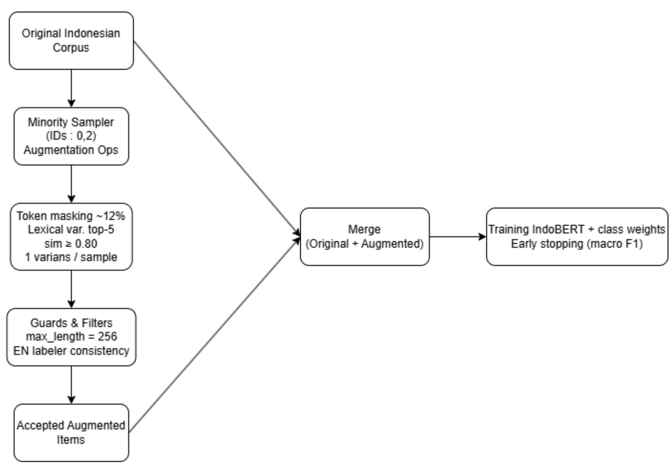

Figure 2: Light-touch augmentation pipeline

We use a light-touch augmentation protocol that is shown in **Figure 2** and is, first of all, a protocol that prioritizes minority classes, introduces controlled lexical and masking noise, and removes any variants that violate length constraints or cross-lingual label consistency by the time they are merged with the training corpus.

IndoBERT Model & Training Protocol

IndoBERT-p2 is the model employed, and it has been fine-tuned for ternary sentiment classification. This involves adding a single softmax classification head to the pre-trained model and then training it on Indonesian texts that have been aligned with the labels projected from the English-side annotator.

In order to maintain class proportions in different partitions, stratified sampling is used: 10% of the corpus is kept as a final test set, and 10% of the remaining data is used as a validation set. The inputs are tokenized with the IndoBERT tokenizer with a maximum sequence length of 256; the sequences that are longer are truncated, whereas the shorter inputs are padded.

The model is optimized using the AdamW optimizer with a learning rate of 2e-5, a batch size of 16, 3 epochs, and a weight decay of 0.01. Due to the heavy neutral bias (56.32% of the data) that is significantly pronounced, class-frequency weights are used during training to correct the distribution, and early stopping on validation macro-F1 is employed to stop overfitting.

Evaluation Protocol

The evaluation protocol is tightly structured to measure the model's ability to generalize beyond the sample. Only the validation split is used for model selection, and the stratified test set is kept aside for the final evaluation to avoid selection bias The model's performance was finally measured through four standard metrics, which were derived from the standard outputs of a confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Setiawan et al., 2025).

1. Accuracy
   Measures the overall correctness of the model, calculated as the ratio of all correct predictions to the total number of samples (1).

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

2. Precision
   Evaluates the accuracy of the positive predictions. It is the ratio of true positives to the total number of instances predicted as positive (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3. Recall
   Determines the model's ability to identify all relevant instances. It is calculated as the ratio of true positives to the total number of actual positive instances (3).

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4. F1-Score
   provides a single metric that balances Precision and Recall by calculating their harmonic mean (4).

$$F_1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

The Macro-F1 score represented the main criterion to evaluate the quality of predictions in this research. The reason for its selection lies in its ability to provide a fairer and more reliable measure than accuracy for imbalanced datasets, since it gives the same weight to all classes and is a good indicator when the classes are imbalanced. Macro-F1 is computed as the unweighted average of the F1 scores of each class (neutral, negative, and positive).

FINDING AND DISCUSSION

RESEARCH RESULT

Under the suggested framework, the model is able to carry out strong performance on the test set that is completely held out and fully unseen. Overall, the metrics show that the model is very accurate and behaves in a balanced way with respect to classes. In numbers, the model is able to obtain an Accuracy of 0.9826, a Weighted F1 of 0.9826, and a macro-F1 of 0.9816, which highlights that the model is not only correct most of the time but also performs equally well for all labels.

