# AMAZON KINDLE STORE REVIEWS ANALYSIS USING IBM

**WATSON SERVICES**

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

## BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Submitted by

**SRIYAGUDLA 19UK1A0518**

**VAMSHI MAMIDALA 19UK1A0516**

**RAMYA BOLLEPALLY 19UK1A0553**

Under the esteemed guidance of

## Mr. A. ASHOK KUMAR

**(**Assistant Professor)

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING VAAGDEVI ENGINEERING COLLEGE

(Affiliated to JNTUH, Hyderabad) Bollikunta, Warangal – 506005 **2019– 2023**

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINNERING

**ENGINEERING VAAGDEVI ENGINEERING COLLEGE**

## BOLLIKUNTA, WARANGAL – 506005 2019 – 2023



**CERTIFICATE OF COMPLETION INDUSTRY ORIENTED MINI PROJECT**

This is to certify that the UG Project Phase-1 entitled “**AMAZON KINDLE STORE**

**REVIEWS ANALYSIS USING IBM WATSON SERVICES”** Is being submitted by-

***GUDLASRIYA(H.NO:19UK1A0518),MAMIDALAVAMSHI(H.NO:19UK1A0516),BOLL***

***EPALLYRAMYA(H.NO:19UK1A0553),***in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2022-23, is a record of work carried out by them under the guidance and supervision.

**Project Guide Head of the Department**

**Mr. A. ASHOK KUMAR Dr. R. Naveen Kumar**

(Assistant Professor) (Professor)

**External**

# ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr.P.PRASAD RAO,** Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.R.NAVEEN KUMAR**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **A.ASHOK KUMAR,** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

**SRIYAGUDLA (19UK1A0518) VAMSHI MAMIDALA(19UK1A0516)**

**RAMYA BOLLEPALLY(19UK1A0553)**

# ABSTRACT

Amazon Kindle Store is an e-book e-commerce store for all the book reading hobbyists. Online reviews are a category of product information created by users based on personal handling experience. Online shopping websites endow with platforms for consumers to review products and carve up opinions. The problem is most of the comments from customer reviews about the products are contradicted to their ratings. Many customers will post their comments and forgot to rate the product or not engrossed to rate it.

Sentiment mining plays a very important role in business to understand the opinion of customers to improve the products. Customer also depends on the opinion of others who have bought the products already. Reviews or feedback becomes the deciding factor to buy or sell a product. A rating of the products gives a speedy clarification to pact with the product. We will be using Natural language processing to analyse the sentiment (positive or a negative) of the given review.

# TABLE OF CONTENTS:-

1. [INTRODUCTION 1](#_TOC_250010)
   1. OVERVIEW… 1
   2. [PURPOSE 1](#_TOC_250009)
2. [LITERATURE SURVEY 2](#_TOC_250008)
   1. EXISTING PROBLEM 2
   2. PROPOSED SOLUTION 2-3
3. [THEORITICAL ANALYSIS 4](#_TOC_250007)
   1. BLOCK DIAGRAM 4
   2. [HARDWARE /SOFTWARE DESIGNING 4](#_TOC_250006)
4. EXPERIMENTAL INVESTIGATIONS 5
5. [FLOWCHART… 6](#_TOC_250005)
6. RESULTS… 7-8
7. [ADVANTAGES AND DISADVANTAGES… 9](#_TOC_250004)
8. [APPLICATIONS 9](#_TOC_250003)
9. [CONCLUSION 10](#_TOC_250002)
10. [FUTURE SCOPE… 10](#_TOC_250001)
11. [BIBILOGRAPHY 11](#_TOC_250000)
12. APPENDIX (SOURCE CODE) 12-25
    1. **OVERVIEW**

# INTRODUCTION

The objective of this paper is to categorize the positive and negative feedback of the customers over different products and build a supervised learning model to polarize large amounts of reviews. A study on amazon last year revealed more than 80% of online shoppers trust reviews as much as personal recommendations. Any online item with a large amount 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. Quite simply, more reviews look more convincing. People value the consent and experience of others and the review on a material is the only way to understand others' impression on the product. Opinions, collected from users’ experiences regarding specific products or topics, straightforwardly influence future customer purchase decisions. Similarly, negative reviews often cause sales loss. For those understanding the feedback of customers and polarizing accordingly over a large amount of data is the goal. There are some similar works done over amazon dataset. In opinion mining over a small set of dataset of Amazon kindle product reviews to understand the polarized attitudes towards the product.

