**CONTENT**

**1. INTRODUCTION TO SENTIMENTAL ANALYSIS**

**2. INTRODUCTION TO PYTHON**

**3. ANALYSIS FLOWCHART**

**4. WEB AUTOMATION**

**5. DATA VISUALIZATION**

**6. ADVANTAGE AND DISADVANTAGE**

**7 DIFFICULTY FACED**

**8. APPLICATIONS**

**9. DECLARATION**

**INTRODUCTION TO SENTIMENTAL ANALYSIS**

**Sentiment analysis** (sometimes known as **opinion mining** or **emotion AI**) refers to the use of natural language processing, [text analysis](https://en.wikipedia.org/wiki/Text_analytics), [computational linguistics](https://en.wikipedia.org/wiki/Computational_linguistics), and [biometrics](https://en.wikipedia.org/wiki/Biometrics) to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to [voice of the customer](https://en.wikipedia.org/wiki/Voice_of_the_customer) materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from [marketing](https://en.wikipedia.org/wiki/Marketing) to [customer service](https://en.wikipedia.org/wiki/Customer_relationship_management) to clinical medicine.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation , affective state (that is to say, the emotional state of the author or speaker), or the intended emotional communication (that is to say, the emotional effect intended by the author or interlocutor).

## Types

A basic task in sentiment analysis is classifying the *polarity* of a given text at the document, sentence, or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry", "sad", and "happy".

First steps to bringing together various approaches -- learning, lexical, knowledge-based, etc -- were taken in the 2004 [AAAI](https://en.wikipedia.org/wiki/AAAI) Spring Symposium where linguists, computer scientists, and other interested researchers first aligned interests and proposed shared tasks and benchmark data sets for the systematic computational research on affect, appeal, subjectivity, and sentiment in text.[[1]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-1) Early methods such as the General Inquirer [[2]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-2) often inspired by research in content analysis for psychology [[3]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-3) or somewhat later implementations such as the system described in a patent by Volcani and Fogel,[[4]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-4) identify individual words and phrases in text with respect to different emotional scales. Their system presents synonyms that can be used to increase or decrease the level of evoked emotion in each scale. Later efforts turned to a more polar view of sentiment, from positive to negative, such as work by Turney,[[5]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-Turney02-5) and Pang[[6]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-PangAl02-6) who applied different methods for detecting the polarity of product reviews and movie reviews respectively. This work is at the document level. One can also classify a document's polarity on a multi-way scale, which was attempted by Pang[[7]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-PangLee05-7) and Snyder[[8]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-SnyderBarzilay07-8) among others: Pang and Lee[[7]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-PangLee05-7) expanded the basic task of classifying a movie review as either positive or negative to predict star ratings on either a 3- or a 4-star scale, while Snyder[[8]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-SnyderBarzilay07-8) performed an in-depth analysis of restaurant reviews, predicting ratings for various aspects of the given restaurant, such as the food and atmosphere (on a five-star scale).

Even though in most statistical classification methods, the neutral class is ignored under the assumption that neutral texts lie near the boundary of the binary classifier, several researchers suggest that, as in every polarity problem, three categories must be identified. Moreover, it can be proven that specific classifiers such as the [Max Entropy](https://en.wikipedia.org/wiki/Maximum_entropy_probability_distribution)[[9]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-Vryniotis13-9) and the [SVMs](https://en.wikipedia.org/wiki/Support_vector_machine)[[10]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-KoppelSchler06-10) can benefit from the introduction of a neutral class and improve the overall accuracy of the classification. There are in principle two ways for operating with a neutral class. Either, the algorithm proceeds by first identifying the neutral language, filtering it out and then assessing the rest in terms of positive and negative sentiments, or it builds a three-way classification in one step.[[11]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-11) This second approach often involves estimating a probability distribution over all categories (e.g. [naive Bayes](https://en.wikipedia.org/wiki/Naive_Bayes_classifier) classifiers as implemented by Python's [NLTK](https://en.wikipedia.org/wiki/Nltk) kit). Whether and how to use a neutral class depends on the nature of the data: if the data is clearly clustered into neutral, negative and positive language, it makes sense to filter the neutral language out and focus on the polarity between positive and negative sentiments. If, in contrast, the data are mostly neutral with small deviations towards positive and negative affect, this strategy would make it harder to clearly distinguish between the two poles.

