

# Sentimental Analysis Report

Using Natural Language Toolkit

TEAM MEMBERS :

- RENU BOJJA
- KARMISHTHA PATNAIK

## TABLE OF CONTENTS

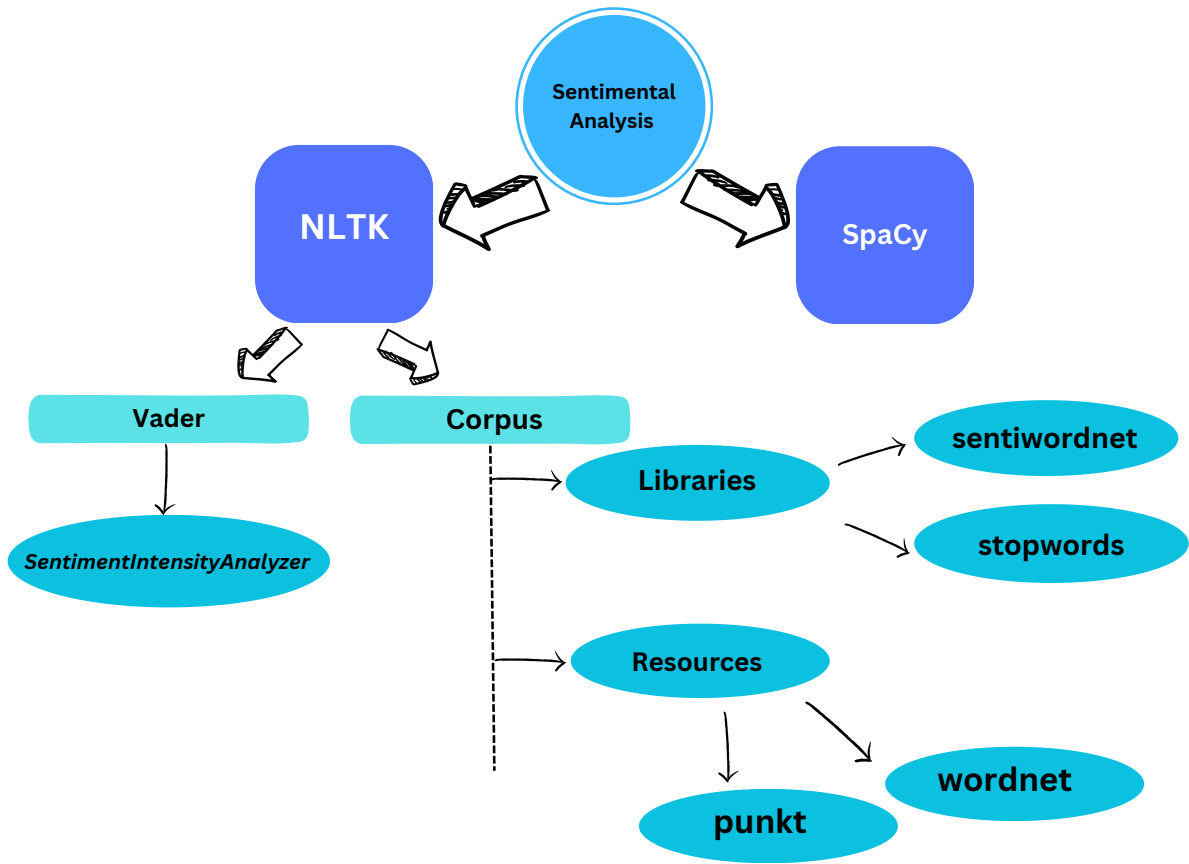
1. INTRODUCTION
2. FLOWCHART ON SENTIMENTAL ANALYSIS
3. NLTK Vs SPACY
4. WHY NLTK OVER SPACY?
5. NATURAL LANGUAGE TOOLKIT (NLTK)
6. GRAPHS: VADER Vs CORPUS
7. WHY CORPUS OVER VADER?
8. WHICH IS THE BEST NLP LIBRARY?
9. CONCLUSION
10. REFERENCES

## INTRODUCTION

This report explores how people express emotions on social media using two tools: VADER and CORPUS. We'll compare these tools to see how well they understand our feelings in tweets and posts.

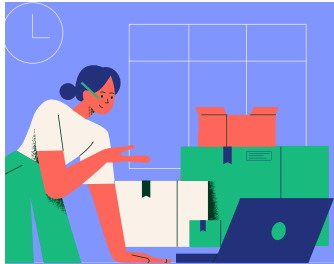
Comparative Analysis Libraries	Target
<ul style="list-style-type: none"><li>• VADER</li><li>• CORPUS</li></ul>	To comprehensively assess and compare the efficacy of two libraries, followed by their practical application in refining social media posts.

