REVIEW SENTIMENT ANALYSIS OF IMDb REVIEWS

A PROJECT REPORT

Submitted by

Uday Agarwal (19BAI10147)
Aryan Tandon (19BAI10148)
Sahil Gauba (19BAI10099)
Parv Bhargava (19BAI10116)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Specialization in

Artificial intelligence and machine learning

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

KOTHRIKALAN, SEHORE
MADHYA PRADESH - 466114

APR 2021

VIT BHOPAL UNIVERSITY, KOTHRIKALAN, SEHORE MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"REVIEW SENTIMENT ANALYSIS OF IMDb REVIEWS"** is the bonafide work of **" Uday Agarwal (19BAI10147), Aryan Tandon (19BAI10148), Sahil Gauba (19BAI10099)** and **Parv Bhargava(19BAI10116)**" who carried out the project work under my supervision.

Certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR
Dr. S Sountharrajan
School of AI &ML division
VIT BHOPAL UNIVERSITY

PROJECT GUIDE
Dr. L. Shakkeera
School of AI & ML division
VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on _____

ACKNOWLEDGEMENT

First and foremost, we would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

We wish to express our heartfelt gratitude to Dr S Sountharrajan, Head of the Department, School of Computer Science and Engineering (AI & ML) for much of his valuable support encouragement in carrying out this work.

We would like to thank our internal guide Dr. L. Shakkeera for continually guiding and actively participating in our project, giving valuable suggestions to complete the project work.

We would like to thank all the technical and teaching staff of the School of Computer Science and Engineering, who extended directly or indirectly all support.

Last, but not the least, we are deeply indebted to our parents who have been the greatest support while we worked day and night for the project to make it a success.

LIST OF ABBREVIATIONS

- POS – Parts of Speech
- NLTK – Natural Language Toolkit
- IDE – Integrated Development Environment
- NLP – Natural Language Processing
- et al – et alia
- etc – et cetera
- www – World Wide Web
- ICIRD - International Conference on Innovative Research and Development
- *IJCSIT - International Journal of Computer Science and Information Technology*
- DOI – Digital Object Identifier
- RAM – Random Access Memory
- API - Application Programming Interface
- SQL - Structured Query Language
- IMDb – Internet Movie Database
- SVM – Support Vector Machine

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Gathering public opinion by analyzing big social data has attracted wide attention due to its interactive and real time nature. For this, recent studies have relied on both social media and sentiment analysis in order to accompany big events by tracking people's behavior. Movie reviews are becoming more important with the evolution of the movie industry. Reviewers are posting reviews directly on movie pages in real time. With the vast amount of movie reviews, this creates an opportunity to see how the industry reacts to a specific product. In this project we propose an adaptable sentiment analysis approach that analyzes IMDb movies' review dataset. By using NLP, we will make the computer truly understand more than just the objective definition of the words. This analysis will help us segregate the data that has good as well as bad movie reviews.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 INTRODUCTION

Can you imagine manually sorting through thousands of tweets, customer support conversations, or surveys, or product and movie reviews?

There's just too much business data to process manually. Sentiment analysis helps businesses process huge amounts of data in an efficient and cost-effective way. In the previous review we've already discussed how sentiment analysis depends on polarity and even the feelings, emotions and intentions of movie reviewers. Sentiment analysis, also known as opinion mining, is essential as analysing customer feedback such as opinions in survey responses, social media conversations and product reviews allows business and brands to listen attentively to their customers and tailor products and services to meet their needs.

Feature based sentiment analysis include feature extraction, sentiment prediction, sentiment classification and optional summarization modules. Feature extraction identifies those product aspects which are being commented by customers, sentiment prediction identifies the text containing sentiment or opinion by deciding sentiment polarity as positive, negative or neutral and finally summarization module aggregates the results obtained from previous two steps.

