

Report On

# **AMAZON PRODUCT REVIEWS SENTIMENT ANALYSIS**

Submitted in partial fulfilment of the requirements of the Course project in  
Semester VII of Final Year Computer Science Engineering (Data Science)

by  
Sahil Gujral (Roll No:15)  
Jidnyasa Patil (Roll No:43)  
Niyati Patil (Roll No:44)

Supervisor  
**Dr. Tatwadarshi P. N.**



**University of Mumbai**

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Science and Engineering (Data Science)**



**(2023-24)**

**Vidyavardhini's College of Engineering & Technology**  
**Department of Computer Science and Engineering (Data Science)**

**CERTIFICATE**

This is to certify that the project entitled “Amazon Product Review System for Sentiment Analysis ” is a bonafide work of "**Sahil Gujral (Roll No.15), Jidnyasa Patil(Roll No.43), Niyati Patil(Roll No.44)**” submitted to the University of Mumbai in partial fulfilment of the requirement for the Course        project in semester VII of Final Year Computer Science Engineering (Data Science).

**Supervisor**

**Dr. Tatwadarshi P. N.**

Dr. Vikas Gupta.  
Head of Department

## Table of Contents

Chapter No	Title	Page No.
<b>1</b>	<b>Abstract</b>	1
<b>2</b>	<b>Introduction</b>	2
<b>3</b>	<b>Methodology</b>	3
<b>4</b>	<b>Technical Term Used</b>	5
<b>5</b>	<b>Results</b>	7
<b>6</b>	<b>Conclusions</b>	8

## **Abstract**

This NLP project is dedicated to the comprehensive examination of customer sentiments within the realm of Mobile Electronics products, as reflected in Amazon US reviews. Sentiment analysis, a key component of natural language processing, is the central focus of our investigation. We aim to delve into the emotional undercurrents hidden within textual customer feedback, providing invaluable insights into the world of consumer preferences and opinions.

To achieve this, we leverage the power of DistilBERT, a state-of-the-art transformer-based language model. DistilBERT enables us to unravel the complex emotional tapestry woven through the reviews of mobile electronics products on Amazon US, offering a nuanced understanding of customer sentiment.

The results of this analysis have far-reaching implications for both businesses and consumers. For businesses, it serves as a window into the minds of their customers, helping them glean important feedback on their products. This, in turn, allows businesses to make data-driven decisions, improve their offerings, and enhance customer satisfaction. For consumers, the analysis offers a trustworthy guide, aiding them in making informed choices when browsing through the vast array of mobile electronics products available on Amazon US.

In essence, this project represents a bridge between the voice of the customers and the world of commerce, facilitating a better understanding of consumer feedback and enabling smarter, more informed decisions based on the wealth of sentiment expressed in these reviews.

## Introduction

In an era where e-commerce reigns supreme, customer reviews have become a critical source of insight, influencing both consumer choices and business strategies. This project embarks on a journey into the heart of Amazon's vast ecosystem, specifically targeting the domain of Mobile Electronics products, to unravel the sentiments concealed within the intricate web of customer reviews. Sentiment analysis, a potent tool in the realm of Natural Language Processing (NLP), forms the cornerstone of our investigation.

The primary objective of this project is to discern and quantify the emotional tone woven into the textual tapestry of customer feedback on Amazon US. By doing so, we aspire to empower both businesses and consumers with a profound understanding of customer sentiment, as it relates to an extensive array of mobile electronics products.

Our choice of the DistilBERT language model represents a deliberate nod to the forefront of NLP technology. DistilBERT, a distilled version of the renowned BERT (Bidirectional Encoder Representations from Transformers) model, boasts the ability to comprehend the subtle nuances of human language. This powerful transformer-based model equips us with the means to dissect customer reviews with precision, revealing the sentiments that underlie them.

The significance of this project extends far beyond mere academic or technical curiosity. For businesses, it offers a window into the collective mind of their customers. The insights gleaned from this analysis can drive data-driven decisions, product improvements, and ultimately, enhanced customer satisfaction. On the other hand, for consumers navigating the labyrinth of choices in mobile electronics, this analysis stands as a trustworthy guide. It empowers them to make informed decisions by considering the wealth of sentiment expressed in these reviews.

In summary, this project serves as a conduit between the voice of the customers and the world of commerce. It seeks to foster a deeper understanding of customer feedback, providing the necessary tools for both businesses and consumers to make smarter, more informed decisions. Through the lens of sentiment analysis, we embark on a journey to uncover the emotions that shape the landscape of Mobile Electronics products on Amazon US.

## Methodology

### Data Collection:

In the first phase of this project, we will amass a comprehensive dataset of Amazon US reviews specifically related to Mobile Electronics products. This dataset is the lifeblood of our sentiment analysis endeavor, and it will encompass a range of attributes essential for our analysis. These attributes include product ratings, the textual body of the reviews, and other pertinent metadata, such as the reviewer's information, product details, and review date. This holistic approach to data collection ensures that our analysis is grounded in a rich, real-world context, providing a wealth of information to uncover and explore customer sentiments.