* Know fundamental concepts and techniques of natural language processing (NLP)
* Gain a broad understanding of text data.
* Know how to pre-process/clean the data using different text preprocessing techniques.
* Know how to build a neural network.
* Know how to build a web application using the Flask framework

### PURPOSE

As the commercial sites of the world are almost fully online platforms, people are trading products through different e-commerce websites. And for that reason reviewing products before buying is also a common scenario. Also nowadays, customers are more inclined towards the reviews to buy a product. 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 and deep learning based algorithms, reading thousands of reviews to understand a product is rather time consuming where we can polarize a review on a particular category to understand its popularity among the buyers all over

* 1. **EXISTING PROBLEM**

# LITERATURE SURVEY

Given a dataset containing of various attributes,use the features available in dataset and define a supervised classification algorithm which can identify wether they getting reviews correct predicted reviews or not. The problem is most of the comments from customer reviews about the products are contradicted to their ratings. Many customers will post their comments and forgot to rate the product or not engrossed to rate it.

### PROPOSED SOLLUTION

All Information in the world can be broadly classified into mainly two categories, facts and opinions. Factsare objective statements about entities and worldly events. On the other hand opinions are subjective statements that reflect people’s sentiments or perceptions about the entities and events . Maximum amount of existing research on text and information processing is focused on mining and getting the factual information from the text or information. Before we had WWW we were lacking a collection of opinion data, in an individual needs to make a decision, he/she typically asks for opinions from friends and families. When an organization needs to find opinions of the general public about its products and services, it conducted surveys and focused groups. But after the growth of Web, especially with the drastic growth of the user generated content on the Web, the world has changed and so has the methods of gaining ones opinion. One can post reviews of products at merchant sites and express views on almost anything in Internet forums, discussion groups, and blogs, which are collectively called the user generated content . As the technology of connectivity grew so as the ways of interpreting and processing of users opinion information has changed. Some of the machine learning techniques like Naïve Bayes, Maximum Entropy and Support Vector Machines has been discussed in the paper . Extracting features from user opinion information is an emerging task.

Preparing Review Database

POS Tagging Feature Extraction

Opinion Word Extraction Opinion Word Polarity Identification

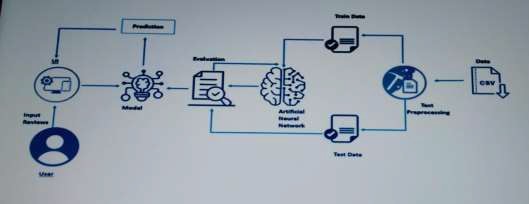
Opinion Sentences Polarity Identification Summary Generation

**Basic Step of Feature Extraction**

a generic model of feature extraction from opinion information is shown, firstly the information database is created, next POS tagging is done on the review, next the features are extracted using grammar rules such as adjective + noun or so on, as nouns are features and adjectives are sentiment words. Next Opinion words are extracted followed by its polarity identification. Some models also calculate sentence polarity for accuracy. Lastly the results are combined to obtain a summary. Many algorithms can be used in opinion mining such as Naive Bayes Classification, Probabilistic Machine Learning approach to classify the reviews as positive or negative, have been used to get the sentiment of opinions of different domains such as movie, Amazon reviews of products. In our work we have used reviews of iPhone 5 extracted from Amazon website. We studied all the reviews and got to know that there are many reviews in which the user talks about the service provided by amazon and its sellers. So we decided to classify reviews into service, product and feature based reviews. We also found that the sentiment of each review is very obvious, the review rating provided by the user mirrors what the user writes as his/her review, i.e. if the user writes something bad definitely the overall rating the user gives is either 1 or 2 out of 5. This is from our study of a set of amazon reviews on iPhone 5. Our work mainly concentrates on feature extraction and finding out the sentiment of the particular feature. We have used POS tagging technique on sentence level. In our approach we have made certain rules using the tags of particular word and using list of words with respective sentiment value to find the feature and then getting the appropriate sentiment from it. The Sentiment model that we have proposed is designed based on the uncertainty of the amazon reviews. Our work also include summarization in the form of charts for overall view of the sentiments of the users on the product or a particular feature.

