

Sentiment Analysis for Amazon Products using Isolation Forest



S. Salmiah, Dadang Sudrajat, N. Nasrul, Tuti Agustin, Nisa Hanum Harani, Phong Thanh Nguyen

Abstract: To text examination for efficiently recognize, evaluate, study full of affective states, extricate use of normal language preparing is known as Sentiment Analysis. The applications in which use advertising to client administration to clinical drug like applications use sentiment analysis on the web and web-based social networking, human services materials and audits and study reactions. Many sites like Amazon urged users to post the review of the product on its site. But Amazon provides the limit of content to post the reviews. For different applications the review helps to analyze the product although the review for several products will differ. This research works on the data that is recovered from Amazon and apply and expand the present work in the field of sentiment analysis and natural language processing. The work uses Machine Learning algorithms and characterize into positive or negative surveys.

Keywords: setiment analysis; natural language processing; Amazon; machine learning.

I. INTRODUCTION

Now in these days social media is the very efficient platform to promote any services and products. to connect with the clients and users many companies uses Facebook and Twitter accounts. To get information about product and services users also use social media platform. For purchasing the services and goods social media plays a very important role and it change the way of selling and purchasing the products. In Natural Language Processing (NLP) Sentiment analysis is another field of research lately. The objective of assessment investigation is to distinguish and remove the subjectivity in content and furthermore arrange the suppositions and feelings. The methodologies of estimation examination study the suppositions, outlooks, opinions, evaluations, emotions towards organizations, affiliations, subject's events of individuals.



Figure 1: Sentiment analysis

Manuscript published on 30 August 2019.

* Correspondence Author (s)

S. Salmiah, Universitas Sumatera Utara, Medan, Indonesia.

Dadang Sudrajat, STMIK IKMI Cirebon, Indonesia.

N. Nasrul, Universitas Halu Oleo, Indonesia.

Tuti Agustin, Department of Civil Engineering, Faculty of Engineering, Sebelas Maret University, Indonesia.

Nisa Hanum Harani, Applied Bachelor Program of Informatics Engineering, Politeknik Pos Indonesia, Indonesia.

Phong Thanh Nguyen, Department of Project Management, Ho Chi Minh City Open University, Vietnam.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

II. LITERATURE SURVEY

1. “Sentiment Analysis of Customer Product Reviews Using Machine Learning[1]”

The author of this paper described that from a huge set of reviews it is necessary to separate positive and negative review. From the text to extract subjective data or information sentiment analysis use as a computational study.

2. “Sentiment Mining Of Online Reviews Using Machine Learning Algorithms [2]”

The author of this paper introduced a sentiment analysis method that is based on review of user and this method use hybrid algorithm. Based on a specific item the public review analysis is very complicated task. This approach include many perspective like score measure, algorithm classification, preprocessing etc.

3. “A Survey on Sentiment Analysis in NLP [3]”

The author of this paper proposed a combined view of sentiment analysis. This paper also described Common Knowledge Base. For analyzing sentiment analysis Hour Glass Model defined. For creating a system of intelligent opinion-mining Hour glass model was used. Semantic knowledge can handle by this model.

4. “Sentiment Analysis and Classification: A Survey [4]”

The author of this paper shows that there is lots of scope in the field of sentiment analysis and there are many challenges also. Sentiment analysis has many applications in the field of e commerce. It can summarize the review, classify the review and can handle the real time applications. This paper descried sentiment analysis tools and classification methods.

5. “Reviews on Opinions Mining and Summarization Methods for Analysis of Unstructured Textual Data for Decision Making [5]”

The author of this paper proposed the method in which based on the opinions of people that expressed freely. This paper also discussed the previous approaches that already existing in opinion mining. It takes the review of available approaches in the field opinion mining and analyze these approaches to achieve the decision making efficiently.

6. “A Survey on Sentiment Analysis Applied in Opinion Mining [6]”

The authors of this paper give a review of different methods and techniques in different field of sentiment analysis. In is basically based on technique of opinion mining like Supervised, Machine learning techniques, CBR, Supervised etc.