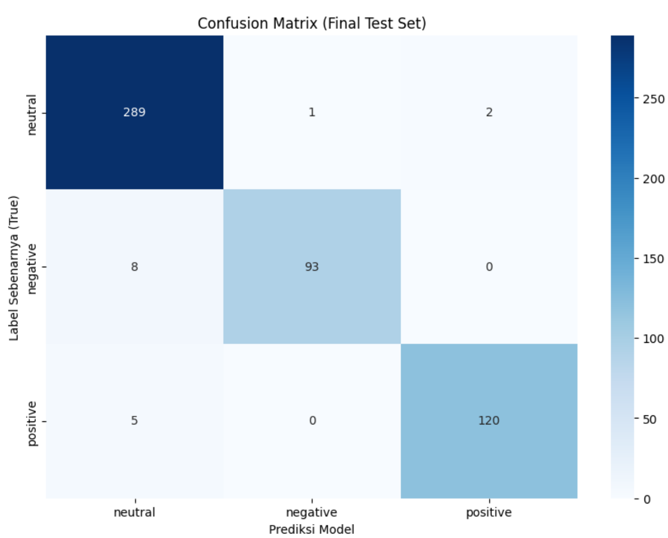

Figure 3: Confusion Matrix on the Held-Out Test Set

The confusion matrix (**Figure 3**) illustrates the results to be very close to the diagonal, which is a strong indication of the overall correctness being pretty high. The small number of residual errors can be found specifically at the boundary between neutral and mild positive or negative cases, which is a usual pattern for low-affect content (e.g., policy summaries). Importantly, there are no instances of extreme polarity flips (e.g., classifying negative as positive, or vice-versa), which means that the decision boundary for strong sentiment is stable. The F1-scores for each class were also on a high level: 0.9789 (neutral), 0.9846 (negative), and 0.9841 (positive).

Baselines & Ablations

To disentangle the effects of cross-lingual auto-labeling (RoBERTa-EN) and back-translation (BT), four distinct model configurations were compared, with results summarized in **Table 1**.

Table 1. Component-wise performance comparison

Model Configuration	Accuracy	Macro-F1	Δ Macro-F1	F1-negative	F1-neutral	F1-positive
IndoBERT (Baseline)	95.9%	95.3%	-	93.9%	96.9%	95.1%
IndoBERT + BT	96.7%	96.3%	+1.0 pp	95.2%	97.2%	96.7%
IndoBERT + CL	96.9%	96.7%	+1.4 pp	97.2%	96.9%	96.0%
IndoBERT + CL + BT	98.2%	98.1%	+2.8 pp	98.4%	97.9%	98.4%

The key findings, summarized in **Table 1**, reveal that the cross-lingual transfer approach is by far more effective than the Indonesian-only baseline. Just by introducing the Cross-Lingual (CL) auto-labeling component (IndoBERT + CL) to the IndoBERT (baseline) setup, a Macro-F1 improvement of +1.4 percentage points (pp) can be observed, thus the score is raised from 95.3% to 96.7%. The main source of the performance increase, therefore, seems to be the cross-lingual auto-labeling step.

The best setup, IndoBERT + CL + BT (the complete model), is the one that leads to the highest performance with 98.2% Accuracy and 98.1% macro-F1. As stated in **Table 1**, this means a total increment of +2.8 percentage points (pp) in Macro-F1 compared to IndoBERT (baseline) is recorded. An examination for each class reinforces that these enhancements are the places where the most significant number of mistakes were previously made. For example, the F1-negative score has gone up from 93.9% (baseline) to 97.2% (with CL) and 98.4% (with CL + BT), respectively. Similarly, the F1-positive score was raised from 95.1% (baseline) to 96.0% (with CL) and 98.4% (with the full pipeline) ultimately.

DISCUSSION

Interpretation of Main Findings

The combined numbers (Accuracy 0.9826, Macro-F1 0.9816) are strong evidence that the model is very accurate and also fair in its performance across imbalanced labels. The examination of the confusion matrix (**Figure 3**) complements this statement. There are very few residual errors, and those are mainly located at the border between neutral and mildly positive/negative instances, which is a usual distribution for low-affect content (e.g., policy summaries). What is more, there are no extreme polarity flips (e.g., classifying negative as positive, or the other way around) observed, which suggests that the model has learned a stable and semantically coherent decision boundary for strong sentiment.

Comparative Benchmarking

Our baseline model performance (IndoBERT-only), which made 95.9% Accuracy and 95.3% macro-F1, is in line with the state-of-the-art research that has been done previously. For instance, a research by (Setiawan et al., 2025), that also employed IndoBERT for sentiment analysis of complex public policy topics (Constitutional Court decisions) gave an accuracy of 95% as a result. The outcome of this work shows that our baseline is a solid point of departure, and it can be compared with the existing benchmarks. There is also a paper by (Manoppo et al., 2025) on the 12% VAT increase, where they report slightly lower figures, such as an F1-score of 84.37%, which probably points to the differences in the complexity of the domain.

By contrast, the major rise to a Macro-F1 of 98.1% (+2.8 pp) after the use of cross-lingual auto-labeling (CL) is a strong indication of the dominance of this pipeline. This result serves as a real-world test and confirmation of the survey arguments, e.g., (Přibáň et al., 2024; Xu et al., 2022), where it is claimed that a translate-then-classify approach is very viable for languages with limited resources.

