# Global Sentiment and Emotion Analysis of COVID-19 Tweets

A Two-Stage Pipeline with Zero-Shot Emotion Tagging and Fine-Tuned DistilBERT Sentiment Classification

Alessandro Bottardi


## Abstract

An end-to-end pipeline is presented for large-scale analysis of COVID-19 tweets comprising cleaning, exploratory data analysis (EDA), zero-shot emotion tagging, fine-tuning of a compact transformer for sentiment analysis, and geo-spatial visualization. Using a publicly available dataset with train/test splits, the data distribution and textual features were characterized, a pre-trained DistilRoBERTa was applied for multi-class emotions (e.g., joy, fear, sadness, anger, neutral, disgust), and a five-way DistilBERT classifier (Extremely Negative $\rightarrow$ Extremely Positive) was fine-tuned and evaluated. Finally, predictions were aggregated by inferred country and rendered on interactive world maps.


# 1 Research Questions

**RQ1**: Which *sentiments* and *emotions* are most prevalent in COVID-19 related tweets across time and geography?
**RQ2**: To what extent can a compact transformer (DistilBERT) fine-tuned on five sentiment classes perform on noisy, short texts?
**RQ3**: How do emotion/sentiment distributions vary geographically when user-provided locations are aggregated to country level?

# 2 Methods

## 2.1 Data and Splits

The provided train/test CSVs (e.g., `Corona_NLP_train.csv`, `Corona_NLP_test.csv`) were used.
Each row contains tweet text and a sentiment label in {*Extremely Negative*, *Negative*, *Neutral*, *Positive*, *Extremely Positive*}. User location is given as a free-text field.

## 2.2 Preprocessing

A `clean_text` field was produced via lowercasing, whitespace normalization, URL and user mention removal, and light punctuation retention (e.g., “!”, “?”). No external geocoding was employed; instead, location strings were mapped to countries through rule-based normalization (accent stripping, token windows), city-to-country dictionaries covering major global cities, and aliases (e.g., “UK”, “U.S.”, US states/abbreviations), enabling offline execution.

## 2.3 Exploratory Data Analysis (EDA)

Beyond class counts, EDA computed text length, uppercase ratio, punctuation frequency, daily timelines were plotted. Unigrams and bigrams (overall and per class) were obtained using `CountVectorizer`.

## 2.4 Zero-Shot Emotion Tagging

Emotions were inferred on the test split using a pre-trained DistilRoBERTa classifier (e.g., joy, fear, sadness, anger, neutral, surprise, disgust) without labels. These predictions were later aggregated geographically to produce a world map of dominant emotions.

## 2.5 Fine-Tuning DistilBERT for Sentiment

**Architecture and notation.** A transformer encoder maps token embeddings $X \in \mathbb{R}^{n\times d}$ to contextual states via multi-head self-attention. For one head with query/key/value projections $Q = XW_Q$, $K = XW_K$, $V = XW_V$,

$$\mathrm{Attn}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V. \qquad (1)$$

The [CLS] representation was fed to a linear layer yielding logits $z \in \mathbb{R}^C$ over $C$=5 sentiment classes.

**Training objective.** Cross-entropy was optimized with label $y$:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{c=1}^{C} \mathbf{1}[y=c] \log \frac{e^{z_c}}{\sum_j e^{z_j}}. \qquad (2)$$

DistilBERT is a distilled variant of BERT; knowledge distillation typically combines supervised loss with a distillation term using teacher logits $t$ and temperature $\tau$:

$$\mathcal{L}_{\mathrm{KD}} = \mathrm{KL}\Big(\mathrm{softmax}(\tfrac{t}{\tau}) \,\Big\|\, \mathrm{softmax}(\tfrac{z}{\tau})\Big), \qquad (3)$$

$$\mathcal{L} = \alpha\,\mathcal{L}_{\mathrm{CE}} + \beta\,\mathcal{L}_{\mathrm{KD}} + \gamma\,\mathcal{L}_{\mathrm{cos}}. \qquad (4)$$

In this work, fine-tuning of the student model on labeled tweets was performed (teacher supervision was implicit from pre-training).

## 2.6 Evaluation

The held-out test split was used to compute accuracy, macro-F1, class-wise precision/recall/F1, and a confusion matrix. Zero-shot emotion distributions were compared to supervised sentiment distributions qualitatively via maps.

