Analysis of Amazon Cell Phone Reviews

U15CS705R - COMPREHENSION AND TECHNICAL REPORT

Activity 2

submitted by

ANANTHA PRIYA K (1517102011)

APSARA S (1517102014)

ARSHIA THASNIM S (1517102016)

VII Semester

COMPUTER SCIENCE AND ENGINEERING

SONA COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

ANNA UNIVERSITY: CHENNAI 600 025

December 2020

ANNA UNIVERSITY: CHENNAI-600 025

BONAFIDE CERTIFICATE

This is to certify that the technical report entitled "**ACTIVITY 2 ANALYSIS OF AMAZON CELL PHONE REVIEWS"** is the bonafide report of **"ANANTHA PRIYA K (1517102011), APSARA (1517102014), ARSHIA (1517102016)"** of B.E Computer Science and Engineering during the year 2020 – 2021.

SIGNATURE

Dr.B.SATHYABHAMA M.Tech,Ph.D.,
HEAD OF THE DEPARTMENT
Professor
Department of Computer Science and Engineering,
Sona College of Technology,
Salem.

SIGNATURE

Dr.A.C.KALADEVI M.E,Ph.D.,
STAFF IN-CHARGE
Professor
Department of Computer Science and Engineering,
Sona College of Technology,
Salem.

Submitted for comprehension and technical report (U15CS705R) examination held on 09.01.2021

_____
Examiner

Analysis of Amazon Cell Phone Reviews

Anantha Priya K, Arshia Thasnim S, Apsara S

Final Year Students, Department of Computer Science and Engineering,
Sona College of Technology, Salem – 636005

priyaanantha3@gmail.com, arshiathasnim1999@gmail.com, apsaranaughty@gmail.com

ABSTRACT

In recent days, online shopping is becoming more and more popular, so the sellers ask their customers to share their reviews of the product they bought. Because of this, millions of reviews are being generated every day. So it will be difficult for the customers to analyze all the reviews and take their decision.

Over 90 per cent of the consumers read online reviews before they decide to purchase any mobile phone from any e-commerce website. Online Mobile applications have revolutionised the way consumers purchase mobile phones online as these apps have all the information regarding any mobile phone at users fingertips. Amazon is one of the best mobile applications which is considered as a treasure trove of all mobile reviews, and their review system is accessible across all channels presenting reviews in an easy-to-use format. So, there should be a system which analyses thousands of reviews of unlocked mobile phones sold on Amazon.com to find insights with respect to reviews, ratings, price and their relationships. If the customers get a clear review of the products and services it will be easier for the customer to purchase the product.

This paper identifies the problem of classifying reviews by their overall semantic which is positive and negative and then a web application has been integrated .

Keywords: Natural Language Processing, text classification, User Reviews, Sentiment analysis, Web application

I. INTRODUCTION

Online marketplaces are becoming popular during these days and reviewing the products before shopping is more common since the customers are mostly inclined to reviews before buying a product. So the online sellers ask their purchasers to share their reviews about the mobile phones they buy. Every day millions of reviews are counted on the Internet about mobile brands, services and quality. This has made the use of the Internet as the most important source of getting ideas and opinions about a product or a service. However, as the number of reviews available for a product grows, it is becoming more mandatory, the different opinions about the same mobile phone on one hand and ambiguous reviews, on the other hand, makes customers get confused to take their decisions.

Here analysing these contents seems difficult for all the online business. The analysis and classification is a computational study which attempts to address this problem by extracting subjective information from the given texts in natural language processing (NLP) techniques such as opinions and reviews *Figure 1*. Different approaches have been used to overcome this problem from natural language processing, text analysis, computational linguistics and biometrics. In recent years, AI methods have got popular in the semantic and review analysis for simplicity and accuracy.

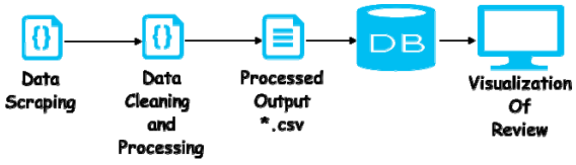


Figure 1: TEXT PROCESSING LAYOUT

Amazon is one of the e-commerce platforms that people are using daily for their online purchases where they can read thousands of reviews commented by the other customers about the mobile phones they wish to get. These reviews provide valuable opinions about mobile phones such as its quality and recommendations which helps the customers to understand almost every detail of a mobile phone and give them a clear opinion about it. This is not only beneficial for costumers but also helps the sellers who are producing their own

products to understand the costumers and their needs better and better. This report is considering the problem for online reviews using supervised approaches to determine the overall semantics of customer reviews by classifying them into positive and negative reviews.

