

Sentiment Lab — Twitter Sentiment Classification Baseline

A transparent first baseline for tweet sentiment, with its gaps named.

62%

test accuracy on 61 tweets

302

positive and negative tweets used

1,661

TF-IDF features

BUILT WITH

NLP

Sentiment Analysis

Scikit-learn

TF-IDF

NLTK

Python

Text Classification

How it works



Key decisions

01

A baseline on purpose

TF-IDF with Naive Bayes is fast and easy to read. Its value is showing what a better model has to beat.

02

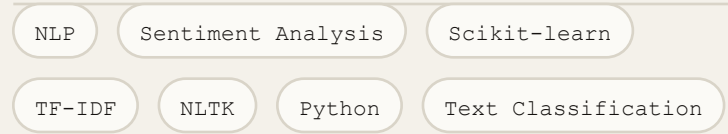
Binary scope, stated as a simplification

Dropping Neutral and Irrelevant made the first pass tractable, at the cost of two of the four real classes.



Stack, status and what is next

BUILT WITH



KNOWN LIMITS

- The validation file was loaded but never evaluated.
- Trained on about 300 rows, not the full dataset.
- Neutral and Irrelevant tweets are dropped, not handled.

NEXT STEPS

- Evaluate on the validation set.
- Train on the full dataset.
- Handle all four classes.

SOURCE

<https://github.com/wailethewave/Sentiment-Lab-NLP>

