1. Title
2. What is it?

* It is the use of natural language processing (NLP) and text analysis to identify and quantify subjective information pertaining to the text captured responses such as attitude, beliefs, and opinions of the sample group
* Sentiment analysis has been around for about half a century.
* Has rapidly gained popularity in recent years due to breakthrough and machine learning technology and the ever-growing adoption of social media

1. Why is this useful?

* Keeps the firms ahead of their competitors and allows them to shape their business strategies in a way to keep their customers happy and attract new customers.
* Evaluating marketing strategies through feedback from reviews such as customer reviews, Social Media post or other marketing study practices (i.e. focus groups, test markets, customer services).
* This is very useful for determining sentiment about public policy based on social media posts & for determining the best market strategies

1. Looking at the reviews we can see one is five stars while the other is one. After briefly looking at each one we can see that by the comments left by the customer with a negative review that they were not happy with certain features of the iPhone 7, specifically with the battery and the heat generated off the phone. If we look at the positive review the customer was happy with the software features and speed of the phone.
2. RESOURCES NEEDED (IMPLEMNTATION)
3. METHODOLOGY (WORKFLOW)
4. **Find good data sets for sentiment analysis**
   1. **Amazon customer reviews, restaurant chains reviews.**
   2. **Reviews with a star rating such as 3.5 out of 5, is ideal because it can be used to train the model. For instance uses of certain word correspond to specific ratings.**
5. **Train a model with suitable data sets**
   1. **There are Several approaches text analysis;**
   2. **We will be implementing the**
      1. **‘Bag-of-Words-Model’**
      2. **‘Naïve Bayes Classifier’**
      3. Text, letter

         Description automatically generated?\
6. **Analyze the two methods and;**
   1. **Decipher what makes them different from one another?**
   2. **Which is a more efficient model?**
   3. **Which learns faster? Slower?**
   4. **Which is most accurate? Why?**
7. SENTIMENT ANAYLISIS PROCESS
8. Text Input is the collecting and loading of text data for use in machine learning models.
9. Tokenization is a way of separating a piece of text into smaller units called tokens. Here, tokens can be either words, characters, or subwords.
10. Stop words are a set of commonly used words in a language. Examples of stop words in English are “a”, “the”, “is”, “are” and etc. Stop words are commonly used in Text Mining and Natural Language Processing (NLP) to eliminate words that are so commonly used that they carry very little useful information.
11. Negation handling is an automatic way of determining the scope of negation and inverting the polarities of opinionated words that are actually affected by a negation. The portion of the sentence that negation affects is called the vicinity or scope of negation.
12. Stemming is the process of reducing a word to its word stem by removing suffixes, or to the roots of words known as a lemma. Stemming uses an algorithm that simply cuts off the ends of words attempting to find the root word. Lemmatization uses extensive libraries to determine the morphology of words and can therefore be used contextually in models.
13. Text classification is the process of categorizing the text into a group of words. By using NLP, text classification can automatically analyze text and then assign a set of predefined tags or categories based on its context. We will be comparing at least two text classifiers. The bag-of-words model and a Naive Bayes classifier.
14. Sentiment classification is the automated process of identifying opinions in text and labeling them as positive, negative, or neutral, based on the emotions authors express within them.