II.LITERATURE REVIEW

In [1] the authors have proposed to find opinions of the customers, the organization conducted surveys about the product which the customers bought and also took the survey for the services. After the drastic growth of the web, it was easier for the customer to post the reviews of the product at the merchant site and express their views. This is from the set of reviews from Amazon for the product iPhone 5.After each review, it is very obvious that the reviews written by the customers were mostly on the services provided by Amazon and it's the seller. So it is decided to classify the reviews into service, product and feature-based reviews. The review rating given by customers which is completely bad gives either 1 or 2 out of 5. This mainly concentrates on finding out the features and getting approximate sentiment which is positive and negative from it.

In [2] the authors have proposed on classifying the reviews based on sentiments i.e positive and negative reviews. Millions of reviews are generated everyday which makes it difficult for the manufacturer to keep track of the Customers opinions. We use Sentimental analysis which is Natural Language Processing (NLP) technique to identify and extract information of text sources and it helps in analyzing the written reviews of the customers and classifying them into positive and negative opinions. These sentiments have a different interpretation in different domains. For example, "This camera sucks" expresses Negative opinion but "This vacuum cleaner sucks" is a positive opinion. For this, they proposed a lexicon-based algorithm for classification. For each sentence of the review, they identify whether it is expressing positive or negative comment and then a final review of the product is identified.

In [3] the authors have proposed that from the consumer reviews of the product in Amazon we collect the datasets which include the type, name of the product and also the text review and rating of the product. For better understanding, first, they extracted the rating and reviews part since it is the main part of this paper. They also used two methods traditional and 50-d glove dictionary. Since both methods were not accurate, we used Naive Bayes which is the first way of representing texts and reviewing the text using glove dictionary. Also using the Long short term memory (LSTM)they have tried to input the original text and found more reviews were of length less than 100 words. So the maximum length is 100 words and words which were more than 100 were removed and the data was trained using LSTM. It gave the resulting training and test accuracy for all the products.

In [4] the authors have proposed that one of the most popular shopping sites is Amazon. In Amazon after every purchase, the buyer is asked to give a review about the product to make it useful for future customers. In The SemEval task, the participants classify among three classes (positive, neutral, negative). we have decided to further improve the performance of the web is and observed improvement. Inspired by this ESWC is been applied for binary polarity detection of customer review from Amazon. Using that experience the robustness, the importance of ensemble learning is understood. And numerous research efforts have also studied the same task on reviews and we have addressed the scalability issues.

In [5] the authors have proposed that the categorization of the positive and negative feedback of the customer and building supervised model to polarize a large number of reviews is done. Both manual and active is used to label the dataset's data scraping from the amazon URL is used to get and process data. Active learning system tries to solve data labelling bottleneck by querying for unlabeled instance to be properly labelled by an expert or oracle.After getting the data it is processed and we extract features. The combination of two kinds of approach to extract features are the bag of words and TF-IDF and chi-square approach the description of the basic theory behind the model is done and the accuracy is over 90%.

In [6] the authors have proposed that the sentiment analysis which has proven to understand the popularity of a product is considered.Multiple studies about sentiment analysis have done using ML, NB, SVM and decision trees resulted in a relatively good performance. The results from multinomial naive Bayes and linear support vector machine were also satisfying but the TF-IDF vectorization limits the data, tokenizing and training the data gave best results. in this study, the assumption is made that the stars are in correspondence with the sentiment of the review. Some cases may have 1 star with a positive review this is checked with another dataset and

classify the review by a model trained on this data. Cross-validation eliminates the risk that an easy to classify is biasing your result. As a result, the most suitable network for binary sentiment analysis on amazon review is LSTM.

