A Project Synopsis on

**Amazon Sentimental Analysis**

Submitted to Manipal University, Jaipur

Towards the partial fulfillment for the Award of the Degree of

**BACHELORS OF TECHNOLOGY (Bold, UPPERCASE, 14pt.)**

In Computers Science and Engineering

2018-2022

By

Ashray Gupta

189301068



Under the guidance of (12pt.)

Prof. Mahesh Jangid

**Department of Computer Science and Engineering**

**School of Computing and Information Technology**

**Manipal University Jaipur**

**Jaipur, Rajasthan**

Introduction

Recent years have seen an increasing amount of research efforts expanded in understanding sentiment in textual resources. As we can see the statistics from web of knowledge in Figure one, the papers published on sentiment analysis have been increasing for the past years. One of the subtopics of this research is called sentiment analysis or opinion mining, which is, given a bunch of text, we can computationally study people opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. Applications of this technique are diverse.

For example, businesses always want to find public or consumer opinions and emotions about their products and services. Potential customers also want to know the opinions and emotions of existing users before they use a service or purchase a product. Last but not least, researchers uses these information to do an in-depth analysis of market trends and consumer opinions, which could potentially lead to a better prediction of the stock market.

Sentiment analysis of product reviews, an application problem, has recently become very popular in text mining and computational linguistics research. Here, we want to study the correlation between the Amazon product reviews and the rating of the products given by the customers. We use both traditional machine learning algorithms including Naive Bayes analysis, Support Vector Machines, K-Nearest Neighbor method and deep neural networks such as Recurrent Neural Network (RNN), Recurrent Neural Network (RNN). By comparing these results, we could get a better understanding of these algorithms. They could also act as a supplement to other fraud scoring detection methods.

Motivation

So far, there are a lot of work related to product reviews, sentiment analysis or opinion mining. For example, Xu Yun el al from Stanford University applied existing supervised learning algorithms such as perceptron algorithm, naive bayes and supporting vector machine to predict a review’s rating on Yelp’s rating dataset. They used to hold out cross validation using 70% data as the training data and 30% data as the testing data. The author used different classifiers to determine the precision and recall values.

Maria Soledad Elli and Yi-Fan extracted sentiment from the reviews and analyze the result to build up 1 a business model. They claimed that this tool gave them pretty high accuracy. They mainly used Multinomial Naive Bayesian (MNB) and support vector machine as the main classifiers. Callen Rain proposed extending the current work in the field of natural language processing. Naive Bayesian and decision list classifiers were used to classify a given review as positive or negative.

Deep-learning neural networks is also popular in the area of sentiment analysis. Ronan Collobert et al used a convolutional network for the semantic role labeling task with the goal avoiding excessive task-specific feature engineering. On the other hand, in paper, the authors proposed proposed using recursive neural networks to achieve a better understanding compositionality in tasks such as sentiment detection.

In this paper, we want to apply both traditional algorithms including Multinomial Naive Bayesian, Supporting Vector Machine and Logistics Regression along with VADER and Text Blob. By comparing the accuracy of these models, we would like to get a better understanding how these algorithms work in tasks such as sentiment analysis.

Project Objective

# PROJECT/ROLE DESCRIPTION

* Go to amazon.in and scrape the customer reviews of 10 mobile phones.
* Develop a Sentiment Analysis model using the data provided as training data.
* Using the model, you developed to evaluate the sentiment of the customer reviews scraped by you.

# Objectives Achieved

* Successfully made a code to scrap Amazon reviews by just giving the URL of the product.
* Feasibility of changing the number of pages of reviews for scraping.
* Fine Grained Sentiment Analysis performed on the dataset.
* Applied the feature-based ML models as well as Rule based methods on both the Amazon Data
* Made visuals of Word Clouds of words in both title of the reviews as well as comments of the reviews separately.
* Data Visualizations to present Sentiment Analysis results.

# METHODOLOGY

### -Scrap Amazon Reviews

1- Put the urls of the mobile phones whose reviews are to be scraped in "urls.txt".

2- Run the "Reviews\_Scrap.ipynb".

3- The data will be saved in "data.csv".

### -Sentiment Analysis on Amazon Data

1- Put the "data.csv" file in the same directory.

2- Run the "SentimentalAnalysis.ipynb".

# RESULTS

|  |  |
| --- | --- |
| **Name of Mobile Phones used for Analysis** |  |
| 1. Apple iPhone 11 Pro Max (64GB) - Midnight Green | |
| 1. Apple iPhone XR (64GB) - Black | |
| 1. OnePlus 7T Pro (Haze Blue, 8GB RAM, Fluid AMOLED Display, 256GB Storage, 4085mAH Battery) | |
| 1. OnePlus 8 5G (Glacial Green 6GB RAM+128GB Storage) | |
| 1. Oppo A52 (Twilight Black, 6GB RAM, 128GB Storage) with No Cost EMI/Additional Exchange Offers | |
| 1. Redmi 9 (Sky Blue, 4GB RAM, 64GB Storage)| 3 Months No Cost EMI on BFL | |
| 1. Redmi Note 9 Pro Max (Interstellar Black, 6GB RAM, 128GB Storage) - 64MP Quad Camera & Latest 8nm Snapdragon 720G & Alexa Hands-Free | |
| 1. Samsung Galaxy M11 (Black, 4GB RAM, 64GB Storage) with No Cost EMI/Additional Exchange Offers | |
| 1. Samsung Galaxy M21 (Blue, 4GB RAM, 64GB Storage) | |
| 1. Samsung Galaxy M31s (Mirage Blue, 6GB RAM, 128GB Storage) | |
|  |  |
| **Fine Grained Sentiment Analysis** |  |
|  |  |
| Logistic Regression | Train accuracy -0.4614285714285714  Test accuracy -0.45666666666666667 |
| SVC | Train accuracy -0.49857142857142855 Test accuracy - 0.5033333333333333 |
| Multinomial Naïve Bayes | Train accuracy - 0.45071428571428573 Test accuracy - 0.4483333333333333 |
| Text Blob | Positive accuracy for Title = 30.0% via 2000 samples Negative accuracy for Title = 10.0% via 2000 samples  Positive accuracy for Content = 72.0% via 2000 samples Negative accuracy for Content = 14.02% via 2000 samples |
| VADER Analysis | VADER Analysis Title  Positive - 17.3%  Negative - 11.7%  Neutral - 71.0%  VADER Analysis Comments  Positive - 17.3%  Negative - 11.7%  Neutral - 71.0% |

# REFERENCES

1. Analysis for Amazon Reviews by [Wanliang Tan](https://www.semanticscholar.org/author/Wanliang-Tan/35496772), [Xinyu Wang](https://www.semanticscholar.org/author/Xinyu-Wang/48630537)
2. A Practitioner's Guide to Natural Language Processing — Processing & Understanding Text