

# 1.INTRODUCTION

## 1.1 Problem Statement

Sentiment analysis is a classic NLP problem because of both the research challenges it poses and its viability in real-life industrial applications. However as the users need of getting more accurate and precise information increases, such higher level sentiment analysis becomes redundant. Consider the following review: "The camera of the phone was nice but battery life is horrible". Here, a reviewer provides information about certain important aspects of the phone like camera and battery. If we are able to extract such aspects & the sentiment associated with them, we can extract much more precise and relevant information from a review which might prove to be beneficial for future consumers.

The aim of our project is to extract these aspect terms from a particular review. After extracting them, we predict their polarities i.e to determine whether a positive sentiment is associated with it or a negative sentiment. The dataset which we choose for our project is real time reviews from e-commerce web sites. We will build our model on the Amazon review dataset and test it on the scraped reviews.

## 1.2 Objective

Customer satisfaction is the most important goal for any service based organizations. The keen interest to know what customers feel about the services or products is important for the organization to further improve their quality and gain more customer base. It is also useful for new consumers to decide which product to buy etc. This project is aimed to improve the efficiency of existing models and provide a web portal to give the aspect entity sentiments expressed by consumers on products.

## 2. Literature Survey

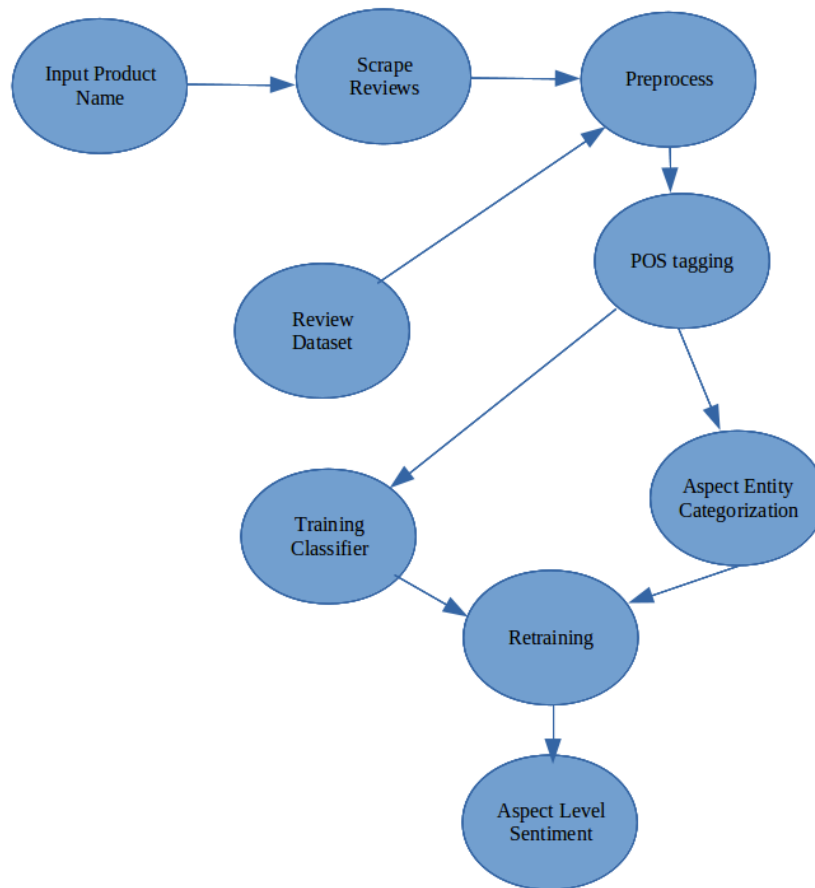
<b>Title</b>	<b>Year</b>	<b>Methodology</b>	<b>Summary</b>
Aspect Based Sentiment Analysis for Review Rating Prediction	2016	WordNet, Naive Bayes Classifier	This paper infers about various classifiers that can be implemented to get the sentiment polarities. Out of which Naive bayes give the best accuracy (70.2 %)
Supervised Learning Based Approach to Aspect Based Sentiment Analysis	2016	Support Vector Machines, Maximum Entropy	This paper infers about the technique maximum entropy which is used along with Support Vector Machines. It uses a non linear svm to classify the sentiments. It uses nltk package to pre processs the data.
Using Word Vectors with Typed Dependencies for Opinion Target Expression Extraction	2016	Word Vectors, Bag of Words	This paper infers about a technique of word vectors which converts each word into a vector of n dimensions. These are converted into a bag which contains these vectors. Then a classifier is trained to detect the sentiments.