### Data Preprocessing:

The collected data is likely to be raw and unstructured, and thus, it will undergo a series of preprocessing steps to prepare it for sentiment analysis. This data preparation is a crucial step in ensuring the accuracy and reliability of our sentiment analysis model. Text cleaning will involve tasks like removing HTML tags, special characters, and any irrelevant information. Tokenization breaks down the reviews into individual words or tokens, which is necessary for the model to understand the text. Furthermore, we'll eliminate stop words, which are commonly used but don't carry significant meaning, to reduce noise in the data. Any missing or erroneous data points will be handled through imputation or data cleaning techniques to maintain the quality and integrity of the dataset.

### DistilBERT Model:

For the core of our sentiment analysis, we've chosen the DistilBERT language model. DistilBERT, a lightweight variant of BERT (Bidirectional Encoder Representations from Transformers), offers comparable performance to BERT while being computationally more efficient. Its efficiency makes it a suitable choice for processing a large corpus of textual data while delivering high-quality results. The DistilBERT model is capable of understanding the complex relationships and nuances within the language, which is essential for sentiment analysis.

### Fine-Tuning:

To tailor the DistilBERT model for our specific sentiment analysis task, we will engage in a process called fine-tuning. Fine-tuning entails adjusting the model's pre-trained weights using our labeled sentiment data. By exposing the model to our dataset, it learns to recognize patterns and characteristics that are indicative of sentiment. This optimization process ensures that the DistilBERT model becomes highly specialized in identifying sentiment in the context of Mobile Electronics product reviews.

### Training and Evaluation:

After fine-tuning, we will divide our dataset into separate training and testing sets. The training set will be used to teach the fine-tuned model to recognize sentiment, while the testing set will be used to evaluate its performance. During training, the model will learn how to classify reviews as positive or negative based on the labeled sentiment data. The

model's performance will be assessed using various evaluation metrics, such as accuracy, precision, recall, and F1-score. This evaluation stage ensures that our model is effective in its sentiment classification task and can be relied upon for real-world applications.

**Sentiment Analysis:**

With a successfully trained and evaluated DistilBERT model, we will be ready to conduct sentiment analysis on new, unseen reviews for Mobile Electronics products. The model will predict the sentiment of these reviews, classifying them as either positive or negative. This final step enables us to extract valuable insights from customer feedback, providing a nuanced understanding of how consumers perceive and evaluate Mobile Electronics products on Amazon US. These insights, in turn, can inform business decisions and guide consumers in making informed choices based on the sentiments expressed in these reviews.

## Technical Terms Used

### **Sentiment Analysis:**

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves assessing and determining the sentiment or emotional tone conveyed in a piece of text, such as a review, comment, or social media post. The primary objective is to classify the text as either positive, negative, or sometimes neutral. It is a valuable tool for understanding public opinion, customer feedback, and user sentiments, making it relevant for various applications, including market research, brand management, and customer service. Sentiment analysis helps organizations gauge the overall sentiment surrounding their products, services, or brand and make data-driven decisions based on customer feedback.

### **Transformer:**

The transformer is a deep learning architecture that has had a profound impact on the field of natural language processing (NLP). It revolutionized the way sequential data, like text, is processed efficiently. One of the key innovations in transformers is the self-attention mechanism, which allows the model to weigh the importance of different words or tokens in a sequence when making predictions. This mechanism enables transformers to capture long-range dependencies in the data, making them highly effective for tasks like machine translation, text generation, and sentiment analysis. Transformers have become the foundation for many state-of-the-art NLP models, including BERT, GPT, and DistilBERT, and have significantly advanced the capabilities of language understanding and generation tasks.

### **Hugging Face:**

Hugging Face is a well-known and widely respected organization in the field of NLP. It has gained recognition for its contributions to the development and maintenance of a repository of pre-trained language models. Hugging Face's efforts have made it easier for researchers, developers, and data scientists to access, fine-tune, and utilize these models for a variety of NLP tasks. Their platform provides a user-friendly and open-source environment for experimenting with and deploying NLP models. This accessibility has democratized the use of advanced NLP models, making them accessible to a broader audience and accelerating progress in the field of NLP.

### **DistilBERT:**

DistilBERT is a specific instance of a transformer-based language model. It is a "distilled" version of the BERT (Bidirectional Encoder Representations from Transformers) model, which means it has been optimized for efficiency while retaining competitive performance in NLP tasks. DistilBERT achieves this efficiency by reducing the model's size and computational demands, making it suitable for use in scenarios where computational resources are limited. Despite its smaller size, DistilBERT is still capable of understanding the contextual relationships in language and can be fine-tuned for various NLP tasks, including sentiment analysis, text classification, and more. This



balance of efficiency and effectiveness makes it a popular choice for many NLP applications.