## FLOWCHART ON SENTIMENTAL ANALYSIS



# NLTK VS SPACY


NLTK	SpaCy
<p>NLTK (Natural Language Toolkit): NLTK is a Python library used for natural language processing tasks. It provides a wide range of tools and resources for tasks like tokenization, stemming, tagging, and sentiment analysis.</p> <p>NLTK offers the <code>word_tokenize()</code> function to split a sentence into words.</p>	<p>SpaCy is another Python library for NLP that focuses on efficiency and production-ready capabilities. It provides fast tokenization, named entity recognition, dependency parsing, and other linguistic annotations.</p> <p>SpaCy's tokenization with sentence parsing:</p>
<pre>from nltk.sentiment import SentimentIntensityAnalyzer nltk.download('vader_lexicon')  # Load the VADER sentiment analyzer sid = SentimentIntensityAnalyzer()  # Sample text text = "NLTK is an amazing library for NLP!"  # Sentiment analyzer from NLTK sid = SentimentIntensityAnalyzer() sentiment_scores = sid.polarity_scores(text)  print(sentiment_scores)</pre> <p><code>{'neg': 0.0, 'neu': 0.595, 'pos': 0.405, 'compound': 0.6239}</code> [nltk_data] Downloading package vader_lexicon to /root/nltk_data...</p> <p>NLTK's advantage in providing a dedicated sentiment analysis tool compared to SpaCy's default functionalities, where sentiment analysis might require more manual implementation or integration with other libraries or models.</p>	<pre>import spacy  # Sample text text = "SpaCy is a powerful NLP library."  # Load SpaCy's English model nlp = spacy.load("en_core_web_sm")  # Analyzing sentiment using SpaCy (which doesn't have a built-in sentiment analysis) doc = nlp(text)  # Extracting sentiment based on rules (example: using the presence of certain words) positive_words = ['powerful', 'awesome', 'amazing', 'great'] negative_words = ['bad', 'terrible', 'awful', 'horrible']  sentiment_score = 0  for token in doc:     if token.text.lower() in positive_words:         sentiment_score += 1     elif token.text.lower() in negative_words:         sentiment_score -= 1  if sentiment_score &gt; 0:     print("Positive sentiment detected!") elif sentiment_score &lt; 0:     print("Negative sentiment detected!") else:     print("Neutral sentiment detected!")</pre> <p>Positive sentiment detected!</p> <p>SpaCy doesn't have a direct sentiment analysis module like NLTK's Vader. While SpaCy excels in other NLP tasks, NLTK's Vader provides specific sentiment intensity analysis with pre-trained lexicons, making it more straightforward for sentiment analysis purposes in certain scenarios.</p>



# Why NLTK over SpaCy ?

- Learning Curve and Accessibility:** NLTK's step-by-step tutorials make it simple for beginners to understand sentiment analysis. For instance, its sentiment analyzer "Vader" comes with clear usage examples.
- Customization Options:** NLTK allows easy adjustments to sentiment analysis models. For example, you can fine-tune lexicons or algorithms to better fit specific data or project requirements.
- Rich Linguistic Resources:** With NLTK, access extensive corpora like the Brown Corpus, aiding sentiment analysis with a diverse range of text samples for better accuracy.

