UNMASKING PRODUCT SENTIMENT: AN INVESTIGATION INTO SENTIMENT ANALYSIS TECHNIQUES FOR UNVEILING CUSTOMER OPINIONS AND REVIEWS

Mrs. N. Malathi[1], T. Dhana lakshmi[2], M. Praisy Rivritha[3], K. Pujitha Krishna Priya[4],

M. Tulasi[5], G. Mani Deepthi[6]

Asst Professor, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur,

Andhra Pradesh, India.[1]

BTech CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India.[2-6]

Abstract: This study delves into the realm of sentiment analysis techniques with a focus on unraveling customer opinions and reviews concerning products. In an era where online shopping and digital engagement have become ubiquitous, understanding customer sentiment is paramount for businesses to thrive. The project employs a Random Forest Classifier model integrated into a web application for real-time sentiment analysis. Through preprocessing text data and utilizing natural language processing tools, the model discerns between positive and negative sentiments expressed in customer reviews. The findings of this investigation shed light on the efficacy of sentiment analysis techniques in deciphering product sentiment, thereby aiding businesses in making informed decisions to enhance customer satisfaction and product quality.

Keywords: Sentiment Analysis, Customer Reviews, Product Sentiment, Natural Language Processing (NLP), Machine Learning, Random Forest Classifier, Text Preprocessing, Real-time Analysis.

I.   INTRODUCTION

In today's digital age, where consumer behavior heavily influences market trends, understanding customer sentiment has emerged as a crucial aspect of business strategy. With the proliferation of e-commerce platforms and social media, customers express their opinions and experiences with products and services through online reviews. These reviews not only serve as a reflection of customer satisfaction but also play a pivotal role in shaping brand perception and influencing purchasing decisions.

This project focuses on exploring sentiment analysis techniques to uncover product sentiments hidden within customer reviews. Leveraging a Random Forest Classifier model integrated into a user-friendly web application, the study aims to provide businesses with a tool to gauge customer sentiment in real-time.

Through preprocessing textual data and employing feature extraction methods, the model seeks to accurately classify reviews into positive or negative sentiments, thereby empowering businesses to gain actionable insights into customer perceptions.

II.   METHODOLOGY

The methodology adopted in this study encompasses several key steps aimed at harnessing sentiment analysis techniques to unveil product sentiments from customer reviews.

Data Collection: The project utilizes a dataset comprising customer reviews sourced from online platforms. These reviews serve as the primary source of textual data for sentiment analysis.

Text Preprocessing: To enhance the quality of textual data and mitigate noise, text preprocessing techniques such as case folding, punctuation removal, and stop-word removal are applied. This step ensures that the textual data is standardized and optimized for analysis.

Feature Extraction: Leveraging the CountVectorizer tool from the scikit-learn library, textual data is transformed into a numerical format suitable for machine learning algorithms. This process involves converting text into a matrix of token counts, enabling the model to interpret textual features.

Model Training: A Random Forest Classifier model is chosen for its robustness and ability to handle high-dimensional data. The model is trained on the preprocessed textual data, with sentiment labels assigned based on star ratings provided in the dataset.

Web Application Development: The trained model is integrated into a web application using the Streamlit framework, allowing users to input their reviews for sentiment analysis. The application provides real-time feedback on the sentiment expressed in the inputted review, aiding businesses in gauging customer perceptions..