In [7] the authors have proposed to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. A general process for sentiment polarity categorization is proposed with detailed process descriptions. Data used in this study are online product reviews collected from Amazon.com. Experiments for both sentence-level categorization and review-level categorization are performed with promising outcomes. At last, they have also given insight into future work on sentiment analysis.

In [8] the authors have proposed to evaluate the compatibility of Amazon.com reviews with their corresponding ratings. Sentiment analysis is the task of identifying and classifying the sentiment expressed in a piece of text as being positive or negative. On e-commerce websites such as Amazon.com, consumers can submit their reviews along with a specific polarity rating. In some instances, there is a mismatch between the review and the rating. To identify the reviews with mismatched ratings we performed sentiment analysis using deep learning on Amazon.com product review data. Product reviews were converted to vectors using paragraph vector, which then was used to train a recurrent neural network with the gated recurrent unit.

In [9] the authors have proposed the review to give the nearly full picture of sentiment analysis, its sorts and characterization. The primary commitments of this paper include the complex orders of late articles and the outline of the ongoing pattern of research in the sentiment analysis and its related territories.

III.PROPOSED SYSTEM

Electronic commerce is becoming increasingly popular due to the fact that e-commerce websites allow purchasers to leave reviews on different products. Millions of reviews are being generated every day by costumers which makes it difficult for product manufacturers to keep track of customer opinions of their products. Thus, it is important to classify siuch large and complex data in order to derive useful information from a large set of data. Classification methods are the way to tackle such problems.

Sentiment analysis

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) problem which means identifying and extracting subject information of text sources. The purpose of sentim classification is to analyze the written reviews of users and classify them into positive or negative opinions, so the system does not need to completely understand the semantics of each phrase or document. This however is not done by just labelling words as positive or negative. There are some challenges involved. *Figure 3.1,*Classifying words and phrases with prior positive or negative polarity will not always work. For example, the word “amazing” has a prior positive polarity, but if it comes with a negation word like “not”, the context can completely change. The word ”unpredictable” camera has a negative meaning to that camera while “unpredictable” experience is considered as positive for tourists.
Sentiment classification has been attempted in different fields such as movie reviews, travel destination reviews and product reviews. Lexicon based methods and machine learning methods are two main approaches that are usually used for sentiment classification.

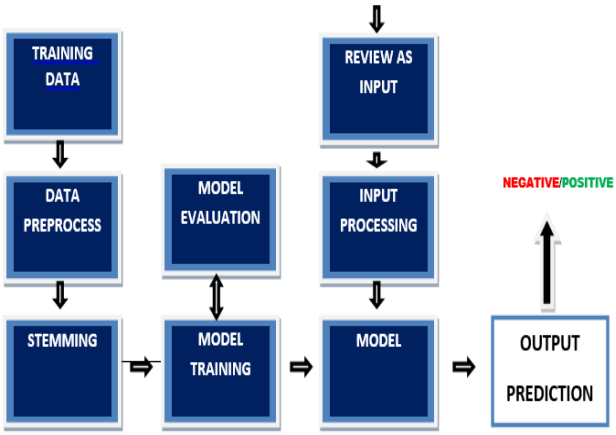


Figure 3.1: TEXT PROCESSING

Sentiment classification using Lexicon based methods

Lexicon based method is another unsupervised approach, which relies on word and phrase annotation. To compute a sentiment score for each text, this method uses a dictionary of sentiment words and phrases . In lexicon-based methods the simplest approach for determining the sentiment of a review document where is to use a count-based approach. If we have a text and a lexical resource containing the positive and negative annotation of words and

phrases, we can assign the polarity of the review. This means that if the number of positive words is more than the negative ones the polarity of the review is

positive. If there are more negative sentiment words than positive sentiment words, the overall sentiment of the text is then negative. However, using only sentiment words and phrases for sentiment classification is not enough. Sentiment lexicon for sentiment analysis is necessary but it is not sucient.
There are some issues involved with this method that are:

Positive or negative sentiment words may have different interpretation in different domains. For instance, the word ”suck” usually have a negative sentiment, but it can also indicate a positive sentiment. For example, ”This camera sucks,” expresses a negative opinion but ”This vacuum cleaner really sucks.” is a positive opinion.

Sarcastic sentences are usually hard to deal with even if they contain sentiment words. For example, ”What a great car! It stopped working in two days.” This is a negative opinion even though it contains the word ”good” which is a positive word.

