

# CHAPTER 1

## INTRODUCTION

People's opinion has become one of the extremely important sources for various services in ever-growing popular social networks. In particular, online opinions have turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities, and manage their reputations. In general, recommender systems are defined as the supporting systems which help users to find information, products, or services (such as books, movies, music, digital products, websites, and TV programs) by aggregating and analyzing suggestions from other users, which means reviews from various authorities, and user attributes. After viewing such reviews they take their decisions. So, such reviews must be correct and proper.

Generally, the reviews are generated in graphical format that is in star ratings. Users just have to see the ratings which are generated by analyzing the ratings given by other users to that product and have to take his/her decisions. Such ratings are easily understandable by any user. But they don't give clear idea of how the product is. They are helpful only in the scenario where if any product is excellent or very poor. The scenario where product is average, star ratings prove bit confuse for any user. They don't get clear views of what the other users think of that product.

If the reviews are in simple English statement it would be easy for any user to understand the feelings of the other users too, about the product. Also star ratings will be there for his/her help. So, the review about any product will give clear idea to any user so that he can easily take his/her decisions in such confusing scenario too.

Our system is a movie review system which will generate reviews related to the movies which are released. Unlike other systems, we are going to generate a common review by analyzing only the comments of the people (no heavy feedback). This will reduce the overhead of any user who is commenting on any movie and will make the system more user-friendly. So, the system will generate better review which will be a simple English statement. Also our system will provide star ratings too.

## CHAPTER 2

### RELATED WORK

#### 1. Comparing sentiment analysis using traditional and deep learning models : -

##### A. Deep Learning Model: -

Two methods are used to implement the learning models one is using ConvNets and the other is a simple long short-term memory (LSTM) recurrent neural network model. Also, ConvNets and LSTM were combined to compare the models. Deep learning models are “Data Hungry” works well on large data sets.

##### B. Traditional Model: -

Bag-of-words, bag-of-n-grams and their TF-IDF variants with a combination of couple linear classifiers to classify sentiments of reviews are used. Pre-trained word embedding models word2vec and GloVe are used to represent review document as a vector. Bag-of-ngrams model where  $n=5$  + TFIDF with Logistic Regression is the best model.

#### 2. Parts of Speech tagging to unravel positive and negative pattern in an document:-

Here an unprocessed text document is taken and is pre-processed by removing special symbol, splitting sentence , Tokenization, stop-word removal, and stemming. After feature extraction for Parts of Speech tagging we use Corpora Brill’s Technique. We obtain a frequency of 83%. But after applying statistical model we can achieve accuracy of over 90% in sentiment analysis.

#### 3. OPINE: -

Opine model is a useful algorithm in “Review Mining”. This algorithm is made on KnowitAllSystem. The MINIPAR algorithm is used to parse the reviews; from these parsed reviews we find the explicit product features using feature assessor and its use of Web Point-wise Mutual Information (PMI). Then from it we find the opinions and cluster the opinions in positive and negative opinions. From the clustered opinion we find the implicit features and then rank the opinions according to their polarity. Higher the PMI perfect the opinion. Naïve Bayes classifier is used for clustering.

#### 4. A Review Classification of Recommender Systems Research

This paper we referred is a research paper. In this, author have identified 164 articles on recommender systems, which are published from 2001 to 2009 to understand the trend of recommender systems research and to provide practitioners and researchers with insight and future direction on recommender systems.

## **CHAPTER 3**

### **Hardware and Software Requirements**

#### **Hardware Requirements**

- Intel 3<sup>rd</sup> gen i3 or higher
- 8GB RAM
- GB hard free drive space

#### **Software Requirements**

- Jupiter notebook
- Web Browser: Microsoft Internet Explorer, Mozilla, Google Chrome or later
- Operating System: Windows XP / Windows7/ Windows Vista/Windows 10

