

**National University of Computer and Emerging Sciences, Karachi**

**SENTIMENT ANALYSIS ON MOBILE PHONE REVIEWS USING BERT AND XLNET**

**Research Proposal**

**Prepared for**

**Ms. Atifa Batool**

**Instructor Technical and Business Writing**

**FAST-NUCES**

**By**

**19K-1517 Hermain Qadir**

**19K-0354 Saman Khan**

**19K-0281 Anusha Saad**

**BSCS – Batch 2019**

**FAST-NUCES**

**May 17, 2022**

# **Acknowledgements**

We would like to thank our supervisors: Mr. Waheed Ahmed, Mr. Ali Shah Fatmi, Mr. Nouman Durrani, and Ms. Atifa Batool, for assisting in formation of the research problem statement. We thank them for their guidance that greatly improved this manuscript.

We would also like to thank FAST National University of Computer and Emerging Sciences for instilling in us the best of knowledge and providing us with instructors who have motivated and guided us in the past three years.

We are immensely thankful to our parents and friends who have time and again given us feedback and comments to improvise the research problem. Any errors made by us should not tarnish the reputation of these esteemed individuals.

# **Abstract**

In the past decade, e-commerce has seen massive surge. As a result, online purchasing power has taken over traditional shopping methods and this has subsequently led to increase in online customer reviews. The same reviews have a substantial impact on potential buyers who determine worth of the product by its reviews, both complaints and recommendations. Our research proposes performing sentiment classification with newer and more accurate machine learning models, i.e. BERT and XLNET. The models, once trained, will predict sentiment of textual reviews, the emotion depicted by customer, and the aspect or main topic of the review. The accuracy of the model will be determined by techniques of F-score, recall, and precision.

# **Table of Contents**

|  |  |
| --- | --- |
| Acknowledgements……………………………………………………………………. | i |
| Abstract………………………………………………………………………………… | ii |
| I: Introduction………………………………………………………………………….. | 4 |
| Motivation……………………………………………………………………….. | 4 |
| Background……………………………………………………………………… | 4 |
| Purpose Statement……………………………………………………………….. | 4 |
| Problem Statement………………………………………………………………. | 4 |
| Scope and Limitations…………………………………………………………… | 4 |
| II: Literature Review…………………………………………………………………… | 5 |
| III: Methodology……………………………………………………………………….. | 6 |
| Data Collection…………………………………………………………………... | 6 |
| Data Cleaning and Preprocessing………………………………………………… | 6 |
| Feature Extraction………………………………………………………………... | 6 |
| Modelling………………………………………………………………………… | 6 |
| Evaluation……………………………………………………………………….. | 6 |
| Data Visualization………………………………………………………………. | 6 |
| IV: Conclusion………………………………………………………………………… | 7 |
| References……………………………………………………………………………... | viii |

# **I: Introduction**

## **Motivation**

Mobile phones and internet have become increasingly accessible and with the Covid-19 pandemic spanning to over two years, electronic commerce has overshadowed traditional methods of shopping for goods. The online reviews and feedback on e-commerce websites attract potential buyers and are possibly the make-or-break for any manufacturer. Each product has hundreds, or more, of reviews. To expect any manufacturer to go through each review, and conclude how customers feel about a product, is unreasonable due to the time and effort it would take up. This is why the textual reviews left by buyers on e-commerce websites are no longer disposed because the same reviews can be utilized in sentiment analysis, also called opinion mining, to generate immensely useful insights on customer preferences.

## **Background**

Conventionally, sentiment analysis has been thought of as technique limited to classifying reviews as positive or negative. Much of the work and research undertaken in the field has explored how Naïve Bayes and Decision Trees algorithms are used in sentiment analysis to determine polarity of reviews written by customers. The technique has been applied on an array of domains ranging from mobile phone reviews, restaurant reviews, to hotel reviews.

## **Purpose Statement**

The research intends to apply aspect-based Sentiment Analysis on a set of Amazon reviews to classify the feature(s) in a review and conclude a polarity for the same features which will guide manufacturers on how to improvise the product to meet customer requirements.