7. A Survey on Text Mining and Sentiment Analysis for Unstructured Web Data [8]

The authors of this paper described the basic idea of slant investigation. This paper also works on about Common Knowledge Base approach. For slant examination the researcher used Hour Glass Model.

8. "Trend Analysis in Social Networking using Opinion Mining a Survey [9]"

The authors of this paper proposed that to get the information about the communication between different groups of people the social media plays a very important role. This approach is based on opinion of users.

9. "A Survey on Sentiment Analysis and Opinion Mining [10]"

The author of this paper presents that for showing the opinions, attitudes, emotions and appraisals of peoples sentiment analysis is necessary. It shows the opinions towards organizations, people, attributes, topics, events and issues.

III. EXISTING METHODOLOGY

In Natural Language Processing (NLP) Sentiment analysis is another field of research in recent years. The goal of sentiment analysis is to identify and extract the subjectivity in text and also classify the sentiments and opinions. The approaches of sentiment analysis study the suppositions, frames of mind, sentiments, assessments, feelings towards administrations, associations, themes occasions of people [12]. Text is classified according to following basis in sentiment analysis:

1. Based on the sentiment expressed polarity. It can neutral, positive or negative
2. Based on outcome polarity. It can in medical text improvement versus death
3. Based on favorable and unfavorable on any topic. Example is political database
4. Based on any good or bad news.
5. Based on give support to any condition or stand in opposition
6. Based on advantages and disadvantages.

IV. IMPLEMENTATION TECHNOLOGY

1. Jupyter Notebook

For creating the Jupyter notebook documents, Jupyter Notebook is a interactive computational environment which is based on web. It is also known as Python Notebook. The term "notebook" collectively related with many different attributes like Jupyter document format that depend on context, Jupyter Python web server and Jupyter web application. A document of Jupyter Notebook is known as JSON document. This document follows the versioned schema, and contains the ordered list of input and output cells; these cells have text, code, rich media, mathematics and plots. The document has extension of ".ipynb" [13].

2. Python Notebook interface

It provides many open source libraries. It is based on REPL. Some libraries are:

- IPython
- ØMQ
- Tornado (web server)
- jQuery
- Bootstrap (front-end framework)

3. Jupyter kernels

For handling the several requests like code completion, inspection, code execution a Jupyter kernel is a responsible program. It also provides reply to these operations. Over the network through the use of ZeroMQ, Jupyter talk to other components. In this way it can be on remote machine or on same machine [11].

V. PROPOSED METHODOLOGY

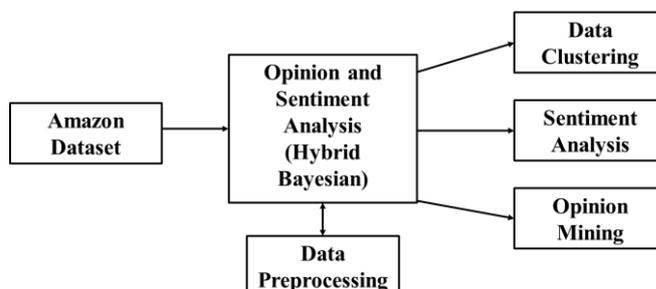


Figure 2: Proposed Methodology

VI. POSITIVITY AND AMAZON

As taking the example of Amazon the products of Amazon get many positive reviews than negative reviews. It seemed well and good to acquire the most reviews that get 1 and 2 stars consider as negative review and it limit the quantity of the positive review until the two approached one another and the most continuous sense standard rose to half. it may happen that some positive review will lost and all the negative review were saved. But as most incessant features can extract, in the positive set of review the vital feature will preserved [12-14]. Endeavor to produce explicit highlights only for them, the best characterizing framework will just train on negative review, and then show the positive label.