* 1. **BLOCK DIAGRAM**

# THEORITICAL ANALYSIS



### HARDWARE / SOFTWARE DESIGNING

The following is the Hardware required to complete this project:

* + - Internet connection to download and activate
    - Administration access to install and run Anaconda Navigator
    - Minimum 10GB free disk space
    - Windows 8.1 or 10 (64-bit or 32-bit version) OR Cloud: Get started free, \*Cloud account required.

Minimum System Requirements To run Office Excel 2013, your computer needs to meet the following minimum hardware requirements:

* + - 500 megahertz (MHz)
    - 256 megabytes (MB) RAM
    - 1.5 gigabytes (GB) available space
    - 1024x768 or higher resolution monitor

The following are the software required for the project:

* + - Google Colaboratory Notebook and Jupyter Notebook
    - Spyder and Pycharm Community
    - Microsoft Excel 2013

# EXPERIMENTAL INVESTIGATION

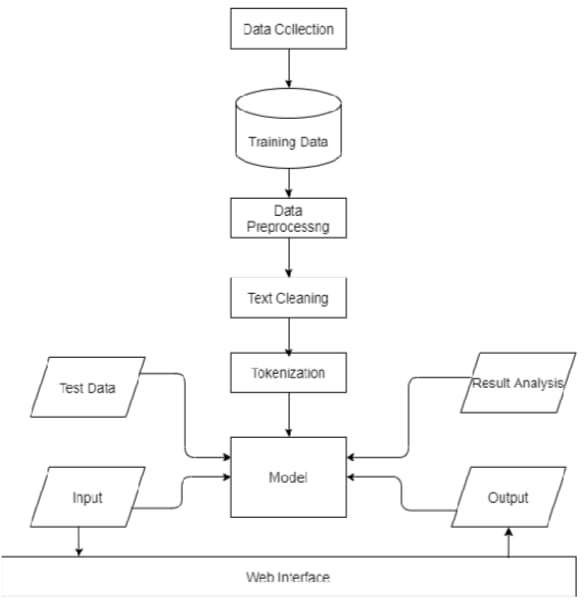
In this project, we have used Amazon Kindle Store Reviews Dataset.This dataset is a csv file consisting of labelled data and having the following coloumns-

“reviewerID”: D of the reviewer “asin”: ID of the product “reviewerName”: name of the reviewer

“helpful”: helpfulness rating of the review “reviewText”: text of the review “overall”: rating of the product “summary”: summary of the review “reviewTime”: time of the review “unixreviewTime”: unix timestamp

For the dataset we selected, it consists of more than 50,000 kindle book reviews. From the format used analysing the review polarity we used review test & Overall from it.

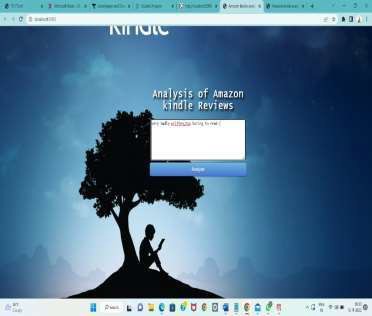
# FLOWCHART



1. **RESULT**



## HOMEPAGE

## REVIEW ANALYSIS PAGE OUTPUT



**REVIEW ANALYSIS**



**OUTPUT**

**ADVANTAGES**

# ADVANTAGES AND DISADVANTAGES

So many books to choose from Free books

Access to libraries online collections Cheaper books

Internet, music, and games Dictionary

Translations Electronic markers

No book light required Large print

Long battery life Search function Paperless Convenience

### DISADVANTAGES

It’s harder to share No color

Eye strain and retention Its electronic

# APPLICATIONS

1. Kindle unlimited and amazon prime numbers can select and download kindle books directly in the app
2. choose from over six million kindle books 3.understand challenging books

4.Improve your reading comprehension

# CONCLUSION

It is completely impossible to use only raw text as input for making predictions. Hence,we saw that the preprocessing step played a major role in the complete process of NLP. To get better results, accuracy and make the machine take all the text as tokens, pre-processing of data is to be done carefully looking at the type of contents present in it.The most important thing is to be able to extract the relevant features from the given source of data. This kind of data can often come as a good complementary source in order to extract more learning features and increase the predictive power of the models. And the user is able to predict that the given comment is positive or negative.

# FUTURE SCOPE

In future, the work can be extended to perform multi-class classification of reviews which will provide a delineated nature of review to the consumer, hence better judgment of the product. It can also be used to predict the rating of a product from the review. This will provide users with a reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. The proposed extension of work will be very beneficial for the e-commerce industry as it will augment user satisfaction and trust.

