

Social Media Analysis 2.0

Customizing Data Analytics Tools to Accurately Measure
Sentiment



ENODO
GLOBAL

“Accurately measuring the sentiment of large volumes of social media data is critical to understanding Social Risk.”

Problem:

The distinct character and rhetoric of each social, political, and economic conversation across various social media platforms present unique challenges to accurately and consistently measure sentiment. This task requires novel techniques and tailored software tools to recognize the differences in both language and narrative structure. For example, the vocabulary of posts such as tweets significantly varies from writing used in other media such as newspapers, magazines, and blogs. Additionally, non-standard characters like emojis and emoticons combined with character limits of social media platforms prompt authors to express their views using a different blend of words than in more traditional and formal texts.

Sentiment analysis is also impacted by the argument style for a particular topic, which affects the choice of words and characters in a tweet. Posts from opposing sides in a heated Twitter debate, for example, may incorporate similar words, characters, emojis, and styles to support their view points but in critically different contexts. This argument style means the placement of a word relative to those surrounding it matters as much as the occurrence of the word itself. The context of social media messages is even more important for platforms like Twitter that constrain the number of characters that can be used to express an opinion.

Commercial-off-the-shelf (COTS) sentiment analysis tools fail to incorporate these key dynamics. They (1) use basic language models, (2) are not attuned to the nuances of varying social media platforms, and (3) provide customers few, if any, options to customize the models. While these easy-to-use tools have formidable processing power, they lead to poor results when applied to data they were not designed to understand. This problem is magnified when social media is integrated with other data sources in follow-on analysis where the poor sentiment measures propagate across other data sources and bias metrics.

Solution:

To address the shortcomings of social media sentiment analysis, ENODO Global developed a technology that combines best-in-class computational techniques and the most appropriate data to learn from the constantly evolving language of social media. The technology uses sentiment analysis classifiers that are optimized to the specific social media platforms, languages, and topics. It delivers results that are significantly more accurate than generic sentiment analysis platforms.

ENODO analysts start by identifying and collecting data that most appropriately captures the online conversations of interest. Working with clients, we conduct a preliminary investigation on a subset of this data to identify the key sentiments and topics—in the native language—of the social media posts. This approach produces a set of “training” and “testing” data that consists of posts whose sentiments towards the explicit or implied topic are categorized by native speakers. This process ensures that the subsequent models are built on accurate data and rooted in specific social and linguistic issues of the problem at hand.



“ENODO designed a sentiment analysis tool specifically tailored for Arabic language Twitter posts.”

ENODO then uses the training data to design, implement, and refine a computational model of the collected texts in the native, untranslated language of the posts. This iterative procedure uses natural language processing (NLP) tools to process the data, applies deep-learning techniques to extract the most salient features from processed data, and combines a variety of machine learning techniques to create a predictive model of sentiments. We continually refine these processes to optimize the predictive accuracy of the model, which is validated across different statistical metrics. To further enhance sentiment accuracy, we continually update the classification model as new language and narratives emerge on social media.

COTS sentiment tools employ relatively static models based on snapshots of social media language from a set time and ignore or minimize new words, characters, and structures. That approach works for traditional media. However, it fails to incorporate the dynamics of living social media conversations that make those platforms such attractive and powerful methods of communications. We constantly “retrain” our models to match the evolving nature of online debates and the priorities of our clients. This enables us to incorporate new vocabulary and narrative structures into our analysis, which enables our models to learn from new data. Updated models continue to refine our technology to ensure that the accuracy of our approach keeps pace with the speed of change of social media platforms and conversations.

Case Study:

To demonstrate the power of ENODO’s sentiment analysis, we compared the results of our model to IBM’s AlchemyAPI sentiment classifier. The case study analyzed a sample set of over 10,000 Arabic tweets developed by academic researchers. Native Arabic speakers categorized the sentiments of tweets as positive, negative, neutral, or observational. Although the tweets originated from the same language and social media platform, the topics they covered varied widely. Even with this added difficulty of subject variability, ENODO’s model significantly outperformed the IBM alternative.

Table 1 illustrates how ENODO’s customized sentiment analysis technology outperformed the IBM tool. The cells in the table represent the percentage of tweets whose sentiments was correctly predicted by each model. The ENODO technology was on average 30% percent more accurate than the IBM alternative across all tweets. This improvement was particularly pronounced for negative tweets, where the ENODO technology was ten times more accurate than IBM’s tool. Even for positive and neutral tweets, the accuracy gains realized by ENODO’s technology were robust.



“ENODO’s technology significantly outperformed IBM’s sentiment analysis tool.”

	IBM Accuracy	ENODO Accuracy	Improvement
All Tweets	49%	79%	30%
Negative Tweets	8%	80%	72%
Neutral Tweets	58%	79%	21%
Positive Tweets	56%	70%	14%

Table 1 – Sentiment Classifier Accuracy

The increased accuracy of ENODO’s technology is the result of training our models on actual tweets combined with using a word-embedding technique that accounts for word order. This approach worked especially well for negative tweets in the test data that used a broad range of narrative styles, which required fine-tuned models to classify their subtler statements. Examples of negative tweets that ENODO accurately classified and the IBM tool misclassified are illustrated in Table 2.

Negative Sentiment Arabic Tweet and Language Translation	ENODO	IBM
نحن امام انهيار لشرعية المجلس الأعلى للقوات المسلحة We are facing a collapse of the legitimacy of the Supreme Council of the Armed Forces	Negative	Positive
محمود_سعد اعلامي قدر بعد اذاعة فيديو خيانة #حمددين قال لم اسمعه جيدا ولم يعلق... Mahmoud Saad is a filthy journalist after broadcasting Hamdeen's betrayal video and said he didn't hear well and didn't comment on it. He is not a journalist, he is an...	Negative	Neutral

Table 2 – Example Negative Tweet Classifications

Conclusion:

ENODO’s technology significantly increases sentiment analysis accuracy for Twitter posts and can be applied to various social media sites. It provides just one example of how we combine advanced technology and new forms of data through a social dynamics lens, to measure and forecast how different demographic groups perceive and debate critical topics. This dynamic approach to sentiment analysis produces accurate, in-depth insights into current perceptions on social, economic, and political debates. It can be used to develop a baseline assessment to forecast changes in perceptions and can be adapted to uncover how social identities shape the narratives and which topics will resonate across different social groups. ENODO’s technology enables organizations that are impacted by public perception to harness large amounts of online data and accurately interpret the social pulse. Moreover, it helps clients to proactively shape online conversations and perceptions of their target audience before making strategic decisions.





ENODO
GLOBAL

ENODOGLOBAL.COM | INFO@ENODOGLOBAL.COM

This electronic representation of ENODO Global, Inc. intellectual property is provided for non-commercial use only. Unauthorized posting of ENODO electronic documents to a non-ENODO website is prohibited and are protected under copyright law. ENODO retains all rights to the information contained in this document and assumes no liability or fault for any material contained in this document, or derived from its subsequent use, whether directly attributed, implied, or inferred by any users of this submission.