1.2 OVERALL SYSTEM ARCHITECTURE

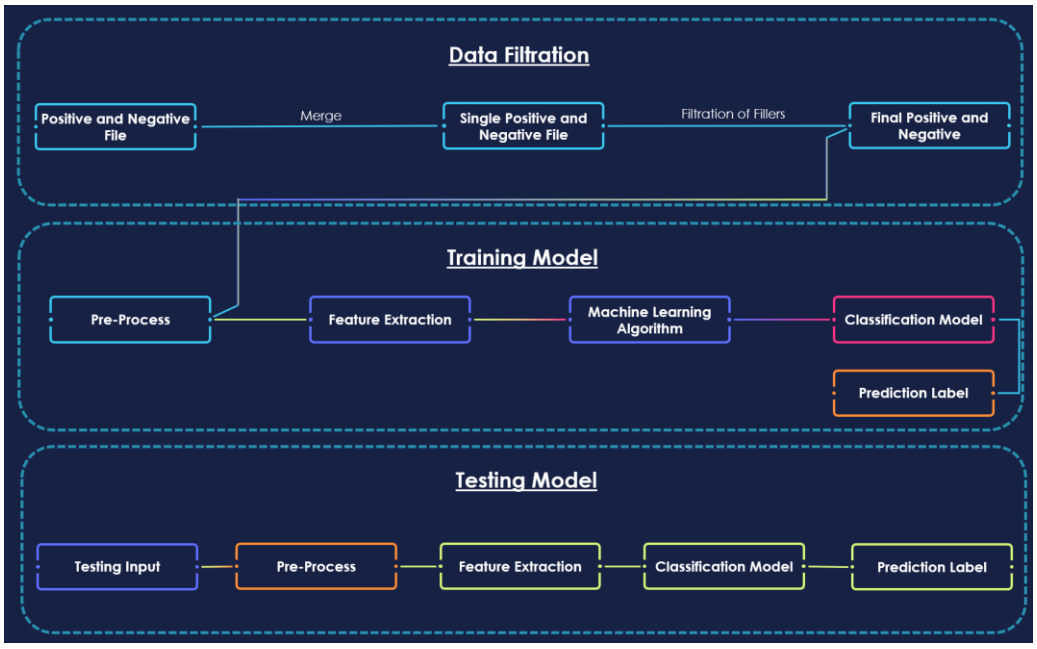

Fig. 1.2

Data Filtration: Data Filtration includes importing all positive and negative datasets from file and combining them into a single file. The data sets may contain lots of unwanted symbols, and number. These factors need to be corrected or solved to increase the efficiency. Therefore, in this process the unwanted symbols and number are removed.

Training Model: This includes Fetching the datasets from the file and extracting all the corresponding words (feature words) like adjective, adverb and verb. Then datasets are labelled a respectively as “pos” for positive and “neg” for negative. Eventually, frequency distribution is performed over collected words and 5000 words for training are selected. Again, the shuffling of data is performed using random seed for better training.

Testing Model: Here user can test and analysis the respective model by performing preprocessing over the input data. The preprocessing contains the removal of the symbol and number. Mapping to user input using saved featured (based on training dataset). Then feed to saved model for prediction.

1.3 EXISTING WORK WITH LIMITATIONS

Method	Proposal	Classification	Text Level	Prediction Accuracy	Pros	Cons
OPINE	2005	Unsupervised rule-based approach	Word	87%	Domain independent	Difficulty in availing OPINE system, thus rare to get applied in real life
Sentiment Analysis: Adjectives and Adverbs are better than Adjectives alone	2006	Linguistic Approach	Document	Pearson correlation of 0.47	Adjectives are given more priority (adjectives express human sentiments better than adverbs alone)	None
Opinion Digger	2010	Unsupervised Machine Learning Method	Sentence	51%	Rates product at aspect level	Requires rating guidelines to rate. Works only on known data
Sentiment Classification using Lexical Contextual Sentence Structure	2011	Rule based Approach	Sentence	86%	Said to be domain independent	Depends solely on WordNet
Independent Latent Dirichlet Allocation	2011	Probabilistic Graphical Model	Document	73%	Faster in comparing and correlating sentiment and rating	Correlation between identified clusters and feature or ratings are not explicit always
A joint Model of Feature Mining and Sentiment Analysis for Product Review Rating	2011	Machine Learning	Document	71% (in 3 categories); 46.9% (in 5 categories)	Automatic calculation of feature vector	Use of WordNet

Table 1.3

Here is the table showing the existing work on Sentiment Analysis with their respective limitations and their year of proposal.

