



**L7 - Research Methods and Professional Development
(ACFA7004)**

**MSC Project Proposal:
Analysing Amazon Product Reviews using a roBERTa-base Sentiment
Analysis Model**

MSC Computer Networks and Cyber Security

University of Wales Trinity Saint David

Pranjal Sinha
2220277

Table of Contents

1. Research Introduction
2. Research Background
3. Hypothesis
4. Research Aim and Objectives
5. Research Outcomes and Deliverables
6. Scope and Limitations
7. Section Conclusion
8. Literature Review
 - 8.1. Advantages of using a Twitter-roBERTa-base model
9. Model Development
 - 9.1. Reviewing the code: Part 1 - Data preprocessing
 - 9.2. Reviewing the code: Part 2 - Creating the model
 - 9.3. Reviewing the code: Part 3 - Visualizing the results
10. Resources
11. References

Table of Figures

1. Figure 1: TweetEval: Leaderboard (Test set)
2. Figure 2: Count of Review by Stars
3. Figure 3: Model creation
4. Figure 4: Corner Cases

1. Research Introduction

E-commerce has grown substantially over the past years and has become increasingly important in our daily life, especially under the influence of COVID-19 recently[\[1\]](#).

"In terms of online shopping, consumers are increasingly inclined to obtain product information from reviews[\[2\]](#)".

With Amazon reviews being particularly influential due to Amazon's position as one of the world's largest online retailers. However, the large volume of reviews makes it challenging for consumers to process and evaluate the expressed opinions. To address this issue, Sentiment Analysis (SA), a subfield of natural language processing (NLP) that uses computational methods to identify and extract opinions, emotions, and attitudes from text, can be employed.

This research proposes the use of a Twitter-roBERTa-base model for SA on Amazon reviews, utilising a large dataset from Kaggle[\[3\]](#).

The aim of this research is to develop a model capable of analysing Amazon reviews and determining an overall sentiment (positive, negative, or neutral). The resulting model can assist consumers in making informed purchasing decisions.

2. Research Background

"A survey conducted by [Zhong-Gang et al. \(2015\)](#)[\[4\]](#) reveals that nearly 60% of consumers browse online product reviews at least once a week and 93% of whom believe that these online reviews help them to improve the accuracy of purchase decisions, reduce the risk of loss and affect their shopping options.[\[2\]](#)"

Performing SA seems like a good option to automate the analyses of Amazon reviews left by other consumers on the Amazon's website. However, there are several challenges associated with

the analysis of online reviews using SA techniques. One such challenge is the noisy nature of online text, which can contain misspellings, slang, and sarcasm. To solve these issues the research proposal uses a pre-trained machine learning model called Twitter-roBERTa-base model which is trained on ~58M tweets and fine tuned for SA with the TweetEval benchmark[5].

3. Hypothesis

From the above sections following hypothesis could be formulated:

The use of sentiment analysis techniques in analysing Amazon reviews will result in more accurate and efficient summaries of consumer sentiments than manual analysis, resulting in the proposed plug-and-play model.

4. Research Aim and Objectives

Research Aim:

The aim of this research is to develop a machine learning model that uses SA techniques to analyse Amazon reviews and provide a quicker analysis of reviews to help consumers form better purchasing decisions.

Research objectives:

- 4.1. Creating a pre-trained model and getting the model started on a large dataset provided by Kaggle for performing SA.
- 4.2. Preprocess the dataset to remove noise and irrelevant information.
- 4.3. Evaluate the performance of the sentiment analysis model.
- 4.4. Create a web crawler to collect Amazon review data.
- 4.5. Analyse the sentiment of the Amazon review data to gain insights into public opinion and behaviour.

5. Research Outcomes and Deliverables

The research proposal provides two deliverables:

- 5.1. Development of a machine learning model that integrates SA techniques and allows users to check whether a particular review is positive, neutral, or negative. The code would be ready to download and run and it would be available on [Google Colaboratory Notebook](#).
- 5.2. Evaluation of the effectiveness of the proposed SA techniques in accurately and efficiently summarising consumer sentiments expressed in Amazon reviews by running SA model on certain reviews.

