Predicting Review Helpfulness and Rating Using NLP and Machine Learning for Mobile Phone Reviews

[1] Nagalakshmi Manepalli, [2]Sattiraju Bhargavi, [3]Digumarthi Nireesha

[1] Assistant Professor, Dept of CSE, BVC College of Engineering, Palacharla

[2] Assistant Professor, Dept of CSE, ABN PRR College of Science, Kovvur

[3] Assistant Professor, Dept of CSE, Srinivasa Institute of Engineering and Technology, Cheyyeru

Abstract: To build a model to predict the helpfulness of the review and the rating based on the review text. Corpus based and knowledge-based methods can be used to determine the semantic similarity of review text. Natural language processing to analyze the sentiment (positive or a negative) of the given review. A sample web application is integrated to the model built. The consumer reviews serve as feedback for businesses in terms of performance, product quality, and consumer service. In this research, we predict consumer opinion based on mobile phone reviews, in addition to providing an analysis of the most important factors behind reviews being classified as either positive, negative, or neutral. This insight could help companies improve their products as well as helping potential buyers to make the right decision. The research presented in this paper was carried out as follows: the data was pre-processed, before being converted from text to vector representation using a range of feature extraction techniques such as bag-of words, TF-IDF, Glove, and word2vec. We study the performance of different machine learning algorithms, such as logistic regression, stochastic gradient descent, naive Bayes and convolutional neural networks. In addition, we evaluate our model using accuracy, F1-score, precision, recall and log loss function. Moreover, we apply Lime technique to provide analytical reasons for the reviews being classified as either positive, negative or neutral. Our experiments revealed that convolutional neural network with word2vec as a feature extraction technique provides the best results for both the unbalanced and balanced versions of the dataset.

Keywords: sentiment, convolutional, extraction, Corpus, word2vec, TF-IDF, Glove, F1-score, precision, research, experiments.

I. INTRODUCTION

Purchasing a product is an interaction between two entities, consumers and business owners. Consumers often use reviews to make decisions about what products to buy, while businesses, on the other hand, not only want to sell their products but also want to receive feedback in terms of consumer reviews. Consumer's reviews about purchased products shared on the internet have great impact. Human nature is generally structured to make decisions based on analyzing and getting the benefit of other consumer experience and opinions because others often have a great influence on our beliefs, behavior, perception of reality, and the choices we make. Hence, we ask others for their feedback whenever we are deciding on doing something.

Additionally, this fact applies not only to consumers but also to organizations and institutions. In the last few years, consumer ways of expressing their opinions and feelings have changed according to changes in social networks, virtual communities and other social media communities. Discovering large amounts of data from unstructured data on the web has become an important challenge due to its importance in different areas of life. To allow better information extraction from the plethora of data available sentiment analysis has emerged to be able to predict the polarity (positive, negative, neutral) of consumer opinion. This in turn would help consumers to better analyze the textual data providing useful information. We study in this research sentiment analysis of mobile phone reviews taken from the Amazon 1 website, and how these reviews help consumers to have confidence that they have made the right decision about their purchases. Also, the research in this work aims to help companies understand their consumers' feedback to maintain their products/services or enhance them. In addition, giving them insights about them in providing offers on specific products to increase their profits and customer satisfaction.

II. SYSTEM ANALASIS (EXPERIMENTAL INVESTIGATIONS)

Natural Language Processing (NLP): Natural language processing is a branch of Artificial Intelligence that helps the computers to understand, interpret and manipulate the human language. Example of NLP in action are virtual assistants, like Siri and Alexa. NLP understands and translate the human language to machine understandable language.

NLTK (Natural language Toolkit): This is one of the most important NLP library which contain packages to make machines understand human language and perform tasks like stemming, removing stop words etc.

Pre-processing: It refers to the transformations applied to our data before feeding it to the algorithm. The messages have to be pre-processed for the removal of unwanted punctuation, grammar, stop words etc.