## CHAPTER 4

### IMPLEMENTATION AND RESULT

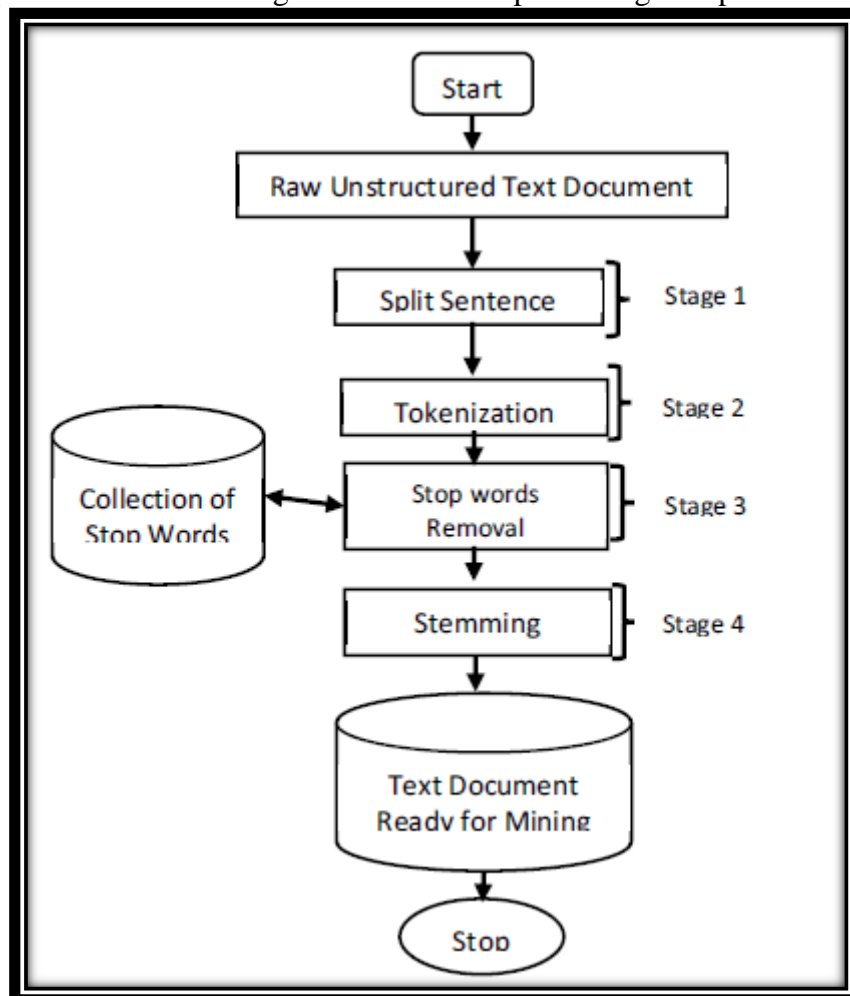
#### Implementation

##### ➤ Data gathering:

The required review input data is collected or manually scrapped from the e-commerce websites like Amazon, Apple official website etc. The data is put together in a text file for further processing.

##### ➤ Data pre-processing

Data pre-processing is all about cleaning the unwanted data like special characters, symbols, stop words etc. from the gathered data. Pre-processing is explained in figure below



##### ➤ Data cleaning:

**Stage 1: Split Sentences** - This algorithm which splits the whole text into sentences

**Stage 2: Tokenization** - This step involves splitting up of the sentence into tokens and these Token are converted into lower case. The process of splitting is called Tokenization.

**Stage 3: Stop-word Removal** - In this stage, stop words are removed. The list of stop words is stored in a separate file.

## Result

After three iterations we received an accuracy of 94.95% in first run

```
Train on 60000 samples, validate on 15000 samples
Epoch 1/3
60000/60000 [=====] - 264s 4ms/step - loss: 0.3283
- accuracy: 0.8544 - val_loss: 0.2641 - val_accuracy: 0.9058
Epoch 2/3
60000/60000 [=====] - 281s 5ms/step - loss: 0.2124
- accuracy: 0.9185 - val_loss: 0.1771 - val_accuracy: 0.9389
Epoch 3/3
60000/60000 [=====] - 293s 5ms/step - loss: 0.1613
- accuracy: 0.9406 - val_loss: 0.1516 - val_accuracy: 0.9495

F1-score: 0.9603399433427762
Confusion matrix:
array([[12155,    635],
       [   345, 11865]], dtype=int64)
```

After three iterations we received an accuracy of 93.97 % in second run

```
Train on 60000 samples, validate on 15000 samples
Epoch 1/3
60000/60000 [=====] - 220s 4ms/step - loss: 0.3560
- accuracy: 0.8365 - val_loss: 0.3116 - val_accuracy: 0.8727
Epoch 2/3
60000/60000 [=====] - 231s 4ms/step - loss: 0.2273
- accuracy: 0.9105 - val_loss: 0.1892 - val_accuracy: 0.9345
Epoch 3/3
60000/60000 [=====] - 267s 4ms/step - loss: 0.1733
- accuracy: 0.9352 - val_loss: 0.1753 - val_accuracy: 0.9397

F1-score: 0.9545381006309658
Confusion matrix:
array([[12076,    700],
       [   424, 11800]], dtype=int64)
```

## **CHAPTER 5**

### **CONCLUSION**

Here, we design and implement a movie-rating and review-summarization system. Sentiment classification is applied to the movie reviews, and rating information is based on sentiment-classification results. In feature-based summarization, product-feature identification plays an essential role, and we propose a novel approach based on LSA to identify related product features. Sentiment analysis is done on the comments given by the users. The design proposed in this paper could fully utilize the Internet content to provide a new product-review summarization and rating service. The design can also be extended to other product-review domains easily.

## **CHAPTER 6**

### **FUTURE SCOPE**

- Can be further extended for classification of reviews related to the products in Online Shops.
- Also can be used for classification reviews for the online videos.
- Also can be used for classification reviews related to the colleges during admission process.
- Also can be used for classification reviews of the candidates in the election.

## REFERENCES

- [1] Roshan D, Dr. T Hanumantha Reddy, " Parts of Speech tagging mechanism to unravel positive and negative patterns in an unstructured text document", International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), at K.L.S. Gogte Institute of Technology,Belagavi, Karnataka,2018,pp.1-6.
- [2] Tamara Katic , Nemanja Milicevic, "Comparing Sentiment Analysis and Document Representation Methods of Amazon Reviews", 16th International Symposium on Intelligent Systems and Informatics, at Subotica, Serbia, 2018,pp.1-6
- [3] Buddhika H. Kasthuriarachchy, Kasum De Zoysa and H.L. Premaratne ,2014 "Enhanced bag-of words model for phrase-level sentiment analysis ",International Conference on Advances in ICT for Emerging Regions (ICTer), at Reid Avenue, Colombo 07, Sri Lanka, 2014, pp.210-214
- [4] Anne Kao and Stephen R. Poteet (Eds), " Extracting Product Features & Opinions from Reviews" , in ACM SIGKDD(Special Interest Group on Knowledge Discovery and Data Mining) ,Seattle, WA, USA, 2004, pp.9-28.