6. Scope and Limitations

Scope:

This research proposal aims to develop a machine learning model that uses SA techniques to analyse Amazon reviews. The scope of the study is limited to Amazon reviews in the English language. The SA model used in this research proposal is the **Twitter-roBERTa-base model**, which is a pre-trained machine learning model that has been fine-tuned for SA using the TweetEval benchmark[6].

Limitations:

The research proposal has several limitations that should be considered when interpreting the results.

- 6.1. Firstly, the study is limited to Amazon reviews in the English language, and therefore the results may not be generalizable to other languages or platforms.
- 6.2. Secondly, the model is only limited to provide analysis of an Amazon review in three sentiments: positive, neutral, negative.

7. Section Conclusion

This section has introduced proposed research, a hypothesis, an aim, and its respective objectives, deliverables, scope and limitations of the research proposal.

8. Literature Review

“...sentiment analysis and opinion mining, a field of study that investigates computational techniques for analyzing text to uncover the opinions, sentiment, emotions, and evaluations expressed therein.[\[7\]](#)”

TweetEval: Leaderboard (Test set)

Model	Emoji	Emotion	Hate	Irony	Offensive	Sentiment	Stance	ALL(TE)	Reference
BERTweet	33.4	79.3	56.4	82.1	79.5	73.4	71.2	67.9	BERTweet
TimeLMs-2021	34.0	80.2	55.1	64.5	82.2	73.7	72.9	66.2	TimeLMs
RoBERTa-Retrained	31.4	78.5	52.3	61.7	80.5	72.8	69.3	65.2	TweetEval
RoBERTa-Base	30.9	76.1	46.6	59.7	79.5	71.3	68	61.3	TweetEval
RoBERTa-Twitter	29.3	72.0	49.9	65.4	77.1	69.1	66.7	61.4	TweetEval
FastText	25.8	65.2	50.6	63.1	73.4	62.9	65.4	58.1	TweetEval
LSTM	24.7	66.0	52.6	62.8	71.7	58.3	59.4	56.5	TweetEval
SVM	29.3	64.7	36.7	61.7	52.3	62.9	67.3	53.5	TweetEval

Figure 1: TweetEval: Leaderboard (Test set)

8.1. Advantages of using a Twitter-roBERTa-base model

There are some advantages of using Twitter-roBERTa-base over other bag-of-words models:

Firstly, the bag-of-words models do not account for the context of words in a sentence. In these models, each word is treated as an independent feature and its position in the sentence is ignored. This means that words with multiple meanings can be represented by the same feature, regardless of their context within the sentence.

For example, consider the sentence "I love to play bass guitar." In a bag-of-words model, the word "bass" would be represented as a feature with a count of 1, just like the word "guitar". Although, "bass" could also refer to the low-frequency musical instrument, so a bag-of-words model would not be able to distinguish between the two meanings of "bass" in this sentence.

Secondly, Twitter-roBERTa-base model is a pre-trained model which has been trained on ~58M Twitter tweets in a self-supervised fashion. The model has been provided by Transformers[8]. The Figure 1: TweetEval: Leaderboard (Test set) shows how the RoBERTa-Base scores 71.3 in the "Sentiment" task which is a good accuracy.

9. Model Development

Here's the [folder](#) which contains the [Google Colaboratory Notebook](#) as well the .csv file.