Stemming: Stemming helps reduce a word to its stem form. It often makes sense to treat related words in the same way. It removes suffices, like "ing", "ly", "s", etc. by a simple rule-based approach. It reduces the corpus of words but often the actual words get neglected.
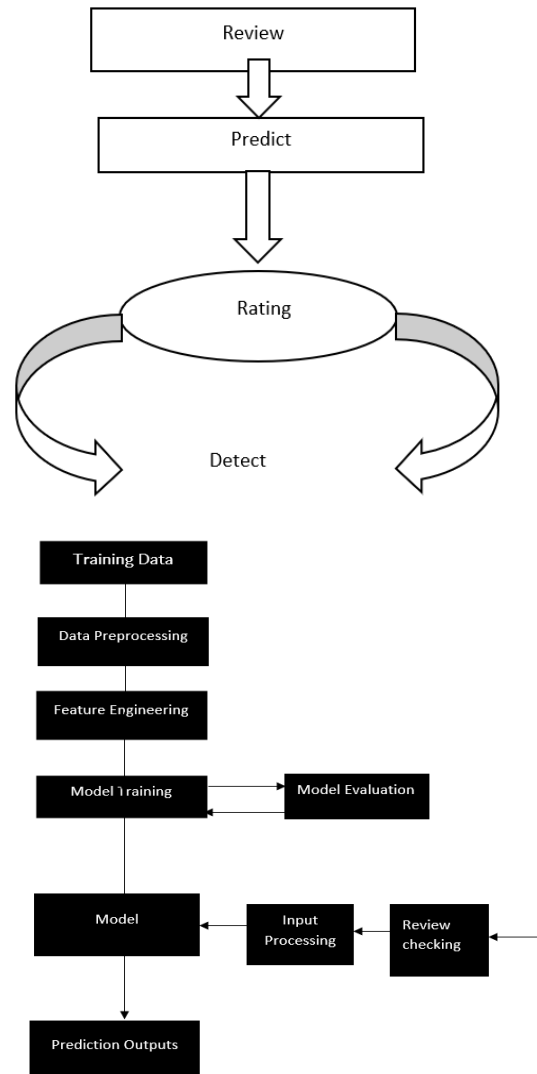
Advantages
A Larger Market
• Customer Insights Through Tracking And Analytics
• Fast Response To Consumer Trends And Market Demand
• Lower Cost
• More Opportunities To "Sell"

Disadvantages
• Lack Of Personal Touch
• Lack Of Tactile Experience
• Price And Product Comparison
• Need For Internet Access
• Credit Card Fraud

III. SYSTEM DESIGN

The DESIGN illustrate a feedback driven system designed to process and analyze user reviews. It begins with the "Reviews" stage, where textual or user generated feedback is collected this data is then pass it to the "Predict" phase in which an algorithm or model process the review content to generate a "Rating" likely reflecting sentiment quality or satisfaction. The rating is then subjected to a "Detect" process that continuously monitors the generated rating for patterns anomalies or trends. This detection phase forms a feedback loop, cycling back into the rating system, suggesting the insights from detecting may influence feature predictions or trigger action such as flagging unusual content or retraining the model. The cyclical

arrows emphasize the dynamic and iterative nature of the system highlighting it potential for learning adoption and improvement overtime.

The work flow outlines the complete machine learning pipeline from the preparation to generating prediction outputs. The process begins with Training data, which undergoes Data preprocessing to clean, normalize, and prepare it for analysis. After the preprocessing, feature engineering is performed to extract and select relevant attribute that improve the models performance. The refined data is then used in the model training phase where emission learning algorithm lens patterns from the data. Following this, model evolution is conducted to assess the model accuracy and effectiveness. This evolution loop feeds back into model training for optimization once the model is finalized it is used to process new inputs incoming data first going through review checking, ensure it is suitable and relevant followed by input processing. Where the data is formatted and transformed before being fed into the model that model then generates prediction outputs, completing the pipeline this design effectively highlights the importance of iterative training and real-time in put validation for robust predictive models.

IV. CONCLUSION

In this research we have developed a comprehensive system that leverages Natural Language Processing (NLP) and mission learning techniques to predict the helpfulness and sentiment based rating of mobile phone reviews. By integrating both croppers based and knowledge based approaches for semantic similarities analysis along with various feature extension method such as TF-IDF, GloVe and Word2Vec, we have the effectively translated and structure textual data into meaningful vector representation for modeling the expectations with several machine learning algorithms especially that Conventional Neural Networks (CNN) paid with Word2Vec was demonstrated so period performance in terms of accuracy, F1-score, precision and recall.

Moreover the inclusion of model interpretability through the LIME technique has added transparency to our prediction allowing the better understanding and trust in model outputs. The system design, which incorporates a robust feedback loop for continuous rating detection and model improvement, ensures adaptability and relevance over time. This predictive

model not only assist customers in making informed purchase decision making also provides businesses with critical insights into customer sentiment and product performance. By bridging the gap between raw review and content actionable intelligence, the proposed approach contributes significantly to domain of opinion mining and consumer analytics.

REFERENCE

[1] (IJACSA) International Journal Of Advanced Computer Science And Applications, Vol. 10, No. 6, 2019.
[2] Https://Scholarworks.Sjsu.Edu/Cgi/Viewcontent. Cgi?Referer=&Httpsredir=1&Article=1346&Context=Etd_Projects
[3] Predicting The Usefulness Of E-Commerce Products' Reviews Using Machine Learning Techniques https://Www.Informatica.Si/Index.Php/Informatica/Article/View/4155
[4] Sentiment Analysis On Mobile Phone Reviews Using Supervised Learning Techniques, I.J. Modern Education And Computer Science, 2019, 7, 32-43.