It can happen that a phrase or opinion does not have a sentiment word, so it makes it hard for the machine to determine the compute a sentiment score for the opinion. ”This washer uses a lot of water” has no sentiment words but it implies a negative opinion about the washer.

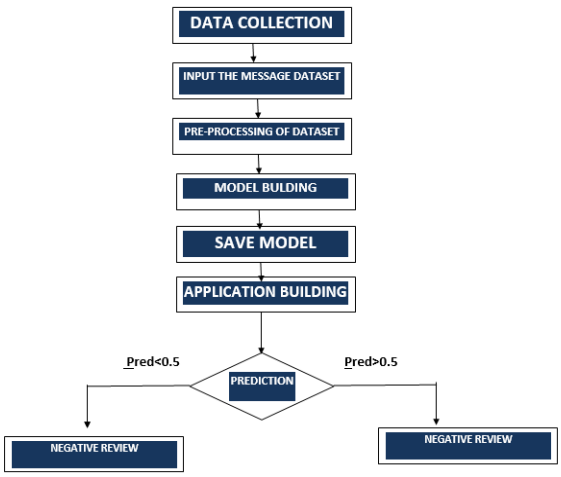


Figure3.2: FLOW CHART OF PREDICTION PROCESS

APPLICATION DEVELOPED

Python Language has been used in building up the model. The main libraries used in this process were Tensorflow, keras, StemmerPorter .The features in this applications are *Figure3.2*:
Preprocessing - The messages have to be pre-processed for the removal of unwanted punctuation, grammar, stop words etc.
Label Encoding - Label Encoder encodes labels with values between 0 and n-1 where n represents the number of distinct labels for the classes. Same value is as assigned to the labels which are repeated earlier. In our experiment, we convert the class labels to binary values, where 0 is positive and 1 is negative.
Stop word removal - When using Natural Language Processing (NLP), our goal is to perform some analysis or processing so that it can respond to the text immediately. A machine cannot understand the human readable form. So, data has to be pre-processed in order to make it machine-readable. This is “pre-processing” of which one of the major forms is to filter out useless data. This useless data (words) is generally referred as stop words in Natural Language Processing(NLP).
Stemming - Stemming is another pre-processing step that normalize sentences. Stemming is a way to account for the variations of words and sentences which often have a same meaning; furthermore, it will help us shorten the sentences and shorten our lookup.
For pre-processing Natural Language Processing (NLP) techniques have been implemented such as Flask application has been used for the Front-end application development. The application’s UI has been developed using HTML, CSS, JavaScript are used.

IV. CONCLUSION

In this research, we proposed an Artificial intelligence application to solve a prevailing problem in Amazon website where the user can get information about the review is positive or negative. Reviews are a very important source of information for a potential customer before deciding to purchase a mobile. To construct our model, we first used NLP for text processing and then integrated our model into Flask and developed a web application. Our accuracy rate was about 95.7%. This application will increase the efficiency of the user to find the mobile phone reviews are good or bad.

REFERENCE

[1] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande,"Amazon Review Classification and Sentiment Analysis", International journal of Computer Science and Information technologies (IJCSIT), Computer Department,Sardar Patel Institute of Technology,Mumbai, published 2015

[2] Sepideh Paknejad,"Sentiment classification on Amazon reviews using machine learning approaches",KTH royal institute of technology School of Electrical Engineering and Computer Science,Stockholm,Sweden, published 2018

[3] Wanliang Tan, Xinyu Wang, Xinyu Xu,"Sentiment Analysis for Amazon reviews", published 2018

[4] Efstratios Sygkounas, Giuseppe Rizzo, Raphael Troncy,"Sentiment Polarity Detection From Amazon Reviews:An Experimental Study", EURECOM,SophiaAntipolis,France, ISMB,Turin,Italy ,13th European Semantic Web conference, Semantic Sentiment Analysis Challenge,Herakloin,Greece May 29-June 2,2016