III. MODELING AND ANALYSIS

Model and Material which are used is presented in this section. Table and model should be in prescribed format.
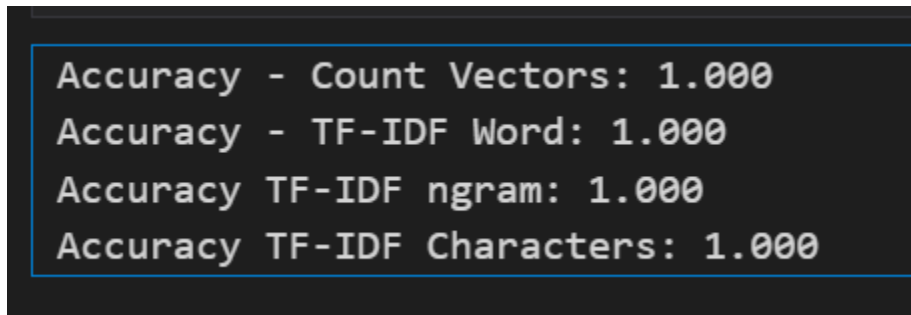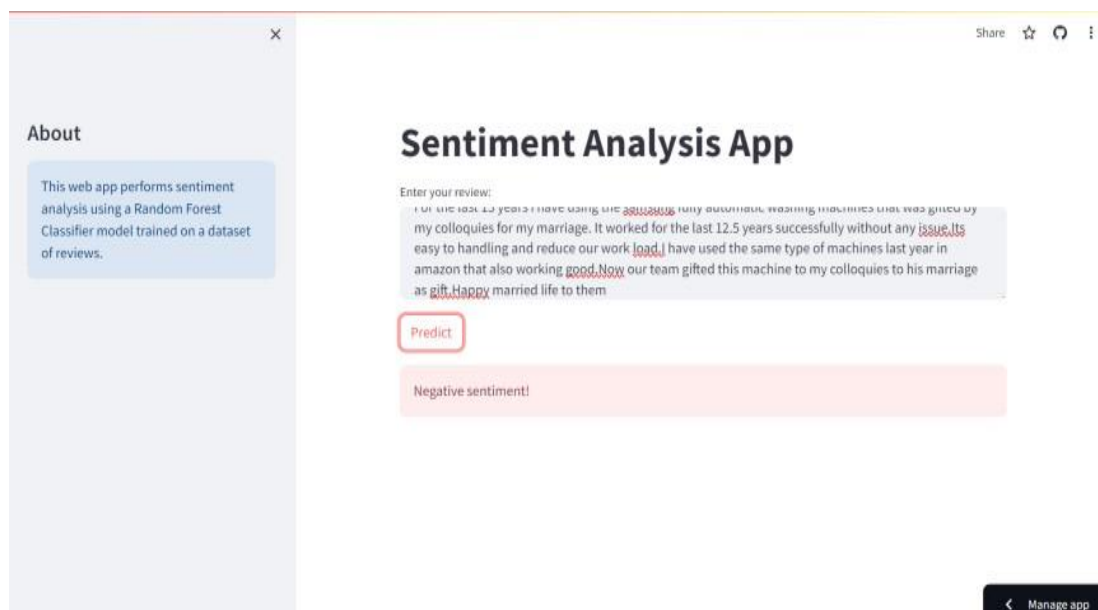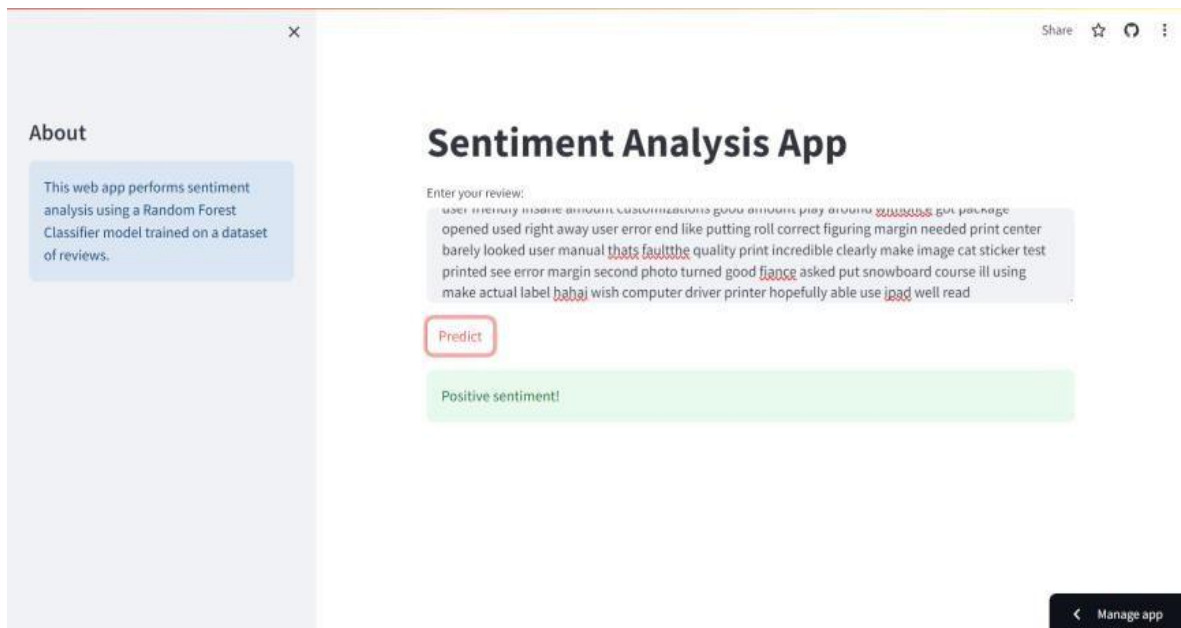