# NATURAL LANGUAGE TOOLKIT(NLTK)

TOOLS	CLASS / LIBRARY
<div data-bbox="344 254 458 296">VADER</div> <div data-bbox="111 327 698 443"><ol style="list-style-type: none"><li>1. Vader stands for <b>Valence Aware Dictionary and sEntiment Reasoner</b>.</li><li>2. Pre-trained sentiment analysis tool</li></ol></div> <div data-bbox="318 468 489 600"></div> <div data-bbox="237 625 575 663">specifically for social media</div> <div data-bbox="99 695 706 774"><p>3.Vader can analyze the sentiment of a piece of text and determine</p></div> <div data-bbox="121 812 509 890"><ul style="list-style-type: none"><li>• Detect Emotion</li><li>• Sentiment intensity score</li></ul></div>	<div data-bbox="805 237 1279 279">SentimentIntensityAnalyzer</div> <div data-bbox="753 310 1328 459"><ol style="list-style-type: none"><li>1. The SentimentIntensityAnalyzer is a class within NLTK's Vader module.</li><li>2.It evaluates the text by considering the :</li></ol></div> <div data-bbox="776 495 1265 810"><ul style="list-style-type: none"><li>• Presence of these words</li><li>• Their intensity</li><li>• Punctuation</li><li>• Capitalization</li><li>• Degree modifiers</li><li>• Emoticons to determine the sentiment expressed in the given text.</li></ul></div> <div data-bbox="753 837 1326 913"><pre>import pandas as pd import nltk from nltk.sentiment.vader import SentimentIntensityAnalyzer</pre></div>
<div data-bbox="308 1029 436 1066">CORPUS</div> <div data-bbox="111 1100 679 1180"><ol style="list-style-type: none"><li>1. <b>Module:</b> A corpus is a big collection of text or speech data.</li></ol></div> <div data-bbox="99 1285 309 1323"><p>2.Components:</p></div> <div data-bbox="121 1356 661 1570"><ul style="list-style-type: none"><li>• <b>Written documents</b></li><li>• <b>Articles</b></li><li>• <b>Books</b></li><li>• <b>Conversations</b></li><li>• <b>Or any language-related material</b></li></ul></div> <div data-bbox="99 1675 354 1713"><p>3.Structured Data:</p></div> <div data-bbox="121 1747 639 1906"><ul style="list-style-type: none"><li>• categorized</li><li>• tagged</li><li>• labeled to enable easier analysis and processing</li></ul></div> <div data-bbox="99 1938 682 2018"><ol style="list-style-type: none"><li>4.Used for Analysis: Linguists, researchers, and developers use corpora to study</li></ol></div>	<div data-bbox="926 1029 1156 1066">SENTIWORDNET</div> <div data-bbox="753 1100 1322 1314"><p>Each word in SentiWordNet is linked to specific synsets (groups of synonymous words representing a distinct concept or meaning). For each synset, SentiWordNet assigns sentiment scores</p></div> <div data-bbox="951 1421 1132 1459">STOPWORDS</div> <div data-bbox="753 1493 1303 1707"><p><b>Common Words:</b> Stopwords refer to the most frequently occurring words in a language that are often removed from text during natural language processing tasks.</p></div> <div data-bbox="753 1738 1319 1953"><p><b>Non-Contextual Words:</b> These words, like "and", "the", "is", "in", etc., don't carry specific meaning or contribute significantly to the understanding of the text's content or context.</p></div>

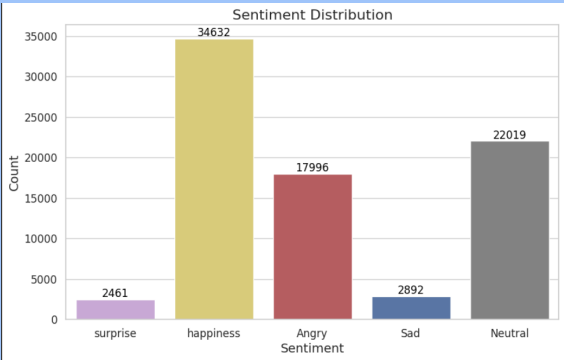
language patterns, develop language models, test algorithms, and improve natural language processing tools.

```
import pandas as pd
from nltk.corpus import sentiwordnet as swn
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import nltk
```