1.4 REAL TIME USAGE

1.4.1 ANALYSING AMAZON PRODUCT REVIEWS

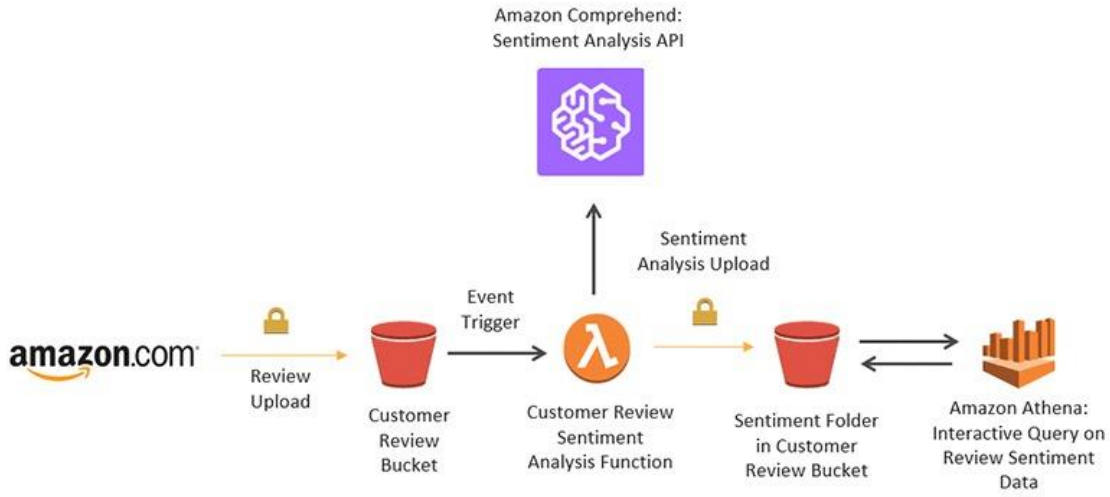

Fig. 1.4.1

Amazon is the biggest e-commerce store on the planet. This means it also has one of the largest product selections available. Many times, companies want to understand the public opinion on their product and figure out what's responsible for the same. or that purpose, they perform sentiment analysis on their product reviews. It helps them in recognizing the primary issues with their products (if there are any). Some products have thousands of reviews on Amazon while some others only have a few hundred. So, analyzing the data from those customer reviews to make the data more dynamic is an essential field nowadays. In this age of increasing machine learning based algorithms reading thousands of reviews to understand a product is rather time consuming where we can polarize a review on particular category to understand its popularity among the buyers all over the world. The objective is to categorize the positive and negative feedbacks of the customers over different products and build a supervised learning model to polarize large number of reviews. A study on amazon last year revealed over 88% of online shoppers trust reviews as much as personal recommendations. Any online item with large number of positive reviews provides a powerful comment of the legitimacy of the item. Conversely, books, or any other online item, without reviews puts potential prospects in a state of distrust.

1.4.2 ANALYSING IMDb REVIEWS

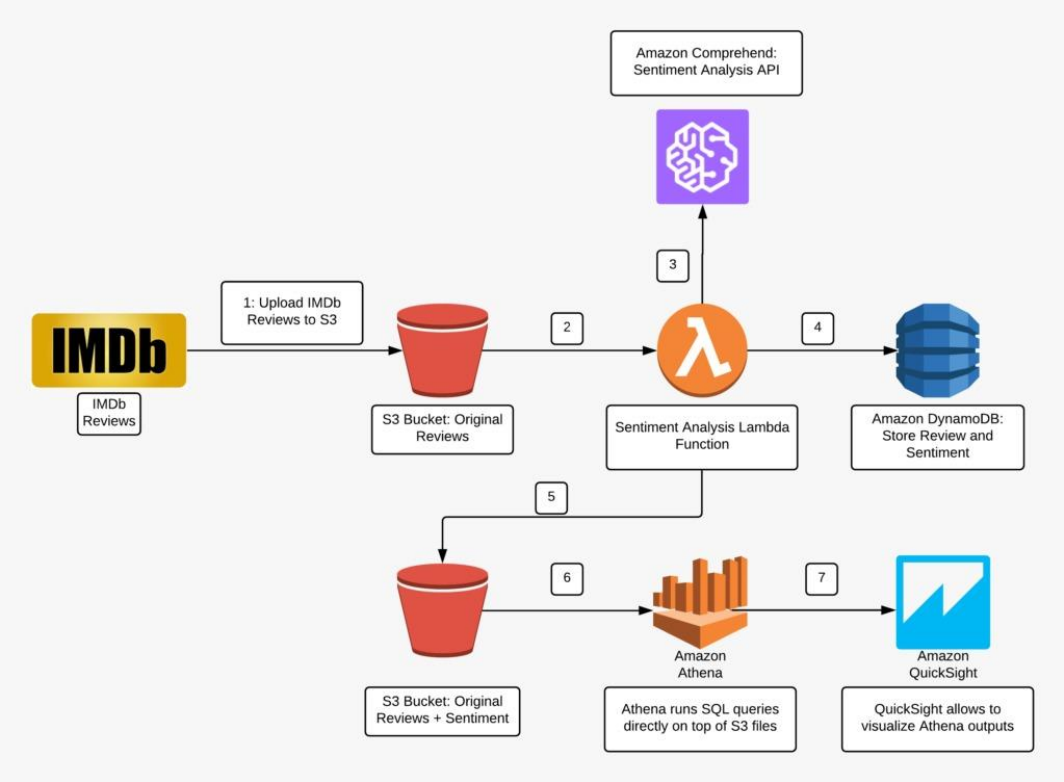

Fig. 1.4.2

The objective is to categorize the positive and negative feedbacks of the customers over different products and build a supervised learning model to polarize large number of reviews. A study on amazon last year revealed over 88% of online shoppers trust reviews as much as personal recommendations. Any online item with large number of positive reviews provides a powerful comment of the legitimacy of the item. Conversely, books, or any other online item, without reviews puts potential prospects in a state of distrust.