**Tokenization:**

Tokenization is a fundamental text processing technique that involves breaking down a sequence of text into smaller units called tokens. Tokens can be words, subwords, or even individual characters, depending on the chosen tokenization strategy. The primary goal of tokenization is to convert raw text into a format that can be processed by language models and other NLP algorithms. By splitting text into tokens, a language model can analyze and understand the structure and meaning of the text more efficiently. Tokenization plays a crucial role in many NLP tasks, as it provides the model with the basic building blocks to learn patterns and relationships within the text data, enabling it to make predictions or classifications.

# Implementation and Results

The image displays two screenshots of a Jupyter Notebook titled "sentiment-analysis-python.ipynb" in a web browser. The browser tabs include "Index Page", "Inbox (2,891) - sahil.203158101...", "1.NLP - Google Docs", "sentiment-analysis-python.ipynb", and "ChatGPT". The notebook's address bar shows a Google Drive link: "colab.research.google.com/drive/16QGRa-dAdjDSOGD9g59H2SMF10hS6wL?usp=sharing#scrollTo=BV7QRUfIdudx".

The first screenshot shows the notebook's title "Sentiment Analysis in Python" and a brief description: "sentiment analysis in python using two different techniques:". It lists three methods: 1. VADER (Valence Aware Dictionary and sEntiment Reasoner) - Bag of words approach, 2. Roberta Pretrained Model from Huggingface, and 3. Huggingface Pipeline. Below this, a section titled "Step 0. Read in Data and NLTK Basics" contains the following code:

```
[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import nltk

plt.style.use('ggplot')

[ ] # Read in data
df = pd.read_csv('/content/Reviews.csv')

print(df.shape)
df = df.head(500)
```

The second screenshot shows a section titled "Extra: The Transformers Pipeline" with the sub-header "Quick & easy way to run sentiment predictions". It contains the following code:

```
[ ] from transformers import pipeline

sent_pipeline = pipeline("sentiment-analysis")

No model was supplied, defaulted to distilbert-base-uncased-finetuned-sst-2-english and revision afef99b (https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english).
Using a pipeline without specifying a model name and revision in production is not recommended.
Downloading (...)ve/main/config.json: 100% 629/629 [00:00<00:00, 30.2kB/s]
Downloading model safetensors: 100% 268M/268M [00:02<00:00, 114MB/s]
Downloading (...)okenizer_config.json: 100% 48.0/48.0 [00:00<00:00, 2.43kB/s]
Downloading (...)solve/main/vocab.txt: 100% 232k/232k [00:00<00:00, 2.68MB/s]

[ ] sent_pipeline('I love sentiment analysis!')

[{'label': 'POSITIVE', 'score': 0.9997853636741638}]

[ ] sent_pipeline('I do not love sentiment analysis!')

[{'label': 'NEGATIVE', 'score': 0.997802197933197}]
```

## Conclusion

In the ever-evolving landscape of e-commerce, customer reviews on platforms like Amazon serve as a vital source of information for both consumers and businesses. This mini-project delves into the realm of sentiment analysis, focusing on Amazon US reviews for Mobile Electronics products, with the powerful DistilBERT language model as our guiding tool. Our journey through this project has unveiled several key insights and implications.

Through the lens of sentiment analysis, we have harnessed the capabilities of DistilBERT to peel back the layers of customer reviews, deciphering the emotional undercurrents concealed within the textual fabric. This process has allowed us to classify reviews as positive or negative, shedding light on the collective sentiment of customers towards Mobile Electronics products available on Amazon US.

The impact of this analysis is two-fold. For businesses, it opens a window into the hearts and minds of their customers, providing invaluable feedback that can be used to improve product offerings, refine marketing strategies, and enhance overall customer satisfaction. Businesses can now make data-driven decisions, thereby strengthening their competitive edge in the marketplace.

On the other hand, consumers also stand to benefit significantly. In a world inundated with choices, the sentiment analysis conducted in this project serves as a trustworthy guide, aiding consumers in making informed decisions when shopping for Mobile Electronics products on Amazon US. By considering the sentiments expressed in these reviews, consumers can navigate the vast array of choices with greater confidence, ensuring that their purchases align with their expectations and preferences.

In essence, this mini-project stands as a bridge between customer voices and the world of commerce. By unlocking the emotional content within Amazon US reviews, we have empowered businesses and consumers alike to gain deeper insights into the world of Mobile Electronics products. As a result, this analysis equips both stakeholders with the knowledge needed to make more informed decisions, ultimately enhancing the quality of products, services, and the overall shopping experience.

As the e-commerce landscape continues to evolve, projects like this demonstrate the power of NLP and sentiment analysis, showing how they can reshape the way we understand and engage with customer feedback in a digital age.