Our RoBERTa-IndoBERT pipeline with 98.1% Macro-F1 can be considered a considerable step forward in the identification of the correct sentiment class of the Indonesian monetary policy discourse when compared to the recent domain-specific benchmarks. It is, therefore, very clear that the Transformer-based cross-lingual auto-labeling method (RoBERTa) is by far the best performer among the traditional machine learning methods on the text domains that are noisy and quite complex in nature.

Impact of Cross-Lingual Labeling and Augmentation

Most of the improvements in the ablation experiments (**Table 1**) are due to the cross-lingual transfer method, which, by a large margin, beats the Indonesian-only baseline. The addition of the Cross-Lingual (CL) auto-labeling module (IndoBERT + CL) alone results in a significant increase in the Macro-F1 by +1.4 percentage points (from 95.3% to 96.7%). This shows that the main performance-enhancing factor is the cross-lingual stage, where the data is automatically labeled with high quality.

The best setup, IndoBERT + CL + BT (the complete model), is the one that yields the highest performance of 98.1% Macro-F1. This is a total improvement of +2.8 pp over the baseline. A per-class analysis confirms that the enhancements correspond to the areas where the most frequent types of errors were previously. For example, the F1-negative score was raised from 93.9% (baseline) to 97.2% (with CL) and finally to 98.4% (with CL + BT). Likewise, the F1-positive score was elevated from 95.1% to 98.4% with the full pipeline. This reinforces that the CL and BT duo work wonderfully in bringing up the minority class representation.

Error Analysis and Practical Implications

In general, misclassifications tend to be concentrated near the boundary between the neutral and mild-polarity (slightly positive or negative) regions, especially in the cases of headline-style expressions and low-affect policy summaries without any obvious sentiment cues.

Table 2. Qualitative Error Analysis for Low-Affect Content

Original Text (Indonesian)	Translated Text (English)	IndoBERT (Baseline) Prediction	RoBERTa (Labeler) Prediction	Analysis / Implication
"10 Negara dengan Penduduk Termiskin Versi Bank Dunia, Indonesia Juara 4 lewat"	"Indonesia Ranks 4th Among the 10 Countries with the Poorest Populations, According to the World Bank"	Neutral (0.514)	Negative (0.754)	RoBERTa successfully captures the implicit negative valence of "poorest populations," which the baseline model misses.

Table 2 shows an example of a representative case. The Indonesian sentence, '10 negara dengan penduduk termiskin versi Bank Dunia, Indonesia juara 4 lewat,' looks like a factually statement only. The IndoBERT (baseline) model appropriately changes it to neutral (Confidence: 0.514). But after Indonesian-to-English translation, the RoBERTa-EN labeler points to it as negative (Confidence: 0.754). The difference is justified semantically: the English translation, "ranks 4th among the poorest populations," gives a negative meaning quite clearly, even though the original Indonesian doesn't have any explicit affective cues. So, this is the proof that the cross-lingual auto-labeling pipeline is more consistent and more like a human in that it can pick up on the implicit sentiment of low-affect content.

From a performance point of view, the IndoBERT + CL pipeline (no augmentation) is generally a good accuracy–efficiency trade-off that is suitable for environments with limited computing power. A Back-Translation (BT) augmentation (that is, the complete IndoBERT + CL + BT model) should be used only when a considerably increased resistance to paraphrastic and slang variants is necessary, for instance, in the case of a nearly–real-time monitoring of the public discourse.

CONCLUSION

The research reveals that a cross-lingual translate-then-label strategy is the most effective approach for tracking public sentiment about Bank Indonesia's monetary policy. The best outcomes on a held-out test set are obtained when Roberta in English is used for automatic supervision and then IndoBERT is fine-tuned on the original Indonesian texts (Accuracy = 0.9826; Macro-F1 = 0.9816), achieving balanced per-class F1 scores and avoiding severe polarity flips. Intra-lingual cross-lingual labeling consistently increases Macro-F1, and light back-translation is particularly beneficial for minority classes, as demonstrated by the comparison with an Indonesian-only baseline. The evidence presented here suggests that the limitations in the number of gold-standard labels and the presence of noise in social media texts can be alleviated by transferring supervisory signals from well-developed English resources, projecting the labels to Indonesian, and performing targeted fine-tuning.

IndoBERT, along with a cross-lingual pipeline without augmentation, is basically a good trade-off between accuracy and efficiency for limited computing environments. When paraphrase, slang, and lexical variation resistance are a focus, the version with back-translation is better. In addition to model metrics, the method allows almost real-time monitoring of people's views of monetary policy and can be a substitute as well as a support to traditional instruments used by analysts and policymakers.

There are still several limitations remaining. The quality of the translation and the projection of labels may cause bias. Low-affect informational posts and the use of subtle sarcasm in the area of neutrality are still difficult. Also, topic drift over time can lower the performance of the model if it is not updated.