IMDb is an entertainment review website where people leave their opinions on different movies and shows. Reviews of shows and movies help production companies in understanding why their title failed (or succeeded). The dataset is from IMDB, which a famous movie online database of information related to world movies and other entertainment sources. A collection of 50,000 reviews from IMDB, where it allows no more than 30 reviews per movie. The constructed dataset contains an even number of positive and negative reviews, so randomly guessing yields 50% accuracy. In the proposed model sentiment analysis is employed on IMDb movie reviews. The input for the proposed model is the set of reviews whose polarity needs to be determined.

1.5 NOVELTY OF THE PROJECT

The project presents the theoretical analysis of methods or proposal of Sentiment Analysis of the reviews of a particular movie. Both the advantages and disadvantages of the discussed methods are considered to add new features in the proposed approach. The new approach follows machine learning technique at document level with combination of adjectives, adverbs, and verbs. The project is completed using PyCharm, Jupyter Notebook, Matplotlib, TextBlob, NLTK and Scikit Learn. The project uses Natural Language Processing and Machine Learning to interpret and classify emotions of different reviewers in subjective data.

In this work, we proposed a domain dependent rule-based method for semantically classifying sentiment from online customer reviews and comments. This method works as it takes a review; checks individual sentences whether sentences are objective or subjective, and decides its semantic orientation by using lexical contextual information at the sentence level. The statistical data of sentiments and their frequency in a review is also displayed.

2. LITERATURE SURVEY

2.1 INTRODUCTION

The problem of polarity categorization that determines whether a review is positive, negative or neutral, has been addressed. A general process for sentiment polarity categorization is discussed with detailed process descriptions. They have also mentioned the flaws related to collection of online data by developers that hinders the process of sentiment analysis have also been mentioned.

2.2 LITERATURE REVIEW

Fang et al [1] have addressed the problem of polarity categorization that determines whether a review is positive, negative or neutral. First, each movie review receives inspections before it can be posted. Second, each review must have a rating on it that can be used as the ground truth. The rating is based on a star-scaled system, where the highest rating has 10 stars and the lowest rating has only 1 star.
It is observed that both the SVM model and the Naive Bayesian model are identical in terms of their performances. Both models are generally superior than the Random Forest model on all vector sets. Even though there are reports of spams on imdb.com, it is still a relatively spam-free website in terms of reviews because of the enforcement of its review inspection process.

Shah et al [2] used both manual and active learning approach to label the datasets. In the active learning process different classifiers are used to provide accuracy until reaching satisfactory level. After getting satisfactory result the labelled datasets were taken and processed. From the processed dataset we extracted features that are then classified by different classifiers. Combination of two kinds of approaches to extract features were used. A supervised learning model is proposed to polarize a large amount of product review dataset which was unlabeled. The proposed model is a supervised learning method and uses a mix of 2 kinds of feature extractor approach for higher accuracy.

Bhatt et al [3] ensured fair results of sentiment thereby saving the users' time to read through long textual description in the reviews. This was done with the help of the following steps, namely: (i) Data Scraping; (ii) Data Cleaning; (iii) Classifying the reviews into service

review and product review; (iv) Extracting features and Parts of Speech (POS) tagging, and (v) Interpreting the Result.

Classification of Reviews along with sentiment analysis increased the accuracy of the system which in turn provided accurate reviews to the user. Data visualization is an important technology in the upcoming future. Facts are objective statements about entities and worldly events and opinions are subjective statements that reflect people's sentiments. Maximum amount of existing research on text and information processing is focused on mining and getting the factual information from the text or information.

Tripathy et al. [4] presented a text classification by using Naïve Bayes (NB) and support vector machine (SVM). The results showed that these two algorithms can classify the dataset with high accuracy compared to the other existing research.

Two different algorithms namely Naive Bayes (NB) and Support Vector Machine (SVM) are implemented. These two algorithms have also been implemented earlier by different researchers and results of all versions of implementation have been compared. It is observed that SVM classifier outperforms every other classifier in predicting the sentiment of a review.

Vijayaragavan et al. [5] discussed an optimal SVM based classification for the sentimental analysis of online product reviews. The paper firstly applied SVM and K-means to cluster the reviews into two groups. Then, the authors employed fuzzy based soft set theory to determine the possibility of customer to purchase the product.