Figure 1: Accuracy

During the training phase, the model learns to identify patterns and features indicative of positive or negative sentiments within the textual data.

By analyzing the frequency of words and phrases across the dataset, the model develops an understanding of linguistic cues associated with varying sentiments.

IV. RESULTS AND DISCUSSION

V.   CONCLUSION

In conclusion, this project underscores the significance of sentiment analysis techniques in unmasking product sentiments embedded within customer reviews. By employing a Random Forest Classifier model integrated into a user-friendly web application, businesses can gain real- time insights into customer perceptions and sentiments.

The project demonstrates the feasibility and efficacy of leveraging machine learning algorithms for sentiment analysis, offering businesses a scalable and efficient solution for monitoring customer feedback and enhancing product quality. By embracing sentiment analysis as a strategic tool, businesses can foster greater customer satisfaction, loyalty, and brand advocacy, ultimately driving long-term success and competitiveness in the market.

In conclusion, NutriLens stands as a promising tool for individuals seeking to achieve their nutritional goals and adopt healthier eating habits.

REFERENCES

[1]. Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media, Inc.
[2]. Chen, J., & Liu, B. (2012). Feature selection for sentiment analysis. In Feature Extraction (pp. 215-241). Springer, London.
[3]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.
[4]. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
[5]. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993-1022.
[6]. Jurafsky, D., & Martin, J. H. (2019). Speech and Language Processing (3rd ed.). Pearson Education.
[7]. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
[8]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
[9]. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
[10].   Chollet, F. (2018). Deep Learning with Python. Manning Publications.
[11].   Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751.
[12].   Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

[13]. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in NaturalLanguage Processing (EMNLP), 1532-1543.

[14]. Ayyappa Chakravarthi M, Dr M. Thillaikarasi, Dr Bhanu Prakash Battula, published "Classification of Image Spam Using Convolution Neural Network" in International Information and Engineering Technology Association (IIETA) - "Traitement du Signal" Volume 39, No. 1

[15]. Ayyappa Chakravarthi M, Dr. M. Thillaikarasi, Dr. Bhanu Praksh Battula, published "Classification of Social Media Text Spam Using VAE-CNN and LSTM Model" in International Information and Engineering Technology Association (IIETA) - Ingénieriedes Systèmes d'Information Volume 25, No. 6.

[16]. Ayyappa Chakravarthi M, Dr. M. Thillaikarasi, Dr. Bhanu Praksh Battula, published a paper"Social Media Text Data Classification using Enhanced TF_IDF based Feature Classificationusing Naive Bayesian Classifier" in IJAST 2020

[17]. Ayyappa Chakravarthi M, Dr. M.Thillaikarasi, Dr. Bhanu Prakash Battula published "Social Media Data Classification – a Survey" in "Journal of Information and Computational Science" Volume 10, issue 3, March 2020