## BIBILOGRAPHY

1. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12(Aug):2493–2537, 2011.
2. M. S. Elli and Y.-F. Wang. Amazon reviews, business analytics with sentiment analysis.
3. C. Rain. Sentiment analysis in amazon reviews using probabilistic machine learning. Swarthmore College, 2013.
4. R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing, pages 1631–1642, 2013.
5. Y. Xu, X. Wu, and Q. Wang. Sentiment analysis of yelps ratings based on text reviews, 2015.
6. Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107- 5110.
7. Chen, Weikang, Chihhung Lin, and YiShu Tai."Text-Based Rating Predictions on Amazon Health & Personal Care Product Review." (2015)
8. Shaikh, Tahura, and DeepaDeshpande. "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews.",(2016)
9. Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER.,
10. Text mining for yelp dataset challenge; Mingshan Wang; University of California San Diego, (2017)

# APPENDIX

Model Building 1.Dataset

2.Jupiter Notebook Application Building

1. HTML file
2. CSS file 3.Flask

4.IBM Watson.

# SOURCE CODE:

HTML Code:

<!DOCTYPE html>

<html lang="en">

<head>

<meta charset="utf-8">

<title>

beautiful and simple website using html only -- fullywrold web tutorials

</title>

</head>

<body background="1.jpg" link="#000" alink="#017bf5" vlink="#000">

<br />

<h3 align="center">

<font face="Lato" size="6">LOGO</font>

&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbs p;&nbsp;&nbsp;&nbsp;&nbsp;

&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbs p;&nbsp;&nbsp;&nbsp;&nbsp;

&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbs p;&nbsp;&nbsp;&nbsp;&nbsp;

&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbs p;&nbsp;&nbsp;&nbsp;&nbsp;

<font face="cinzel" size="4">

<a href="#">HOME</a>&nbsp;&nbsp;&nbsp;&nbsp;

<a href="#">VIDEOS</a>&nbsp;&nbsp;&nbsp;&nbsp;

<a href="#">PORTFOLIO</a>&nbsp;&nbsp;&nbsp;&nbsp;

<a href="#">BLOG</a>&nbsp;&nbsp;&nbsp;&nbsp;

<a href="#">CONTACT US</a>

</font>

</h3>

<br /><br /><br /><br /><br /><br /><br /><br /><br /><br />

<h1 align="center">

<font face="Lato" color="#017bf5" size="7"> BEAUTIFUL AND SIMPLE WEB PAGE

</font>

</h1>

<h3 align="center">

<font face="Lato" color="#000" size="5"> USING HTML ONLY (NO CSS USED)

</font>

</h3>

<br />

<h3 align="center">

<a href="#">

<font face="Lato" color="#000">GET STARTED</font>

</a>&nbsp;&nbsp;&nbsp;&nbsp;

<a href="#">

<font face="Lato" color="#fff">SUBSCRIBE US</font>

</a>

</h3>

</body>

</html>

FLASK PROGRAM

import numpy as np

from flask import Flask, request, render\_template from joblib import load

import joblib

from tensorflow.keras.models import load\_model

from sklearn.feature\_extraction.text import CountVectorizer import tensorflow as tf

from tensorflow import keras

from tensorflow.keras import models from tensorflow.keras import backend

from gevent.pywsgi import WSGIServer import os tf.keras.backend.clear\_session()

app = Flask(\_name\_) model=tf.keras.models.load\_model("amazo.h5") @app.route('/')

def home():

return render\_template('index.html') @app.route('/y\_predict',methods=['POST']) def y\_predict():

'''

For rendering results on HTML GUI '''

d = request.form['Sentence'] print(d)

loaded=CountVectorizer(decode\_error='replace',vocabulary=joblib.load("amazo.save")) d=d.split("delimiter")

result=model.predict(loaded.transform(d)) print(result)

prediction=result>0.5

if prediction[0] == False: output="Positive review"

elif prediction[0] == True: output="Negative review"

return render\_template('index.html', prediction\_text='{}'.format(output)) port = os.getenv('VCAP\_APP\_PORT','5000')

if \_name\_ == "\_main\_":

app.secret\_key = os.urandom(12)

app.run(debug=True,host='0.0.0.0',port = port)

# CODE SNIPPETS

**A.MODELBUILDING**