Machine learning approaches generate results with higher accuracy but it is not being used widely because of complex implementation requirements. Hybrid approach is also not been employed by researchers as it integrates two approaches. Coming towards the languages, English is the most explored with the domain of Sentime

3. SYSTEM ANALYSIS

3.1 MODULE DESCRIPTION

3.1.1 DATA ACQUISITION

The collection of data is an important phase since a proper dataset needs to be defined for analyzing and classifying the text in the dataset. Data acquisition is the process of sampling signals that measure real world physical conditions and converting the resulting samples into digital numeric values that can be manipulated by a computer. There are also open-source software packages providing all the necessary tools to acquire data from different, typically specific, hardware equipment. Data acquisition applications are usually controlled by software programs developed using various general-purpose programming languages such as Assembly, BASIC, C, C++, C#.

Fig. 3.1.1

3.1.2 ACTIVE LEARNING

Active learning is a special case in semi-supervised learning algorithm. The main fact is that the performance will be better with less training if the learning algorithm is allowed to choose the data from which it learns. Active learning system tries to solve data labeling bottleneck by querying for unlabeled instance to be properly labeled by an expert. As manually labeling the dataset is quite an impossible task so that to reduce time complexity, we use a special kind of semi-supervised learning approach known as active learning. In the process of our active learning, we need to provide it some pre labeled datasets as training and testing and take unlabeled dataset. For using active learning, we need to provide some manually labeled reviews as training–testing sets. And it will run some classifiers and will tell us the output. There are situations in which unlabeled data is abundant but manual labeling is expensive. In such a scenario, learning algorithms can actively query the user/teacher for labels. This type of iterative supervised learning is called active learning.

Fig. 3.1.2

3.1.3 TOKENIZATION

Tokenization is the process of separating a sequence of strings into individuals such as words, keywords, phrases, symbols and other elements known as tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks are discarded. The tokens work as the input for different process like parsing and text mining.

In other words, it is the process of converting text into tokens before transforming it into vectors. It is also easier to filter out unnecessary tokens. For example, a document into paragraphs or sentences into words. In this case we are tokenizing the reviews into words. Tokenization can be done to either separate words or sentences. If the text is split into words using some separation technique it is called word tokenization and same separation done for sentences is called sentence tokenization. Here is an image depicting role of tokenization in daily life.

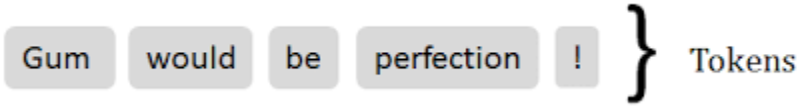

Fig. 3.1.3

3.1.4 REMOVING STOP WORDS

Stop words are those objects in a sentence which are not necessary in any sector in text mining. So, we generally ignore these words to enhance the accuracy of the analysis.

Stop words are the most commonly occurring words which are not relevant in the context of the data and do not contribute any deeper meaning to the phrase. In this case contain no sentiment. A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.

We would not want these words to take up space in our database, or taking up valuable processing time. For this, we can remove them easily, by storing a list of words that you consider to stop words.

NLTK (Natural Language Toolkit) in python has a list of stop words stored in 16 different languages. With the Python programming language, you have a myriad of options to use in order to remove stop words from strings. You can either use one of the several natural language processing libraries such as NLTK, Scikit Learn, TextBlob, P etc., or if you need full control on the stop words that you want to remove, you can write your own custom script.

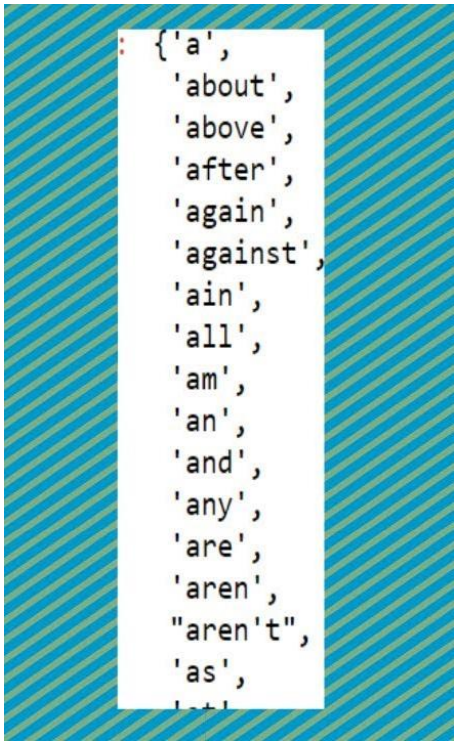

Fig. 3.1.4

3.1.5 PARTS OF SPEECH TAGGING

The process of assigning one of the parts of speech to the given word is called Parts of Speech tagging. It is generally referred to as POS tagging. Parts of speech generally contain nouns, verbs, adverbs, adjectives, pronouns, conjunction and their sub-categories. Parts of Speech tagger or POS tagger is a program that does this job.