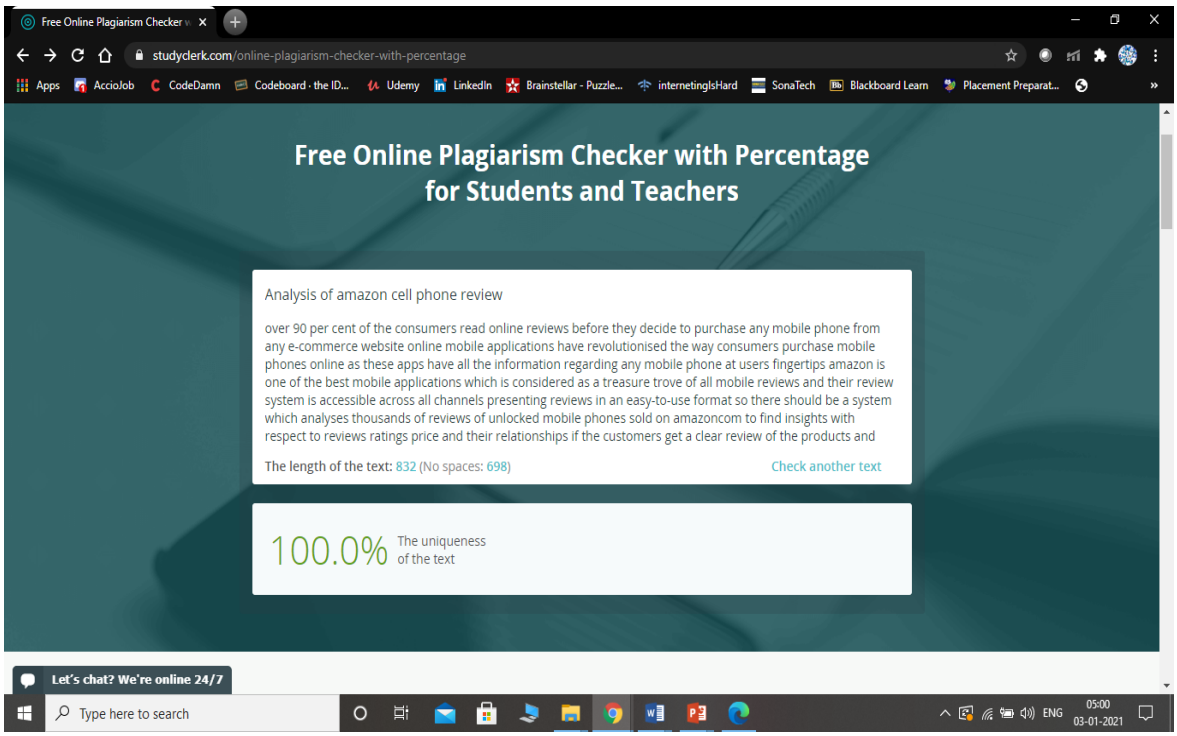
[5] Tanjim UI Haque, Nudrat Nawar Saber, Faisal Muhammad Shah,"Sentiment Analysis on Large Scale Amazon Product Reviews", International Conference on Innovative Research and Development(ICIRD),Computer Science and Engineering,Ashanullah University of Science & Technology,Dhaka, Bangladesh, published 3 June ,2019