A different method for determining sentiment is the use of a scaling system whereby words commonly associated with having a negative, neutral, or positive sentiment with them are given an associated number on a −10 to +10 scale (most negative up to most positive) or simply from 0 to a positive upper limit such as +4. This makes it possible to adjust the sentiment of a given term relative to its environment (usually on the level of the sentence). When a piece of unstructured text is analyzed using [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), each concept in the specified environment is given a score based on the way sentiment words relate to the concept and its associated score.[[12]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-12) This allows movement to a more sophisticated understanding of sentiment, because it is now possible to adjust the sentiment value of a concept relative to modifications that may surround it. Words, for example, that intensify, relax or negate the sentiment expressed by the concept can affect its score. Alternatively, texts can be given a positive and negative sentiment strength score if the goal is to determine the sentiment in a text rather than the overall polarity and strength of the text.

### Subjectivity/objectivity identification

This task is commonly defined as classifying a given text (usually a sentence) into one of two classes: objective or subjective.[[14]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-PangLee08Subjectivity-14) This problem can sometimes be more difficult than polarity classification.[[15]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-MihalceaAl07-15) The subjectivity of words and phrases may depend on their context and an objective document may contain subjective sentences (e.g., a news article quoting people's opinions). Moreover, as mentioned by Su,[[16]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-SuMarkert08-16) results are largely dependent on the definition of subjectivity used when annotating texts. However, Pang[[17]](https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-PangLee04-17) showed that removing objective sentences from a document before classifying its polarity helped improve performance.

### Feature/aspect-based

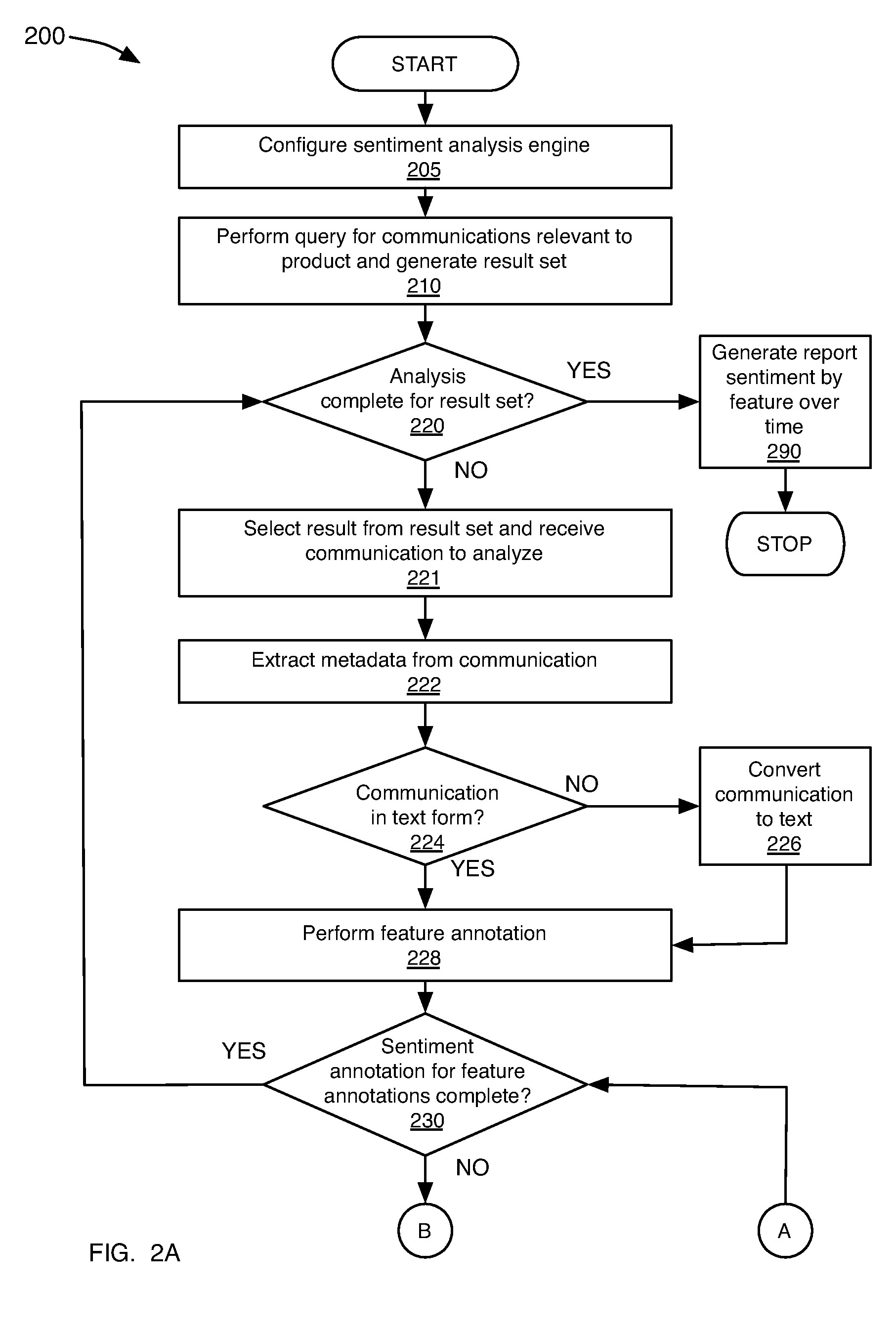
It refers to determining the opinions or sentiments expressed on different features or aspects of entities, e.g., of a cell phone, a digital camera, or a bank. A feature or aspect is an attribute or component of an entity, e.g., the screen of a cell phone, the service for a restaurant, or the picture quality of a camera. The advantage of feature-based sentiment analysis is the possibility to capture nuances about objects of interest. Different features can generate different sentiment responses, for example a hotel can have a convenient location, but mediocre food. This problem involves several sub-problems, e.g., identifying relevant entities, extracting their features/aspects, and determining whether an opinion expressed on each feature/aspect is positive, negative or neutral. The automatic identification of features can be performed with syntactic methods or with [topic modeling](https://en.wikipedia.org/wiki/Topic_model). More detailed discussions about this level of sentiment analysis can be found in Liu's work

