Sentiment Analysis

Zynab AlShareef

December 3rd, 2020

In recent times, with the flow of data on the Internet, especially on social media, we often hear about Sentiment Analysis using machine learning language, and we notice that a group of researchers are heading to this field, so what is Sentiment Analysis? How is it done? What is its importance? In this article we will take a simple overview of this field.

What is sentiment analysis?

Sentiment analysis is a process of analyzing people's feelings and opinions through what they write using natural language processing (NLP) and machine learning techniques and classify them into is positive, negative or neutral. It is not easy for human to do that, also, it is not easy for computer too, but with right tools and Python we can use sentiment analysis to better understand the sentiment of a piece of writing.

Let us look at few example:

1-Saudi Digital Academy is by far the best company I have ever dealt with.
This sentence clearly express a positive opinion determine by “best”.
2-Tomorrow SDA and MCIT will have an event
In this case, we only have a factual statement about “SDA” and “MCIT”. Its polarity is neutral.
3-The weather is bad today
This sentence clearly express a negative opinion determine by “bad”.

Why sentiment analysis is important?

Sentiment analysis is becoming a more and more popular topic as artificial intelligence, machine learning and natural language processing technologies that are blossoming these days. Sentiment analysis is of great importance because it helps organizations and brand owners measure the satisfaction of their customers with a specific product by using feedback analysis in a faster, more comprehensive and more reliable manner than traditional questionnaires. Sentiment analysis determines the positive and negative opinions about a services that provide by companies to their customers.

Steps of Sentiment Analysis:

Using Natural Language Processing to Preprocess and Clean Text Data
1-Collecting Data: collecting data from its source and creating a Dataset
2-Reprocessing: deleting the symbols and unwanted numbers, or letters of other languages
3-Filtering:It includes several things: Spelling correction (Misspelling), delete Repeated Letters, Eliminate the words (Stop Words), which are words that do not indicate a meaning in themselves, such as: pronouns / nouns and nouns connected / prepositions and accusative ... etc. normalizing by replacing some letters with similar letters, and deleting punctuation marks; the

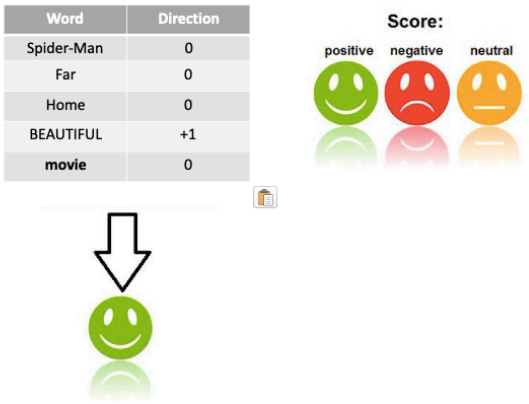
goal is for the text to appear in a single form that is easy to find in the dictionary used in the classification process.
5-Classifying by Machine Learning: the process of classifying texts into positive, negative, or neutral in some cases.
6-Evaluation metrics:As a classification problem, Sentiment Analysis uses the evaluation metrics of Precision, Recall, F-score, and Accuracy
7-Visualize Results: To visualize the results of Sentiment Analysis, many people employ well-known techniques, such as graphs, histograms, and confusion matrices.

How Sentiment Analysis Works?

Basic sentiment analysis of text documents follows a straightforward process:
1-Tokenizing sentences to break text down into sentences, words, or other units.
2-Normalizing words: It entails condensing all forms of a word into a single representation of that word. For instance, “Watched,” “watching,” and “watches” can all be normalized into “watch,” by condensing all forms of a word into a single form.
3-Removing stop words: Stop words are words that may be important in human communication but are of little value for machines. Like “if,” “but,” “or,” and so on.
4-Vectorizing text by turning the text into a numerical representation and Assign a sentiment score to each phrase and component (-1 to +1).
Here is an example of sentiment analysis by human brains . An example of applying data preprocessing steps to a tweet from Twitter that represents the opinion of someone attending a specific movie, these steps are important to extract the opinion of the tweet in the film. To find out the opinion of the tweet, the words are displayed on a dictionary that shows the direction of each word (+1: positive, -1: negative, 0: normal). Then the score of the tweet is calculated.

Step	Explanation	Example
Data Collection	Lets say we have a tweet about specific kind of movie. (Spiderman)	Spider-Man: Far From Home is such a BEAUTIFUL **movie**!
Tokenization	Dividing the sentence to many words - moving from sentence level to word level and this is simplify the processing	Spider-Man: Far From Home is such a BEAUTIFUL **movie**!
Normalization	It aims at unification by removing the appendages and spurs, this step facilitates treatment and is considered a stage of preparation to stemming	Spider-Man: Far - From - Home - is - such - a - BEAUTIFUL - **movie**!
Stop Word Filtering	Remove letters that do not have any meaning, such as prepositions	Spider-Man~~:~~ Far ~~From~~ Home ~~is such a~~ BEAUTIFUL **movie**~~!~~
Stemming	It aims to return the word to its origin and is considered the most difficult type of processing	Spider-Man Far Home BEAUTIFUL **movie**

Analysis of the Tweet words “Spider-Man Far Home BEAUTIFUL movie” shows that the opinion of the tweet is positive due to the presence of the word "BEAUTIFUL" as the picture shows