Steps in Proposed System

- Step 1 : Preprocess raw reviews to cleaned reviews
- Step 2 : Create BoW using CountVectorizer / Tfidfvectorizer in sklearn
- Step 3 : Transform review text to numerical representations (feature vectors)
- Step 4 : Fit feature vectors to supervised learning algorithm (eg. Naive Bayes, Logistic regression, etc.)
- Step 5 : Improve the model performance by GridSearch

Project Flow

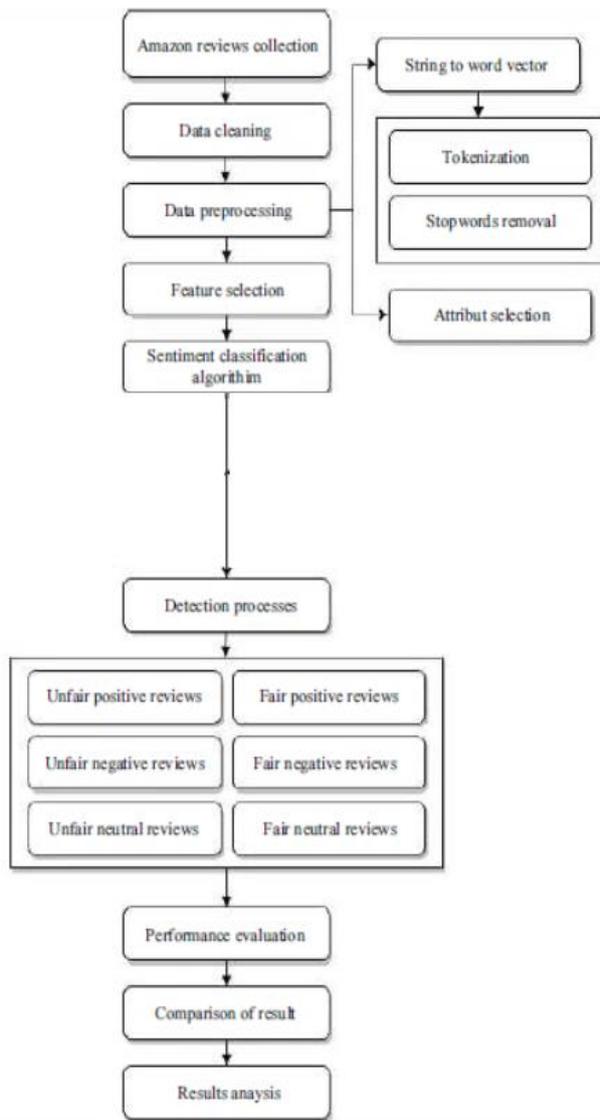


Figure 3: Flow of project

Algorithm to calculate the review orientation

1. Procedure Review Sen()
2. start
3. for the determined each review sentence sen
4. start
5. sen = 0;
6. For the each review belong to word r in sen
7. sen += Word Sense (r, sen);
8. /* Pos = 1 , Neg = -1*/
9. if (sen > 0) sen = Pos;
10. else if (sen = Neg)
11. end for;
12. stop

Algorithm 2:

1. Procedure Word Sense (wrđ, sent)
2. start
3. sen = introduction of the determined words ;
4. if(there are NEG_WRD appeared close and around to the words in sent)
5. sen = opposite(sent);
6. stop

VII. ALGORITHM IMPLIMENTATION

Algorithm 1: Dataset Load Section

| | Product Name | Brand Name | Price | Rating | Reviews | Review Votes | Sentiment |
|--------|---|----------------|--------|--------|---|--------------|-----------|
| 230499 | LG Nexus 5X Unlocked Smart Phone, 5.2" Quartz... | LG Electronics | 399.00 | 5 | I love Nexus phones. My 2012 Nexus 4 still run... | 0.0 | 1 |
| 331895 | Samsung Galaxy Note II N7100 16GB White-Unlock... | Samsung | 618.30 | 5 | the best phone i ever knowCamera is not really... | 1.0 | 1 |
| 287192 | OtterBox Rugged Defender Series Case w/ Holste... | OtterBox | 21.99 | 5 | Came in the box and before it was promised! | 0.0 | 1 |
| 280029 | OtterBox 77-29864 Defender Series Hybrid Case ... | OtterBox | 9.99 | 5 | I love this, before I used to be really nervou... | 0.0 | 1 |
| 53491 | Apple iPhone 6 64GB Factory Unlocked GSM 4G LT... | Apple | 579.95 | 4 | The only compliant is that the battery does no... | 1.0 | 1 |

Table 1: Overview of Dataset