In other words, it is the process of classifying words into their parts of speech and labeling them accordingly. It processes a sequence of words, and attaches a part of speech

tag to each word. When a user reviews a product, the best clue to his/her opinions is given by the adjectives used in the reviews. Thus, it can be said that POS Tagger classifies whether a word is an adjective or a noun.

POS tagging can be really useful, particularly if you have words or tokens that can have multiple POS tags. For instance, the word "google" can be used as both a noun and verb, depending upon the context. While processing natural language, it is important to identify this difference.

Part of Speech	Tag
Noun	n
Verb	v
Adjective	a
Adverb	r

Fig. 3.1.5

3.1.6 FEATURE EXTRACTION

Feature extraction involves reducing the number of resources required to describe a large set of data. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. It identifies aspects of those movies which are being reviewed by viewers, sentiment prediction identifies the text containing sentiment or opinion by deciding sentiment polarity as positive, negative or neutral. Finally, the results obtained from previous steps are aggregated and displayed, which is known as result interpretation.

It addresses the problem of attaining the most informative and compact set of features, to improve the performance of machine learning models. Let's go step-by-step. First, we are talking about 'informative'. This means that we are looking for features that can characterize the behavior of what we are trying to model. For instance, if we want to model the weather, features like temperature, humidity and wind are informative (they are related to the problem). By contrast, the result of a football game will not be an informative feature because it doesn't affect the weather.

Regarding 'compact', what we mean is that we want to exclude irrelevant features from our model. There are several reasons to exclude irrelevant features. In our case, I'd say that the most important is to reduce overfitting.

Fig. 3.1.6

3.2 PROPOSED WORK

The project presents the theoretical analysis of methods or proposal of Sentiment Analysis of the IMDb reviews of a particular movie. Both the advantages and disadvantages of the discussed methods are considered to add new features in the proposed approach. The new approach follows machine learning technique at document level with combination of adjectives, adverbs, and verbs. The project is completed using PyCharm, Jupyter Notebook, Matplotlib, TextBlob, NLTK and Scikit Learn. The project uses Natural Language Processing and Machine Learning to interpret and classify emotions of different reviewers in subjective data.

In this work, we proposed a domain dependent rule-based method for semantically classifying sentiment from online customer reviews and comments. This method works as it takes a review; checks individual sentences whether sentences are objective or subjective, and decides its semantic orientation by using lexical contextual information at the sentence level.

The statistical data of sentiments and their frequency in a review is also displayed.

3.3 MODULE WORKFLOW EXPLANATION

Fig. 3.3

- Import the Dataset: Initially we import the Movie Reviews' Dataset to our model.
- Active learning: In this step, we provide some pre-labelled dataset as training and testing and take unlabeled dataset.
- Data Preprocessing: The data extracted need to be cleaned so that we get proper text review on which analysis can be performed.
- Tokenization: Here we separate the sequence of strings into individuals such as words, keywords, phrases, symbols and other elements known as tokens.
- Stop Words removal: Here we remove all the unnecessary words from a sentence.
- Parts of Speech (POS) tagging: The process of assigning one of the parts of speech to the given word is called Parts of Speech tagging. POS tagger is very useful because of the following two reasons: (a) words like nouns and pronouns usually do not contain any sentiment. It is able to filter out such words with the help of a POS tagger, and; (b) a

POS tagger can also be used to distinguish words that can be used in different parts of speech. For instance, as a verb, “enhanced” may conduct different amount of sentiment as being of an adjective.

- Feature Extraction: In this step the model extracts all the important features which are useful in evaluating the result and ignores the rest as those features might affect the accuracy level of the model.
- Supervised Learning: Here the sentiments are classified.
- Result Interpretation: Final label - positive or negative workflow explanation.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1 IMPLEMENTATION

NLTK (Natural Language Toolkit)

In our project for performing sentiment analysis, we have used the NLTK package of Python which is a suite of libraries and programs for statistical NLP. It is used in the tokenize the text and has been used as a POS (Part of Speech) tagger. POS tagging is the process of marking up a word in a text as corresponding to a particular part of speech based on its definition and its context.

Scikit Learn

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

TextBlob

It is a module which is used for building programs for text analysis. One of the most important aspects of the TextBlob module is the POS tagging. Its goal is to assign linguistic information to substantial units.