## **Problem Statement**

With word-of-mouth no longer being the primary source of a product’s reputation, online reviews impact the sales of a product more than what is perceived. With e-commerce taking over the shopping arena, there is a clear gap as most manufacturers have still not utilized the potential of aspect-based sentiment analysis. Sentiment analysis can assist manufacturer in gaining useful insights and drawing an action-plan on how to improve the product and increase the sales.

In order to bridge the same gap, this research aims to perform sentiment analysis and categorize reviews based on the feature(s) mentioned and deduce the overall average polarity of each feature.

## **Scope and Limitations**

This research involves performing aspect-based sentiment analysis on reviews of mobile phones collected from Amazon. Technique of TF-IDF, i.e. Term-Frequency-Inverse Data Frequency, will classify the features in each subjective and/or comparative review left by a buyer. The general sentiment and ratings will be concluded for various mobile phone brands. In addition to that, the most-talked about feature of each mobile phone brand will be determined. However, the research will only work with supervised learning algorithms of machine learning and not employ algorithms from the branch of reinforcement and unsupervised learning.

# **II: Literature Review**

Presently, e-commerce websites like Amazon, Flipkart, eBay etc. are changing the dynamics of traditional shopping. Consumers leave reviews for products bought which serve as primary data to apply sentiment analysis, or opinion mining. The information obtained from customer reviews assists business and ecommerce managers to improve quality of the products being sold or reviewed [1]. In the past, sentiment analysis has been performed on product reviews from two different perspectives: sentiment analysis using sentiment lexicon and sentiment analysis based on machine learning models [2].

Sentiment lexicon was first explored extensively by Taboada et al. [6] when dictionary of words on the basis of polarity, strength and emotion was used for text mining. Text labeled with emotion and integrated with a UMM (unigram mixture model) was proposed in [7] to classify emotions in textual phrases and sentences. Baseline lexicons evaluated a result of 955 confidence interval using a single-tail t-test. Hybrid of domain-specific emotion lexicon and general-purpose lexicon extract novel features along emotion concepts in textual documents [7]. Fang et al [8] built a bilayer model using emoticons and sentiment words to determine frequently used sentiment words. The results in [8] validate the mutual reinforcement ranking model. However, despite so much of extensive research undertaken, sentiment analysis based on sentiment lexicon is no longer a preferred technique due to high computational time and maintenance costs.

Machine Learning models, both supervised and unsupervised, have recently gained momentum to perform sentiment analysis because of little manual intervention required. In [1, 3], reviews were preprocessed by tokenization, followed by lemmatization using NLP toolkit. Following this, features were extracted by employing TF-IDF (Term Frequency-Inverse Document Frequency). [3] utilized the much-new GloVe simulation to separate features from textual reviews at an intricate level to achieve greater accuracy. Common supervised learning models like Naïve Bayes and Random Forest perform greatly for unary classification only. Bidirectional long-short memory and joint-learned embedding outperformed other models in binary classification [3]. K-medoid algorithm also achieved satisfactory results in varied polarity clustering [1].

One of the most researched and explored areas in unsupervised learning to perform sentiment analysis has been Space Vector Machine (SVM) and POS tagger. POS tagger removed noun and pronoun phrases as they have little to no effect on the sentiment displayed in reviews [4]. The max-entropy POS tagger, coupled with pruning, resolved up to 46 unique tags, becoming the most suitable option for role semantics [4]. On the other hand, SVM, in linear kernel and cubic kernel mode, was up to 60% accurate in one-vs-all (OVA) classification when used in performing sentiment analysis on Daraz reviews in Roman-Urdu text [5]. Cubic kernels fares better marginally in [5].

The various techniques like sentiment lexicon and machine learning models have been worked on data collected from a number of ecommerce websites. The two have often been merged with newer feature extraction methods to achieve greater accuracy and precision in confusion matrix reports. However, much of the work done has been limited to determining and amount, polarity, and type of reviews. No work has been to answer questions like “Does a review length affect product ratings?” or “Is there a relationship between price of a product and its review length?” Moreover, there has been only a little ground work done to apply deep learning models like Convolutional Neural Network (CNN) to perform sentiment analysis which stands a valid chance to prove better than sentiment lexicon and machine learning.