Subsequent research may involve human-in-the-loop validation to be used as a reference for challenging cases, keep track of temporal drift, and use periodic domain adaptation to counter it, expand coverage to more platforms and languages spoken in different regions, be able to integrate uncertainty estimation for risk-aware deployment, and also be able to provide an operational dashboard that shows key metrics, confidence signals, and early-warning alerts. These measures can ensure that accuracy, scalability, and decision relevance are retained in policy analytics that are conducted over an extended period of time.

REFERENCES

Czudaj, R. L., & Nguyen, B. N. (2025). ECB’s central bank communication and monetary policy transmission: predictability from text-based sentiment indicators? *Macroeconomic Dynamics*, *29*. https://doi.org/10.1017/S1365100525000239

Feng, S. Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T., & Hovy, E. (2021). A Survey of Data Augmentation Approaches for NLP. *Computer Science*, 968–988. https://arxiv.org/abs/2105.03075

Gaurav, A., Gupta, B. B., Sharma, S., Bansal, R., & Chui, K. T. (2024). XLM-RoBERTa Based Sentiment Analysis of Tweets on Metaverse and 6G. *Procedia Computer Science*, *238*, 902–907. https://doi.org/10.1016/j.procs.2024.06.110

Gorodnichenko, Y., Pham, T., & Talavera, O. (2025). Central bank communication on social media: What, to whom, and how? *Journal of Econometrics*, *249*. https://doi.org/10.1016/j.jeconom.2024.105869

Kemp, S. (2024, February 21). *Digital 2024: Indonesia*. DATAREPORTAL. https://datareportal.com/reports/digital-2025-sub-section-top-social-platforms?

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. Online. https://huggingface.co/

Li, B., Hou, Y., & Che, W. (2022). Data augmentation approaches in natural language processing: A survey. *AI Open*, *3*, 71–90. https://doi.org/10.1016/j.aiopen.2022.03.001

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. http://arxiv.org/abs/1907.11692

Manoppo, M. R., Kolang, I. C., Nur Fiat, D. N., Mawara, R. M. C., Sumarno, A. D. P., Yusupa, A., & Tarigan, V. (2025). ANALISIS SENTIMEN PUBLIK DI MEDIA SOSIAL TERHADAP KENAIKAN PPN 12% DI INDONESIA MENGGUNAKAN INDOBERT. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, *4*(2), 152–163. https://doi.org/10.69916/jkbti.v4i2.322

Pellicer, L. F. A. O., Ferreira, T. M., & Costa, A. H. R. (2023). Data augmentation techniques in natural language processing. *Applied Soft Computing*, *132*. https://doi.org/10.1016/j.asoc.2022.109803

Picault, M., Pinter, J., & Renault, T. (2022). *Media sentiment on monetary policy: determinants and relevance for inflation expectations*.

Pramono, B. (2025, June 18). *BI-Rate Held at 5.50% Maintaining Stability, Strengthening Economic Growth*. Bank Indonesia. https://www.bi.go.id/en/publikasi/ruang-media/news-release/Pages/sp_2713325.aspx?utm_source=chatgpt.com

Přibáň, P., Šmíd, J., Steinberger, J., & Mištera, A. (2024). A comparative study of cross-lingual sentiment analysis. *Expert Systems with Applications*, *247*(C). https://www.sciencedirect.com/science/article/abs/pii/S095741742400112X

Reuters. (2024, June 14). *Indonesia c.bank intervenes to defend faltering rupiah*. Reuters. https://www.reuters.com/business/finance/indonesia-cbank-says-policy-aims-ensure-inflation-controlled-rupiah-stable-2024-06-14/?utm_source=chatgpt.com

Semary, N. A., Ahmed, W., Amin, K., Pławiak, P., & Hammad, M. (2023). Improving sentiment classification using a RoBERTa-based hybrid model. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1292010

Setiawan, D., Utari Iswavigra, D., & Anggiratih, E. (2025). Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum

Age of Vice Presidential Candidates. *Scientific Journal of Informatics*, *12*(3). https://doi.org/10.15294/sji.v12i3.26360

Xie, Q., Dai, Z., Hovy, E., Luong, M.-T., & Le, Q. V. (2020). *Unsupervised Data Augmentation for Consistency Training*. https://github.com/google-research/uda.

Xu, Y., Cao, H., Du, W., & Wang, W. (2022). A Survey of Cross-lingual Sentiment Analysis: Methodologies, Models and Evaluations. In *Data Science and Engineering* (Vol. 7, Issue 3, pp. 279–299). Springer. https://doi.org/10.1007/s41019-022-00187-3