SVM

Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

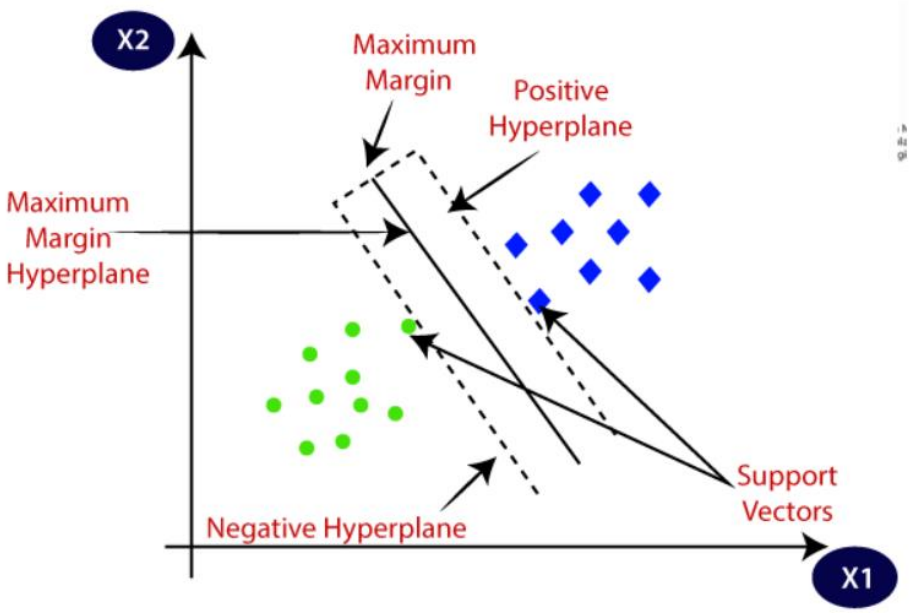

Fig. 4.1.1

Logistic Regression

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression). Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

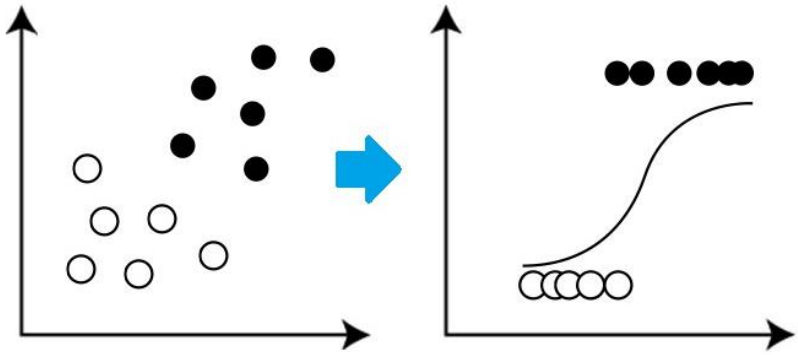

Fig. 4.1.2

4.2 CODING

1. Stop Words: They are the English words which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence. For example, the words like the, he, have etc. Such words are already captured this in corpus named corpus.

```
#Removing Stop Words
final_words = []
for word in tokenized_words:
    if word not in stopwords.words('English'):
        final_words.append(word)
```

2. Tokenization: In Python tokenization basically refers to splitting up a larger body of text into smaller lines, words or even creating words for a non-English language.

```
#Using word_tokenize because it's faster than split()
tokenized_words = word_tokenize(cleaned_text, "english")
```

5. PERFORMANCE ANALYSIS

5.1 TESTING

5.1.1 Before

Output, taking into consideration a small set of reviews.

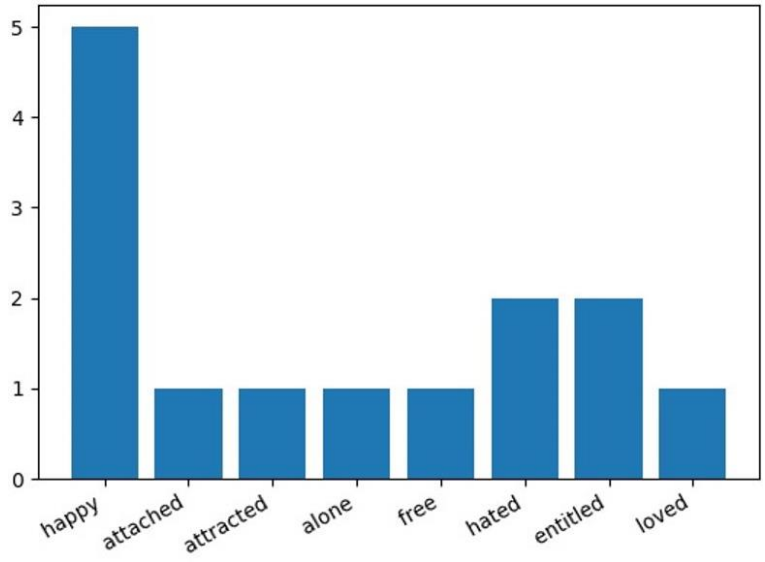

Fig. 5.1.1

5.1.2 After

Output after taking the final dataset containing large set of reviews.