## 2.7 Geospatial Aggregation and Mapping

User locations were mapped to countries using normalization, dictionaries, and aliases, then predictions were aggregated per country, and dominant labels were computed. *Folium* was used to render circle markers at country centroids; marker radius scales with tweet count, and pop-ups show per-country label breakdowns.

# 3 Results

## 3.1 Descriptive Statistics

Figure 1 reports class balance by split. Length distributions and average words by class are shown in Figures 2 and 4.

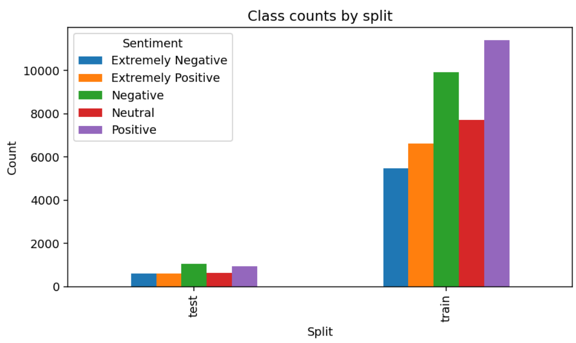


Figure 1: Class counts by split (train vs. test)

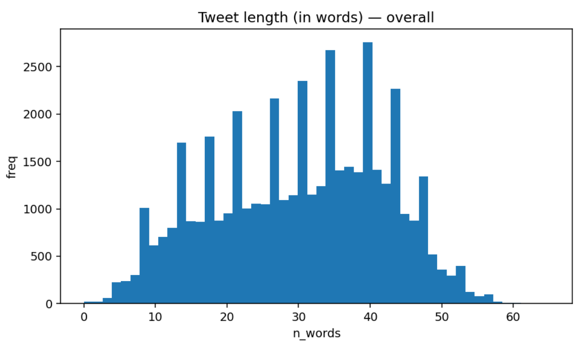


Figure 2: Tweet length histogram (words), overall

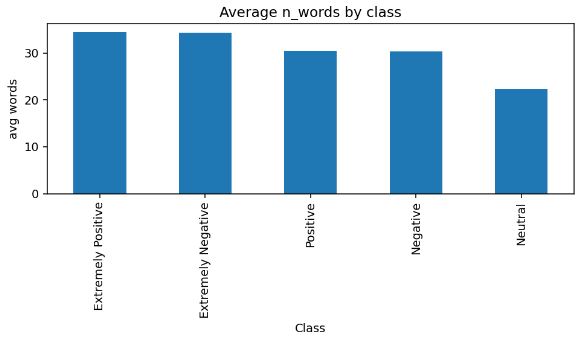


Figure 3: Average words per tweet by class

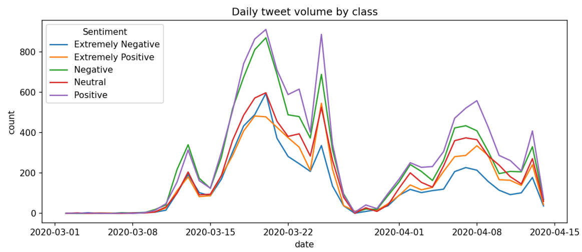


Figure 4: Daily tweet volume by class

## 3.2 Zero-Shot Emotions (DistilRoBERTa)

Emotions on the test set were predicted using a pre-trained DistilRoBERTa classifier. A global distribution and country-level dominant emotions are visualized in Figure 5 (fear appears as the dominant emotion across many countries).

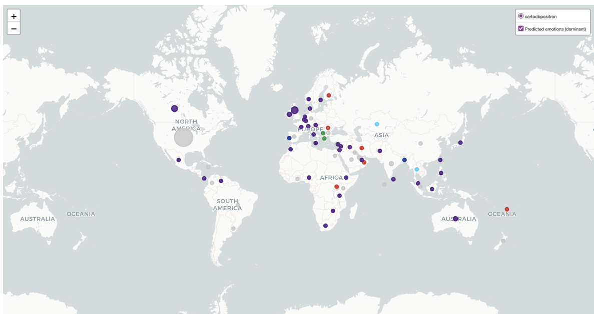


Figure 5: World map of dominant predicted emotions (zero-shot DistilRoBERTa).

## 3.3 Fine-Tuned Sentiment (DistilBERT)

Quantitative results for the five-way polarity classifier are summarized below. The confusion matrix is reported in Figure 6; per-class metrics and overall scores are listed in Table 1.