### 3. METHODOLOGY

#### 3.1 System Design

This project focuses on developing a web based portal which can predict aspect level sentiments of a product give its name as input. The series of steps starts from scraping the reviews of the product from various e-commerce web sites and storing them. Then performing pre processing operations on the scraped data. The next step is to utilize nltk package and perform NLP operations such as stop word removal, spell checking and parts of speech tagging.

The important of the process starts with training a model that categorize aspect entity pairs of given reviews. The aim of the project is to improve the accuracy by using advanced techniques such as Long Short Term Memory neural networks for training the model.



*fig 1: Design*

## 3.2 Implementation

### 3.2.1 Input data

To train the classifier the data will be taken from a fixed data set which comprises of about 2 lakh reviews on electronic products. Testing data will be scraped from e-commerce web sites. While running the web app input data will be scraped from various websites.

### 3.2.2 Preprocessing

We use NLP techniques such as stop word removal, stemming, lemmatization and Parts of Speech tagging. This text data will be handled using data frames. These techniques will consume small amounts of execution time so they can be leveraged to our advantage to optimize the overall processing.

### 3.2.3 Aspect Categorization

There are various techniques to categorize aspects which involves pattern matching in pos tagged sentences or using word vectors to do the same. This will generate limited number of pairs which can be bigrams, trigrams or n-grams. The further step will be sentiment prediction using a classifier algorithm.

### 3.2.4 Sentiment Polarity

In this step we train a classifier on the input data and retrain it every time a new product name is given as an input. For classification we can use SVMs, naive bayes or advanced classifiers such as long short term memory neural networks.

## 3.3 Software required

This project is implemented open source and free softwares such as python, mysql, nltk etc. Language used is python and on top of it other packages are used. Its completely interpreted language and support object oriented features. The greatest advantage is that we can leverage the vast packages it has to offer.

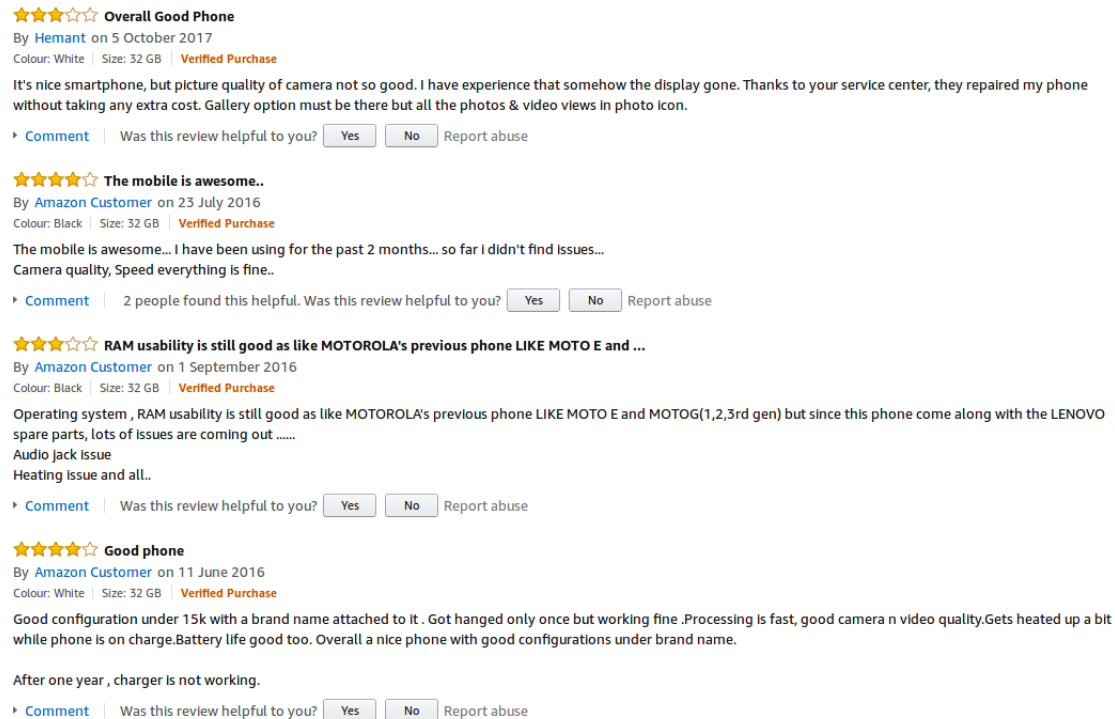
Data scraping is a key task to this project and it is implemented using Scrapy package in python. It is open source and provides great flexibility to use for multiple platforms. BeautifulSoup, parseHub packages can also be used. Flask is used to integrate different modules together as a web application.