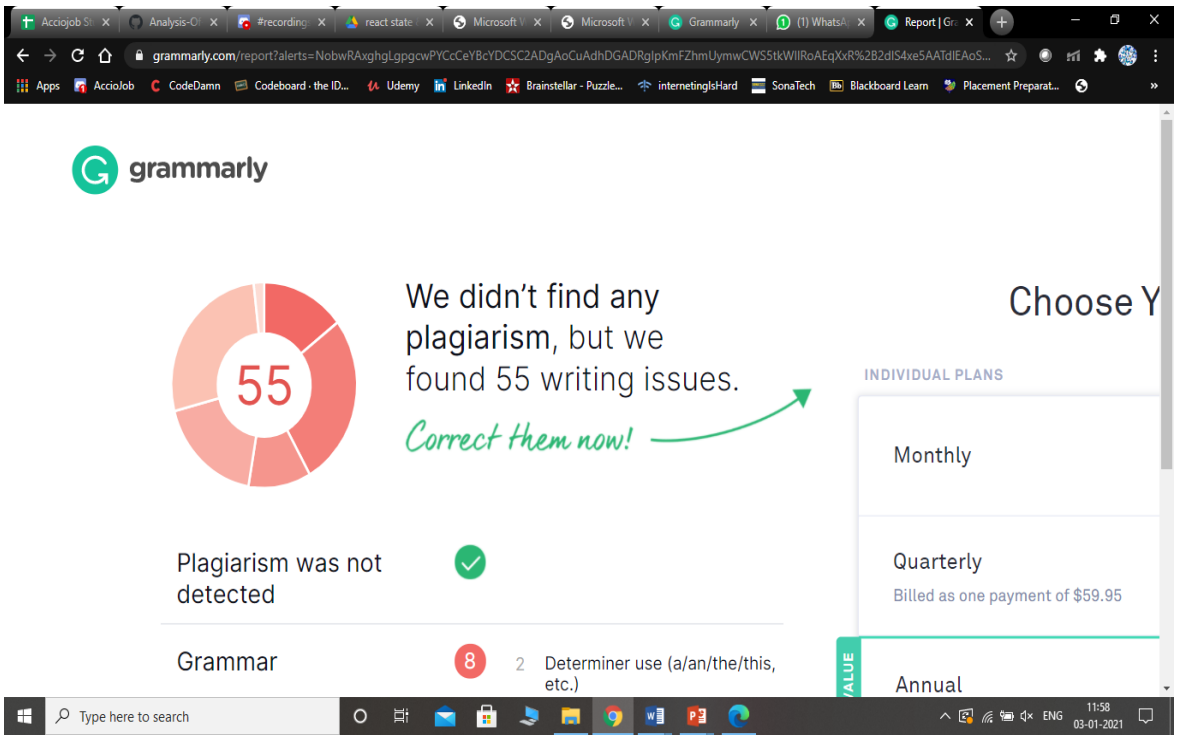
[6] Levent Guner, Emilie Coyne, Jim Smit, "Sentiment Analysis for Amazon.com reviews" ,Big Data in Media Technology(DM2583) , KTH Royal Institute of Technology, Stockholm, published 24 April 2019

[7] Xing Fang, Justin Zhan ,"Sentiment Analysis using product review data",Fang and Zhan Journal of Big Data, published 2015

[8] Nishit Shrestha, Fatma Nasoz,"Deep Learning Sentiment Of Amazon.com reviews and ratings", Department of Computer Science, University of Nevada Las Vegas,Nevada,USA, International journal on Soft Computing, Artificial Intelligence and Applications(IJSCAI), published 1, February 2019

[9] Dishi Jain, Bitra Harsha Vardhan, Saravana Kumar Kandasamy, "Sentiment Analysis Of Product Reviews-A Survey", International journal of scientific and technology research, published 12, December 2019


Free Online Plagiarism Checker
studyclerk.com/online-plagiarism-checker-with-percentage
Free Online Plagiarism Checker with Percentage for Students and Teachers
Analysis of amazon cell phone review
over 90 per cent of the consumers read online reviews before they decide to purchase any mobile phone from any e-commerce website online mobile applications have revolutionised the way consumers purchase mobile phones online as these apps have all the information regarding any mobile phone at users fingertips amazon is one of the best mobile applications which is considered as a treasure trove of all mobile reviews and their review system is accessible across all channels presenting reviews in an easy-to-use format so there should be a system which analyses thousands of reviews of unlocked mobile phones sold on amazoncom to find insights with respect to reviews ratings price and their relationships if the customers get a clear review of the products and
The length of the text: 832 (No spaces: 698)
Check another text
100.0% The uniqueness of the text
Let's chat? We're online 24/7
Type here to search
ENG
05:00
03-01-2021


grammarly.com/report?alerts=NobwRAxghgLgpgcwPYCcCeYBcYDCSC2ADgAoCuAdhDGADRgIpKmFZhmUymwCWS5tkWIlRoAEqXxR%2B2dlS4xe5AATdIEAoS...
grammarly
55
We didn't find any plagiarism, but we found 55 writing issues.
Correct them now!
Choose Y
INDIVIDUAL PLANS
Monthly
Quarterly
Billed as one payment of $59.95
Annual
Plagiarism was not detected
Grammar
8
2 Determiner use (a/an/the/this, etc.)
Type here to search
ENG
11:58
03-01-2021