## GRAPHS: VADER VS CORPUS

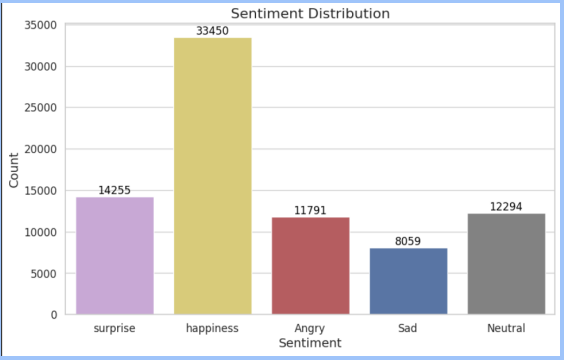
# VADER

## TWITTER

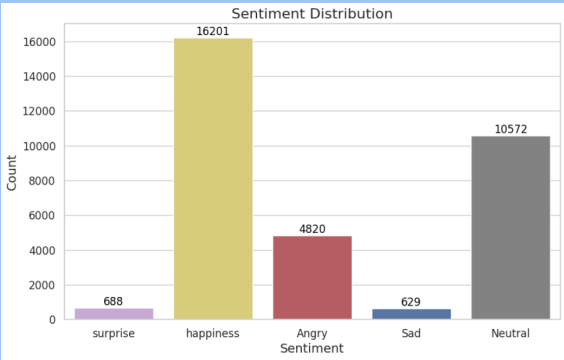


# CORPUS

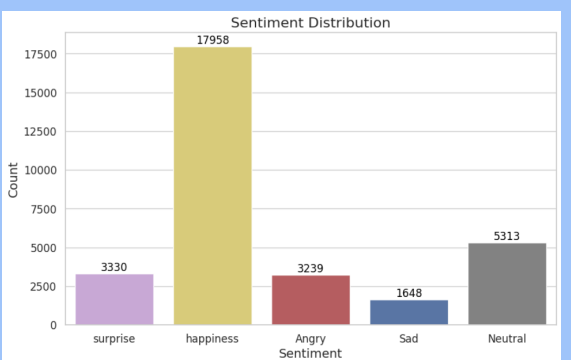
## TWITTER



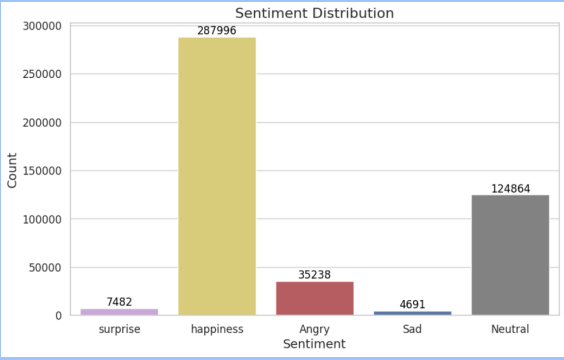
## THREADS



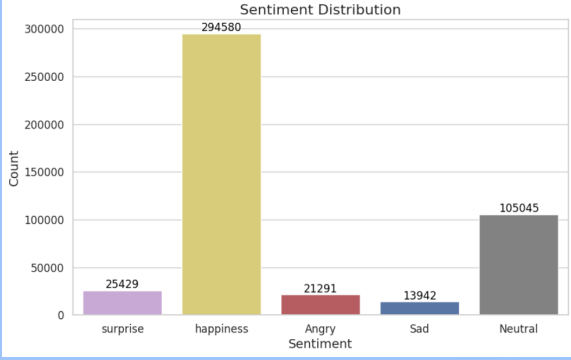
## THREADS

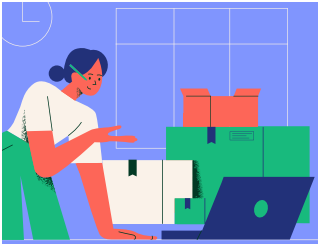


## TIKTOK



## TIKTOK



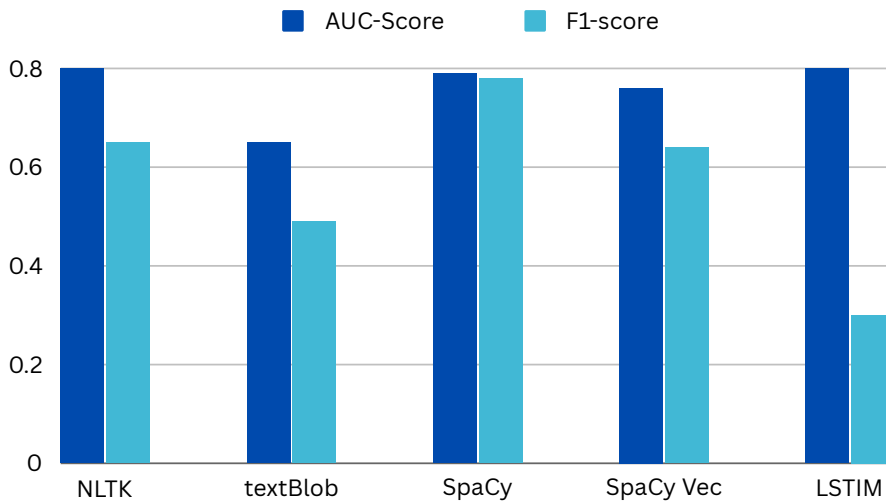


# Why Corpus over Vader?

1. **Contextual Understanding:** Corpus analysis allows for a deeper understanding of sentiment in context. It can capture nuanced emotions specific to certain topics or domains, which pre-defined lexicons like Vader might miss.
2. **Customization:** Corpora can be tailored or curated for specific projects or industries, providing more accurate sentiment analysis for domain-specific content compared to a generalized tool like Vader.
3. **Fine-Grained Analysis:** A corpus enables detailed analysis by considering varied sentiments expressed within different contexts, offering a more comprehensive view than the aggregated scores provided by Vader.
4. **Training Custom Models:** With a corpus, you can train custom sentiment analysis models using machine learning techniques, potentially outperforming general-purpose tools like Vader by adapting specifically to your dataset.

## WHICH IS THE BEST NLP LIBRARY?





## CONCLUSION

In our study, we explored how emotions play out on social media using VADER and CORPUS. By comparing these tools, we wanted to see how well they understood the emotions in tweets and posts. The results not only revealed how good these tools are at grasping sentiments but also highlighted the importance of choosing the right method for understanding the intricate world of digital emotions.

## REFERENCES

1. Guide on Sentimental Analysis:  
<https://monkeylearn.com/sentiment-analysis/>
2. Best libraries in NLP:  
<https://sunscrapers.com/blog/9-best-python-natural-language-processing-nlp/>