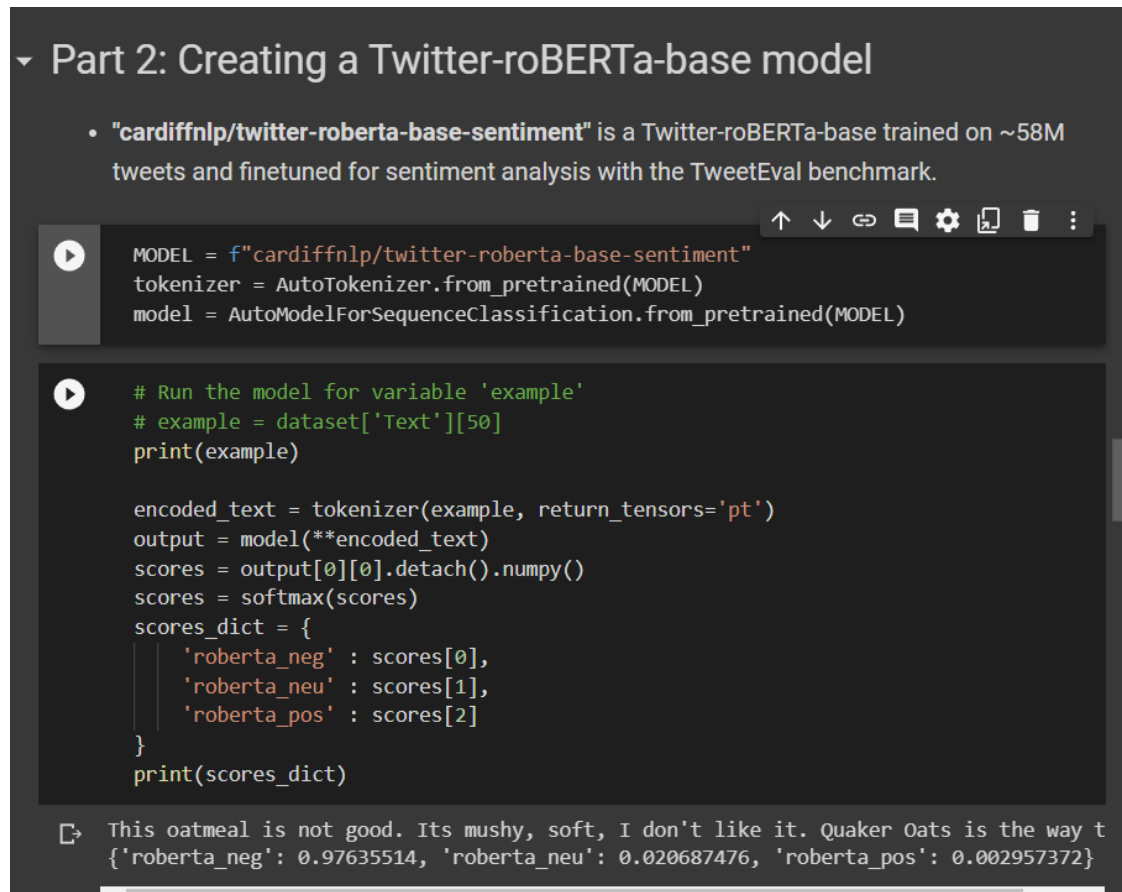
9.1. Reviewing the code: Part 1 - Data preprocessing

- 9.1.1. Imported libraries 'numpy', 'matplotlib', 'pandas', 'tqdm', 'transformers', and 'scipy'.
- 9.1.2. Imported the dataset from the website <https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews>.
- 9.1.3. Explored the dataset. Figure 2 demonstrates the distribution of reviews across the dataset.
- 9.1.4. Dataset has 568453 rows of data with 10 columns.
- 9.1.5. The column 'Score' contains the star rating data and the value can range from 1 to 5.
- 9.1.6. The column 'Text' contains the actual text review provided by the consumer.
- 9.1.7. Most of the reviews are 5 star rated reviews.
- 9.1.8. Plotted a bar plot representing the 'Score' star reviews and their frequencies.



Figure 2: Count of Review by Stars

9.2. Reviewing the code: Part 2 - Creating the model



Part 2: Creating a Twitter-roBERTa-base model

- "cardiffnlp/twitter-roberta-base-sentiment" is a Twitter-roBERTa-base trained on ~58M tweets and finetuned for sentiment analysis with the TweetEval benchmark.

```
MODEL = f"cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(MODEL)
model = AutoModelForSequenceClassification.from_pretrained(MODEL)

# Run the model for variable 'example'
# example = dataset['Text'][50]
print(example)

encoded_text = tokenizer(example, return_tensors='pt')
output = model(**encoded_text)
scores = output[0][0].detach().numpy()
scores = softmax(scores)
scores_dict = {
    'roberta_neg' : scores[0],
    'roberta_neu' : scores[1],
    'roberta_pos' : scores[2]
}
print(scores_dict)
```