**INTRODUCTION TO PYTHON**

**Python** is a [widely used](https://en.wikipedia.org/wiki/Measuring_programming_language_popularity) [high-level programming language](https://en.wikipedia.org/wiki/High-level_programming_language) for [general-purpose programming](https://en.wikipedia.org/wiki/General-purpose_programming_language), created by [Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum) and first released in 1991. An [interpreted language](https://en.wikipedia.org/wiki/Interpreted_language), Python has a design philosophy which emphasizes code [readability](https://en.wikipedia.org/wiki/Readability) (notably using [whitespace](https://en.wikipedia.org/wiki/Whitespace_character) indentation to delimit [code blocks](https://en.wikipedia.org/wiki/Code_block) rather than curly braces or keywords), and a syntax which allows programmers to express concepts in fewer [lines of code](https://en.wikipedia.org/wiki/Source_lines_of_code) than possible in languages such as [C++](https://en.wikipedia.org/wiki/C%2B%2B) or [Java](https://en.wikipedia.org/wiki/Java_(programming_language)) The language provides constructs intended to enable writing clear programs on both a small and large scale

Python features a [dynamic type](https://en.wikipedia.org/wiki/Dynamic_type) system and automatic [memory management](https://en.wikipedia.org/wiki/Memory_management) and supports multiple [programming paradigms](https://en.wikipedia.org/wiki/Programming_paradigm), including [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming), [imperative](https://en.wikipedia.org/wiki/Imperative_programming), [functional programming](https://en.wikipedia.org/wiki/Functional_programming), and [procedural](https://en.wikipedia.org/wiki/Procedural_programming) styles. It has a large and comprehensive [standard library](https://en.wikipedia.org/wiki/Standard_library)

Python interpreters are available for many [operating systems](https://en.wikipedia.org/wiki/Operating_system), allowing Python code to run on a wide variety of systems. [CPython](https://en.wikipedia.org/wiki/CPython), the [reference implementation](https://en.wikipedia.org/wiki/Reference_implementation) of Python, is [open source](https://en.wikipedia.org/wiki/Open_source) softwareand has a community-based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit [Python Software Foundation](https://en.wikipedia.org/wiki/Python_Software_Foundation).

**ANALYSIS SEQUENCE**

**WEB AUTOMATION**

Scrapping Website and copying all necessary Xpath

Scrolling down the page till the information exsist

Downloading all the outer HTML script

Saving the html script to txt file and emojizing the script

Extracting the important data points and information

Displaying the needed information

**DATA VISUALIZATION**

Converting the saved text file into Dataframe

Pre-process the Dataframe

Creating and saving image captions,likes,coments and time

Creating and saving bar or line graph of likes vs time, comment vs time etc.

Creating candle stick graph of required time,date,likes,comments

**DISADVANTAGE**

1) Medium Accuracy.

2) Need high RAM for processing.

3) Slower internet connections.

5) Need large storage.

**ADVANTAGE**

1) Time saving.

2) Super fast when it comes to analysis.

3) Amazing future scope.

4) Makes the search easy.

**DIFFICULTY FACED**

1) Low computer configuration making work slower.

2) Too many wrong approaches.

3) Learning sentimental analysis without guidance was a tuff task.

4) Choosing a proper sentimental analysis algorithm was a tuff task.

-Every time I have to train and test different algorithm it took almost 1-3 minutes.

**APPLICATIONS**

* Allows much more detailed sentiment analytics on each entity
* Addresses cases where multiple, conflicting sentiments are expressed within a single document or sentence
* Allows more in-depth analysis of opinions and actions
* Enables organizations to offer more appropriate, prioritized responses

DECLARATION

**Declaration**

**DECLARATION (WITH SIGNATURE OF STUDENTS)**