**Sentiment Analysis on Textual Reviews**

**Problem Statement**

Every business likes to understand the customers or clients by knowing their opinions about the product. They collect some valuable feedback or reviews about the product from customers to get to know their satisfaction level with the product. The business user has to manually read the reviews which is written in the text form, which is a tedious process. Instead, by performing sentiment analysis on reviews, can figure out if the review is positive or negative. This way the business user with out reading each and every review will get to know the people opinion on products. Finally, business owner can consider several factors like improving the production quality, weather to increase the production or not to increase the production. In this project, I will be performing sentiment analysis on the reviews of Amazon Alexa products so that amazon can know the people have positive or negative opinion on their Alexa products.

**Example of problem**

* In an e-commerce site like amazon there will be various subsections like kids, furniture, Home, sports, Musical instruments. Sellers sell their products on e-commerce sites, if they want to know what people are thinking about the product, they will manually read the reviews and know what people are thinking about the product. Instead we can generate a sentiment analysis on reviews then sellers can easily understand the people’s opinion on products. Like wise, if amazon wants to know how people opinion about the Alexa devices then sentiment analysis on reviews are very user for business.

Below is the review about Amazon Alexa product

“*Amazon you really HIT the perfect sound coming from this little Speaker. You blew my mid with the perfect sound coming out of it. The Bass is very very good as well as the Mid range and Tweeters. I fell in Love with right away! Connectivity with ALL my Bluetooth Devices is instant. I am very pleased with this Echo Studio!!! I am listening to Amazon Music +, Tunein Stations, and all my own music from my ITunes*.” **Positive review**

**“***I tried to love it, but to be honest, it's pretty mediocre. Sound is pretty bad (expected) but better than previous version. Now the real problem is that it pretty much never does what I ask it to do apart from some really basic stuff. Even then, I need to use a "robotic"voice for the echo to understand. Also, the app has a really bad design and is not user-friendly.  
Now the only cool thing about it is its design. Otherwise, it's really not that good*.” **Negative review**

For a business user it is difficult to read such long reviews and understand the opinion of the product weather it is positive or negative. If sentiment analysis is performed on these text reviews and a classification model is applied to predict if the review is positive or negative. Using this process without reading all reviews business owner can decide if people have positive opinion on product or negative opinion. If some deep analysis is performed on the negative reviews then the business owner can concentrate more on what a user is expecting from the product in the next releases.

**Approach to Problem:**

Steps that are to be performed on the problem statement, to approach a solution for the problem are described below.

* Data collection
* Text preprocessing
* Transformation
* Feature selection
* Apply classification model
* Evaluation

Discussion of relevant literature

Sentiment analysis is performed at four different levels like word level, phrase level, sentence level, Document level. Wiebe [1] proposed subjectivity and meaning are both important properties using which word senses can be obtained. Theresa [2] proposed phrase level sentiment analysis to find if the expression is polar or neutral. Yi [3] proposed sentence level polarity classification, his model determines weather a sentence is positive or negative. In the paper Moutami [4] classified the textual reviews of movies dataset into positive, negative and neutral. The author applied tokenization, removal of stop words. The author also applied Term frequency and Inverse document frequency on the text data.

**Outline of Implementation:**

I will be working on the Amazon Alexa product reviews. On this dataset sentiment analysis is performed on the Alexa product reviews and classify if the review is positive or negative. So that huge organization like amazon can know their product opinions like weather Alexa products have positive or negative opinion on people.

Initially, Data is collected and loaded to the Notebook. Then data analysis is performed on the data, this step is performed to understand more about the data. Text processing techniques like tokenization, removal of stop words are applied on the dataset. In the transformation count vectorization, term frequency and inverse document frequency techniques are applied. Some text classification algorithms like Naïve Bayes, Support vector machines and Random forest classifier are applied. I will be testing the dataset based on various models and also apply the various techniques involved in the model.

**Criteria for Evaluating the model**

Some of the criteria used for evaluating the model are

*Accuracy*: when the data set is validated against the test data set and predictions it shows the percentage of the sentences predicted correctly.

*Precision and Recall*: They are the extended versions of accuracy. The exactness of model is measured using precision and the completeness of the model is measured using Recall. And their values are displayed using the confusion matrix.

*F1 score*: It is the harmonic mean of precision and recall.