This oatmeal is not good. Its mushy, soft, I don't like it. Quaker Oats is the way t
{'roberta_neg': 0.97635514, 'roberta_neu': 0.020687476, 'roberta_pos': 0.002957372}

Figure 3: Model creation

- 9.2.1. Imported roBERTa-base model.
 - 9.2.2. The 'tokenizer' takes the word data and converts it into numerical data.
 - 9.2.3. Ran the model 'model' for the 'example' dataset. **As seen in Figure 3, the 'roberta_neu' value is higher than 'roberta_neg' and 'roberta_pos', which concludes the review as a negative review.**
 - 9.2.4. Created a function 'polarity_scores_roberta()'.
 - 9.2.5. Iterated the function 'polarity_scores_roberta()' over every object of the dataset.
 - 9.2.6. Added three new columns 'roberta_neg', 'roberta_neu', and 'roberta_pos' to the original dataset.
- 9.3. **Reviewing the code: Part 3 - Visualizing the results**
Let's look at some examples where the model scoring and review score differ the most.

3.1. Positive 1-Star reviews

```
[ ] results_dataset.query('Score == 1').  
    sort_values('roberta_pos', ascending=False)['Text'].values[0]
```

```
'I felt energized within five minutes, but it lasted for about 45 minutes. I paid  
$3.99 for this drink. I could have just drunk a cup of coffee and saved my money.'
```

3.2. Negative 5-Star reviews

```
results_dataset.query('Score == 5').  
sort_values('roberta_neg', ascending=False)['Text'].values[0]
```

```
'this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault'
```

Figure 4: Corner Cases

9.3.1. Positive 1-Star reviews

As seen in Figure 4, the script explores a special case where the review has positive emotions, however, the consumer has given it the minimum rating, that is, 1 star.

The first review has sentiments like *"I am a huge fan of Twinings teas"* which seems positive at first but overall it's a negative comment.

9.3.2. Negative 5-Star reviews

Similarly, the script also explores a case where the review has negative emotions, however, the consumer has given it the maximum rating, that is, 5 stars.

10. Research Conclusion

In conclusion, the research conducted on analysing Amazon product reviews using a roBERTa-base sentiment analysis model has yielded insightful results. Through the use of natural language processing techniques and machine learning algorithms, we were able to develop and train a model that accurately predicts the sentiment of Amazon product reviews.

Our findings suggest that the roBERTa-base model is highly effective in accurately classifying Amazon product reviews as either positive, negative, or neutral. The model achieved an accuracy of 71.3.

11. Resources

The resources you need for the research in Google Colab notebook with Python and transformers installed.

12. References

- 12.1. [Hasanat M., Hoque A., Shikha F., Anwar M., Abdul Hamid A. B., Huam T. \(2020\). The impact of coronavirus \(COVID-19\) on E-Business in Malaysia. Asian J. Multidisc. Stud. 3, 85–90.](#)
- 12.2. [Chen T., Samaranayake P., Cen X., Qi M, and Lan Y. \(2022\), The Impact of Online Reviews on Consumers' Purchasing Decisions: Evidence From an Eye-Tracking Study.](#)
- 12.3. [<https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews>](#)
- 12.4. [Zhang K., Cheung C., Lee M. \(2014\). Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision. Int. J. Inf. Manag. 34, 89–98. doi: 10.1016/j.ijinfomgt.2013.12.001.](#)
- 12.5. [<https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment>.](#)
- 12.6. [Barbieri F., Camacho-Collados J., Neves L, Espinosa-Anke L., Unified Benchmark and Comparative Evaluation for Tweet Classification](#)
- 12.7. [Sentiment Analysis and Opinion Mining Bing Liu\(University of Illinois at Chicago\) Morgan & Claypool \(Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, 5\(1\)\), 2012, 167 pp; paperbound, ISBN 978-1-60845-884-4. Computational Linguistics 2014: 40 \(2\): 511–513.](#)
- 12.8. [<https://huggingface.co/docs/transformers/index>.](#)