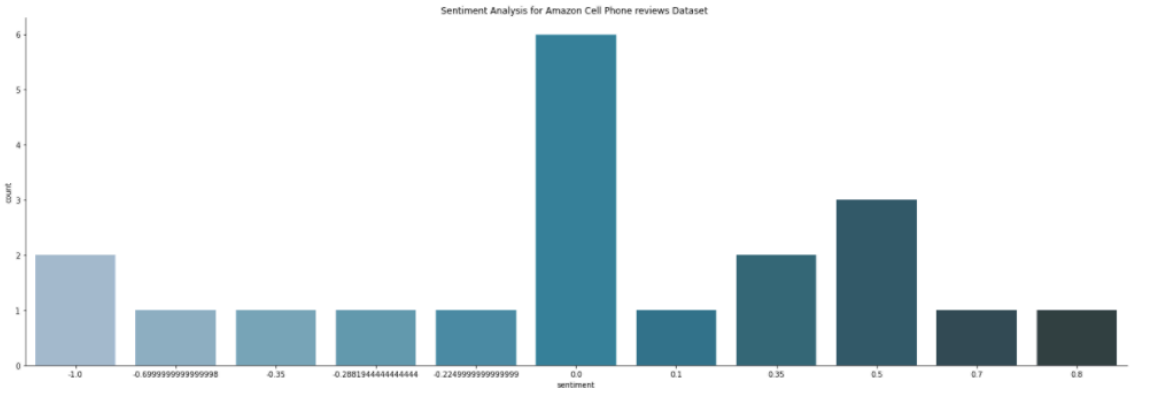
So, why are we using Spiderman movie to explain how human brains do sentiment analysis? The answer is simple: computer sentiment analysis works (almost) the same way.

Applied Sentiment Analysis into Project (Amazon Cell Phone Dataset)

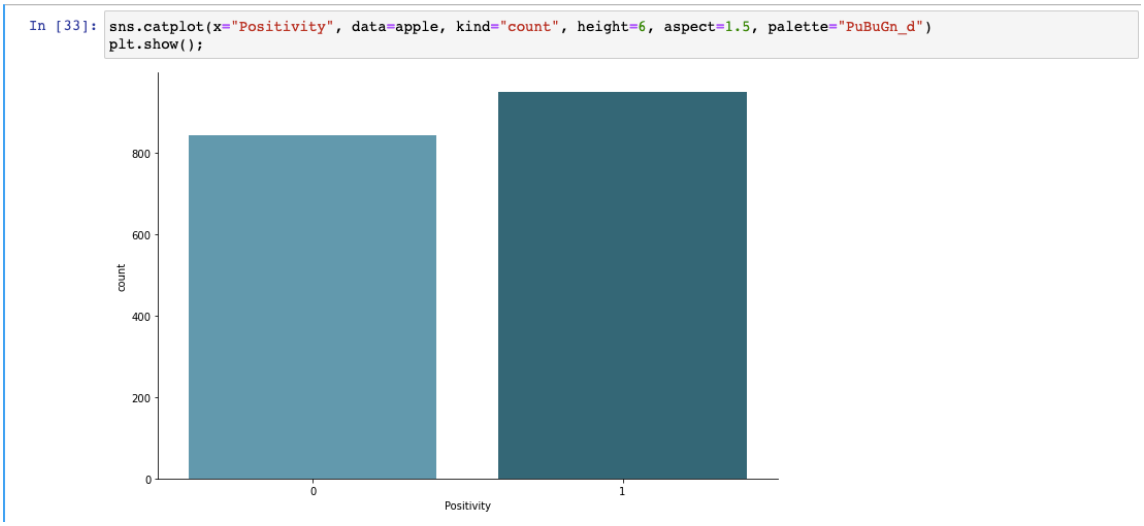
First, I have load Amazon Cell Phones Reviews dataset from Kaggle. Then I follow the steps to find the results. I Convert the sentence to Lower case, then I remove the stop words, Punctuation, Common word, and Rare words. I did the spelling correction. Then, I move to the next step with tokenization, stemming, and lemmatization.

For Sentiment Analysis, I came up with this results for people feedback about some types of cell phones and this is show that most of the words are natural and have 0 scores, and the other words vary between positive and negative score.

```
In [29]: sns.catplot(x="sentiment", data=df_data, kind="count", height=7, aspect=3, palette="PuBuGn_d")
         plt.title("Sentiment Analysis for Amazon Cell Phone reviews Dataset ")
         plt.show();
```



Next step, I choose specific brands to build model, I choose Apple to do the machine learning process, which is very popular company nowadays. I did the sentiment analysis on the apple reviews and come up with this results where the positivity vary between 0 and 1 but people like apple more than the people dose not although they are very close to each other.



Moreover, I build Model using RandomForestClassifier and evaluate the model, and RandomForestClassifier gave accuracy of 88%.

Applications of Sentiment Analysis

When it comes to brand reputation management, sentiment analysis can be used for brand reputation management to analyze web and social media opinions about products and services.These days, consumers use social media to share both their positive and negative experiences with brands which, if collected and analyzed, are a rich source of business information. Sentiment analysis tools can identify positive mentions showing strengths, as well as negative mentions, bad reviews and problems users write about online.

Conclusion

Previously there were opinion polls and questionnaires to find out people's opinions on a particular product, and today Big Data gives us a lot of opinions on the Data Science dish. People transfer to the virtual world through social networks by collecting, analyzing and presenting their data and this is known as Sentiment Analysis. By Symeon Symeonidis, Democritus University of ThraceIn “the last years, Sentiment Analysis has become a hot-trend topic of scientific and market research in the field of Natural Language Processing (NLP) and Machine Learning.”