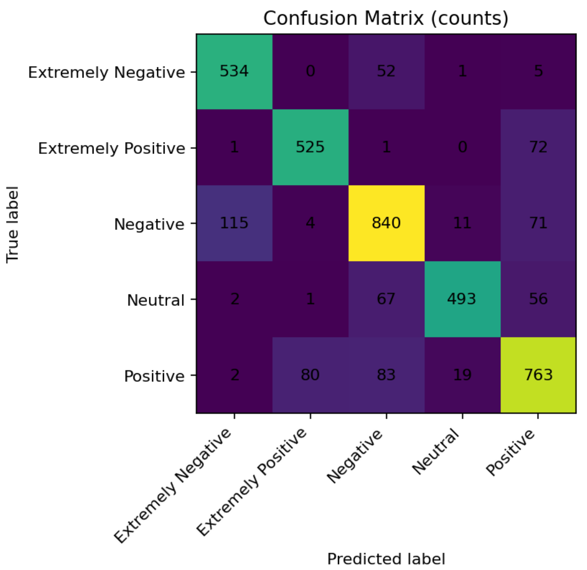


Figure 6: Confusion matrix for the five sentiment classes (counts).

Table 1: Evaluation on the test split (classification report).

| Class | Precision | Recall | F1 | Support |
|---|---|---|---|---|
| Extremely Negative | 0.82 | 0.90 | 0.86 | 592 |
| Extremely Positive | 0.86 | 0.88 | 0.87 | 599 |
| Negative | 0.81 | 0.81 | 0.81 | 1041 |
| Neutral | 0.94 | 0.80 | 0.86 | 619 |
| Positive | 0.79 | 0.81 | 0.80 | 947 |
| **Accuracy** | | **0.83** | | **3798** |
| Macro Avg | 0.84 | 0.84 | 0.84 | 3798 |
| Weighted Avg | 0.83 | 0.83 | 0.83 | 3798 |

### 3.4 Country-Level Maps (BERT)

Two maps were rendered for the fine-tuned DistilBERT pipeline: a map of *predicted* dominant sentiments and a map reflecting *ground-truth* dominant sentiments aggregated by country (Figures 7 and 8).

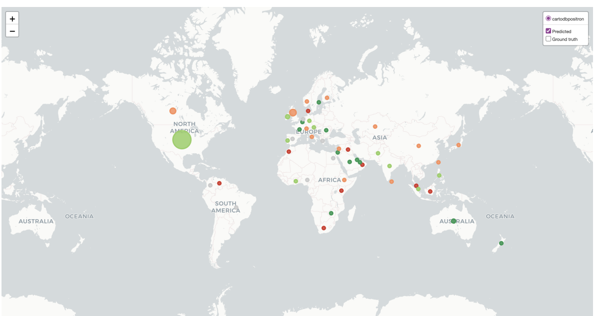

Figure 7: World map of dominant *predicted* sentiments (fine-tuned DistilBERT).

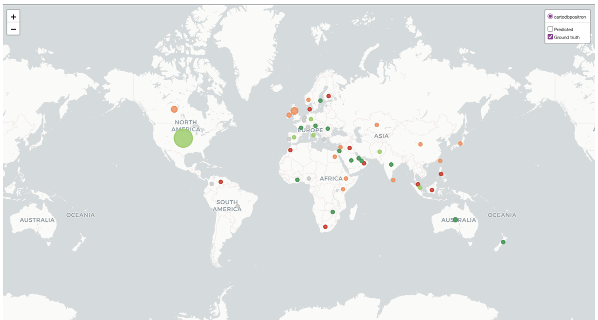

Figure 8: World map of dominant *ground-truth* sentiments aggregated by country.

## 4 Discussion and Limitations

**Emotion geography (zero-shot DistilRoBERTa)** From Figure 5, a clear predominance of *fear* is observable across many countries. This pattern is consistent with the first wave of COVID-19 (first half of 2020), during which uncertainty about transmission, changing public health guidance, and social medias likely evoked fear-related language. Exceptions are visible as pockets of *neutral* or *joy* where the local burden was relatively lower or communications emphasized containment and solidarity; isolated markers of *anger* and *sadness* plausibly map to grief, economic distress, or policy disputes.