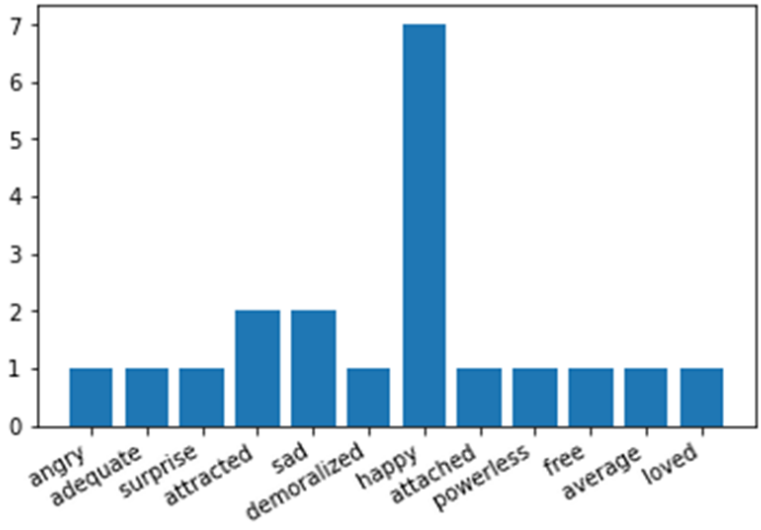

Fig. 5.1.2

We have successfully classified IMDb reviews in an array of sentiments using SVM classifier, and we also predicted the overall sentiment of the review. Each of these labels specifies different sentiments present in each part of the review.

6. FUTURE ENHANCEMENTS AND CONCLUSION

6.1 CONCLUSION

For sentiment analysis a methodology has been designed which integrates existing sentiment analysis approaches. The system is accurate enough for the test case of Avengers reviews on IMDb. Classification of reviews along with sentimental analysis increased the accuracy of the system which in turn provides accurate reviews to the user. This work could be extended to make the system of numbered star rating more useful to users.

6.2 FUTURE WORK

This work could be extended to make the system of numbered star rating more useful to users.

REFERENCES

1. Xing Fang & Justin Zhan – Sentiment analysis using product review data – Journal of Big Data volume 2, Article number: 5 (2015)

2. Faisal Muhammad Shah, Nudrat NAWAL Saber, Tanjim Ul Haque – Sentiment Analysis on Large Scale Amazon Product Reviews ǁ ICIRD At: Bangkok, Thailand (June 2018 DOI)

3. Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande – Amazon Review Classification and Sentiment Analysis || IJCSIT Vol. 6 (6) , 2015, 5107-5110

4. A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques", 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015), Procedia Computer Science, vol. 57, 2015, pp. 821 – 829

5. P. Vijayaragavan, R. Ponnusamy, and M. Aramudhan, "An optimal support vector machine based classification model for sentimental analysis of online product reviews", Future Generation Computer Systems, vol. 111, 2020, 234–240.

6. https://www.imdb.com/title/tt0848228/reviews?ref_=tt_urv

7. K. Ashok Kumar, C. Jagadees, Pravin Kshirsagar, Swagat. M. Marve - Sentiment Analysis of Amazon Product Reviews using Machine Learning || January-February 2020 ISSN: 0193-4120 Page No. 5245 - 5254

8. Krishna V Pad - NLTK Sentiment Analysis || Jun. 17, 19 · AI Zone

9. TextBlob: Simplified Text Processing Release v0.16.0. (Changelog)

10. https://en.wikipedia.org/wiki/Natural_language_processing#:~:text=Natural%20language%20processing%20(NLP)%20is,amounts%20of%20natural%20language%20data.

11. Understanding Sentiment Analysis: What It Is and Why It's Used Understanding Sentiment Analysis: What It Is and Why It's Used, Available at: https://www.brandwatch.com/blog/understanding-sentiment-analysis/

12. Sentiment Analysis Explained, Available at: https://www.lexalytics.com/technology/sentiment-analysis

13. Sentiment Analysis, Available at: https://insightsatlas.com/sentiment-analysis/

14. A. Collomb, C. Costea, D. Joyeux, O. Hasan and L. Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation", Available at: https://liris.cnrs.fr/Documents/Liris-6508.pdf.

15. G.Angulakshmi , Dr.R.ManickaChezian ,"An Analysis on Opinion Mining: Techniques and Tools". Vol 3(7), 2014 www.iarcce.com..

16. Miao, Q., Li, Q., & Dai, R. (2009). AMAZING: A sentiment mining and retrieval system. Expert Systems with Applications, 36(3), 7192-7198.

17. Shaikh, Tahura, and DeepaDeshpande. "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews.",(2016)

18. Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER, 2017