NLTK is a package developed by stanford university and it stands for natural language toolkit. It provides tools such as stop word removal, lemmatizers, stemmers, taggers and many others. Along with NLTK we also use Gingerit -- a package which is used to correct the spellings of sentences based on the context in which the words appear.

Sklearn (also stands for SciKit Learn), TensorFlow packages are used to implement the classification algorithms.

## 4. DATASET

Mobiles and accessories data set taken to train the classifier. It consists of about two lakhs (200000) appx of reviews and their information. For testing and running of the web app real time reviews are scraped and used.



The image shows a screenshot of four Amazon product reviews for a smartphone. Each review includes a star rating, the reviewer's name, the date, product details (colour and size), a 'Verified Purchase' badge, the review text, and interaction buttons (Comment, Yes/No, Report abuse).

**Review 1:**  
★★★★☆ **Overall Good Phone**  
By [Hemant](#) on 5 October 2017  
Colour: White | Size: 32 GB | [Verified Purchase](#)  
It's nice smartphone, but picture quality of camera not so good. I have experience that somehow the display gone. Thanks to your service center, they repaired my phone without taking any extra cost. Gallery option must be there but all the photos & video views in photo icon.  
▶ [Comment](#) | Was this review helpful to you?   [Report abuse](#)

**Review 2:**  
★★★★☆ **The mobile is awesome..**  
By [Amazon Customer](#) on 23 July 2016  
Colour: Black | Size: 32 GB | [Verified Purchase](#)  
The mobile is awesome... I have been using for the past 2 months... so far I didn't find issues...  
Camera quality, Speed everything is fine..  
▶ [Comment](#) | 2 people found this helpful. Was this review helpful to you?   [Report abuse](#)

**Review 3:**  
★★★★☆ **RAM usability is still good as like MOTOROLA's previous phone LIKE MOTO E and ...**  
By [Amazon Customer](#) on 1 September 2016  
Colour: Black | Size: 32 GB | [Verified Purchase](#)  
Operating system , RAM usability is still good as like MOTOROLA's previous phone LIKE MOTO E and MOTOG(1,2,3rd gen) but since this phone come along with the LENOVO spare parts, lots of issues are coming out .....  
Audio jack issue  
Heating issue and all..  
▶ [Comment](#) | Was this review helpful to you?   [Report abuse](#)

**Review 4:**  
★★★★☆ **Good phone**  
By [Amazon Customer](#) on 11 June 2016  
Colour: White | Size: 32 GB | [Verified Purchase](#)  
Good configuration under 15k with a brand name attached to it . Got hanged only once but working fine .Processing is fast, good camera n video quality.Gets heated up a bit while phone is on charge.Battery life good too. Overall a nice phone with good configurations under brand name.  
After one year , charger is not working.  
▶ [Comment](#) | Was this review helpful to you?   [Report abuse](#)

*fig 2: Reviews from amazon*

## REFERENCES

- [1] Susanti Gojali, Masayu Leyia Khodra, Aspect Based Sentiment Analysis for Review Rating Prediction, IEEE publication 2016.
- [2] Lamei Xu , Jin Lin, Lina Wang, Chunyong Yin, Jin Wang, Deep Convolutional Neural Network based Approach for Aspect-based Sentiment Analysis, 2016.
- [3] Pontiki, D. Galanis, H. Papageogiou, S. Manandhar, and I. Androutsopoulos. 2015. Semeval-2015 task 12:Aspect based sentiment analysis. In Proceedings of SemEval 2015, pages 27–35, Denver, Colorado.
- [4] Nipuna Upeka Pannala, Chamira Priyamanthi Nawarathna, Supervised Learning Based Approach to Aspect Based Sentiment Analysis, IEEE international conference, 2016.
- [5] .Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 2009:1–12.
- [6] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, 2014.
- [7] Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. Recursive neural conditional random fields for aspect-based sentiment analysis. CoRR, abs/1603.06679, 2016.
- [8] S. Poria, E. Cambria, and A. Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. Elsevier, pages 42–49, 2016.