**Sentiment geography (fine-tuned DistilBERT)** Figures 7 and 8 show broadly similar patterns between *predictions* and *ground truth*. Countries that experienced severe early outbreaks tend to show increased *Negative/Extremely Negative* dominance, whereas countries with relatively fewer cases or faster containment in March/April 2020 tend to exhibit *Neutral/Positive* dominance. For example, several locations in Asia, especially near China, display negative dominance, while Oceania and selected countries in Europe show more neutral/positive aggregates. Differences between predictions and ground truth appear primarily where sample sizes are small, where location strings are noisy and may be reassigned across borders, or where local idioms/sarcasm complicate polarity detection.

**Classifier performance and error structure** The overall accuracy (0.83) and macro-F1 (0.84) indicate fair performance for a compact transformer on short, noisy tweets. The confusion matrix (Figure 6) shows that most errors occur between *adjacent* polarities: *Extremely Negative* vs. *Negative* and *Extremely Positive* vs. *Positive*. *Neutral* exhibits asymmetrical confusions toward *Negative* and *Positive*, consistent with the nature of polarity in short texts. High precision for *Neutral* (0.94) alongside lower recall (0.80) suggests conservative assignment of neutrality (i.e., fewer false positives, more false negatives). The class supports (592–1041) indicate moderate balance, reducing the risk of majority-class bias.

**Correlations with 2020 trajectories (qualitative)** A qualitative association is visible between higher case/death burden and more negative sentiment, and conversely, between lower burden/early mitigation and neutral/positive sentiment. Instances where a country appears neutral/positive despite substantial case counts can be explained by (i) temporal aggregation over subperiods including recovery phases or policy relief, (ii) topical mixture (e.g., economic support or vaccine updates introducing positive framing), and (iii) demographic/sample biases of Twitter usage.

**Limitations** (i) *Location inference* is heuristic and relies on user-entered strings; without GPS, assignment noise is unavoidable. (ii) *Label noise* is present in crowd-labeled sentiment; sarcasm, figurative language, and domain slang remain challenging. (iii) *Temporal confounding* is likely during early 2020; emotions/sentiments fluctuate rapidly. (iv) *Construct mismatch* exists between emotion taxonomies (fear/joy/etc.) and five-way polarity, so divergences capture meaningful differences rather than model error alone. (v) *Sampling bias*: Twitter users are not representative, media cycles and bots can skew distributions. (vi) *Map rendering*: country centroids and dominance markers conceal intra-country variation; urban hotspots and within-country disparities are not visible. (vii) *Computational power*: DistilBERT was trained without the use of Virtual Machines nor additional GPUs, further analysis could be done increasing the number of epochs or by completely switching to BERT, which would create a much more complete model.

# 5 Conclusion

A modular, resource-efficient pipeline for COVID-19 Twitter analysis was implemented, covering enhanced EDA, zero-shot emotions with DistilRoBERTa, supervised five-way sentiment with fine-tuned DistilBERT, and geo-spatial aggregation. The emotion map indicates that *fear* was the most frequent emotion across many countries during the first half of 2020, with exceptions where neutral/positive affect was more prominent. The DistilBERT maps align qualitatively with contemporaneous trajectories: more severe national impact corresponds to more negative sentiment, while lower incidence or rapid containment corresponds to neutral/positive dominance. Discrepancies between predictions and ground truth are concentrated in low-sample or linguistically ambiguous settings, consistent with the observed error structure (adjacent-class confusions). Future work should incorporate temporal faceting (monthly maps), subnational choropleths, uncertainty visualization, and integration with epidemiological and policy timelines for more rigorous correlation analysis.

# References


[1] A. Vaswani *et al.*, "Attention is All You Need," in *NeurIPS*, 2017.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *NAACL*, 2019.

[3] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv:1910.01108*, 2019.

[4] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *EMNLP: Systems*, 2020.

[5] Folium authors, "Folium: Python Data, Leaflet.js Maps," `https://python-visualization.github.io/folium/`.

[6] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proc. SciPy*, 2010.

[7] D. Demszky *et al.*, "GoEmotions: A Dataset of Fine-Grained Emotions," in *ACL*, 2020.

[8] Aman Miglani, datatattle. *COVID-19 NLP Text Classification*. Kaggle dataset, 2020. Available at: `https://www.kaggle.com/datasets/datatattle/covid-19-nlp-text-classification`.