# **III: Methodology**

The BERT and XLNET models will be implemented on the data sets by Ktrain module which employs Convolutional Neural Network (CNN) as ktrain is a light-weight library to build, train, and deploy neural networks.

The proposed methodology to perform sentiment analysis consists of the following five steps

## **Data Collection**

Data collection will be performed through Kaggle as Kaggle has real-time datasets from Amazon.A dataset of mobile phone reviews from Amazon will be used in the following steps.

## **Data Cleaning and Preprocessing**

In this step, data will be thoroughly cleaned and prepared for null values, empty columns, or columns that do not contribute to deriving results. Followed by this, tokenization using the NLP library will be performed to obtain word tokens and embeddings.

## **Feature Extraction**

Features of the cleaned and preprocessed data will be extracted by TF-IDF in the shape of x\_train, y\_train, x\_test, y\_test tuples.

## **Modelling**

Once the data is split into test and train datasets, BERT Model and XLNET Model will be fine turned and then implemented on test and train data sets separately to obtain values of the target label.

## **Evaluation**

The accuracy of model will be checked by functions of sklearn metrics such as Precision, Accuracy, Recall, F-Score.

## **Data Visualization**

The results obtained will be plotted using seaborn and matplotlib libraries to display count plot, bar charts, histograms, strip-plots, and pie-charts for each brand separately.

# **IV: Conclusion**

Sentiment Analysis on Ecommerce product reviews are a major reason behind the changing trends of online shopping. Sentiment Analysis extracts meaningful insights from knowledge in reviews, ratings, emoticons. This knowledge plays a key role in assisting customer to decide whether the product is generally well-received or not. While other classification models have been implemented a lot, they do not achieve the desired accuracy, i.e. more than 90 per cent. On the other hand, BERT and XLNET are newer models that have been trained 25, 000 million words through Wikipedia, so assumptions are that they would obtain higher accuracy on validation or test data.

# 

# **References**

|  |  |
| --- | --- |
| [1] | U. Rahardja, T. Hariguna and W. M. Baihaqi, "Opinion Mining on E-Commerce Data Using Sentiment Analysis and K-Medoid Clustering," 2019 Twelfth International Conference on Ubi-Media Computing (Ubi-Media), 2019, pp. 168-170, doi: 10.1109/Ubi-Media.2019.00040. (19K-0281 Anusha Saad) |
| [2] | Yang, Li & Li, Ying & Wang, Jin & Sherratt, R.. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. IEEE Access. 8. 1-1. 10.1109/ACCESS.2020.2969854. (19K-1517 Hermain Qadir) |
| [3] | AlQahtani, Arwa S. M., Product Sentiment Analysis for Amazon Reviews (2021). International Journal of Computer Science & Information Technology (IJCSIT) Vol 13, No 3, June 2021, Available at SSRN: <https://ssrn.com/abstract=3886135> (19K-0354 Saman Khan) |
| [4] | Pankaj, P. Pandey, Muskan and N. Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 320-322, doi: 10.1109/COMITCon.2019.8862258. (19K-0354 Saman Khan) |
| [5] | Noor, Faiza & Bakhtyar, Maheen & Baber, Junaid. (2019). Sentiment Analysis in E-commerce Using SVM on Roman Urdu Text. 10.1007/978-3-030-23943-5\_16. (19K-281 Anusha Saad) |
| [6] | Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede; Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics 2011; 37 (2): 267–307. doi: <https://doi.org/10.1162/COLI_a_00049> (19K-1517 Hermain Qadir) |
| [7] | Bandhakavi, Anil Sriharsha & Wiratunga, Nirmalie & P, Deepak & Massie, Stewart. (2016). Lexicon based Feature Extraction for Emotion Text Classification. Pattern Recognition Letters. 93. 10.1016/j.patrec.2016.12.009 (19K-1517 Hermain Qadir) |
| [8] | Feng, S., Song, K., Wang, D. *et al.* A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs. *World Wide Web* 18**,**949–967 (2015). <https://doi.org/10.1007/s11280-014-0289-x> (19K-1517 Hermain Qadir) |