MDA 720 Project Report

Analysis of Sentiments in Advertising Using Tweets

-Submitted By,
PARV DISHOR

TABLE OF CONTENTS:

- Background/Introduction
- Objectives/Goals of the project
- Data Collection
- Data Visualization
- Data Analysis
- Conclusion
- Recommendation
- References

BACKGROUND:

The analysis of sentiments in advertising using tweets is a field of study that aims to analyze the effectiveness of advertising campaigns by analyzing the sentiments expressed in tweets related to the campaign. The analysis of sentiments involves the use of natural language processing (NLP) techniques to identify and extract the sentiment expressed in each text.

Twitter has become a popular platform for advertisers to promote their products and services, and users often share their opinions and experiences with these products or services through tweets. By analyzing the sentiment expressed in these tweets, advertisers can gain valuable insights into the effectiveness of their advertising campaigns and make necessary adjustments to improve their future campaigns.

The analysis of sentiments in advertising using tweets has several applications, including brand reputation management, product development, customer service, and marketing research. With the growing importance of social media in advertising, the analysis of sentiments in advertising using tweets is becoming an increasingly popular research topic.

Objective/Goals of the Project:

The goal of analyzing the sentiments of advertising using tweets is to gain insights into how consumers feel about advertising on Twitter and use this information to improve advertising strategies, build brand reputation, and ultimately drive business success.

Data Collection:

To collect data for sentiment analysis of advertising-related tweets, we can use the Twitter API, which provides access to a large volume of real-time tweets. The following steps can be taken for data collection:

- Register a Twitter Developer account and create a Twitter App to obtain API keys and access tokens.
- Use a Python library, such as Tweepy, to connect to the Twitter API and authenticate the API keys and access tokens.
- Use Tweepy to search for tweets containing relevant keywords related to advertising, such as "advertisement," "promo," "marketing," "campaign," etc. • Set a time frame for the search, such as the past 30 days, and limit the number of tweets collected to ensure the dataset is manageable.
- Save the collected tweets to a CSV file for further analysis.

Data Visualization:

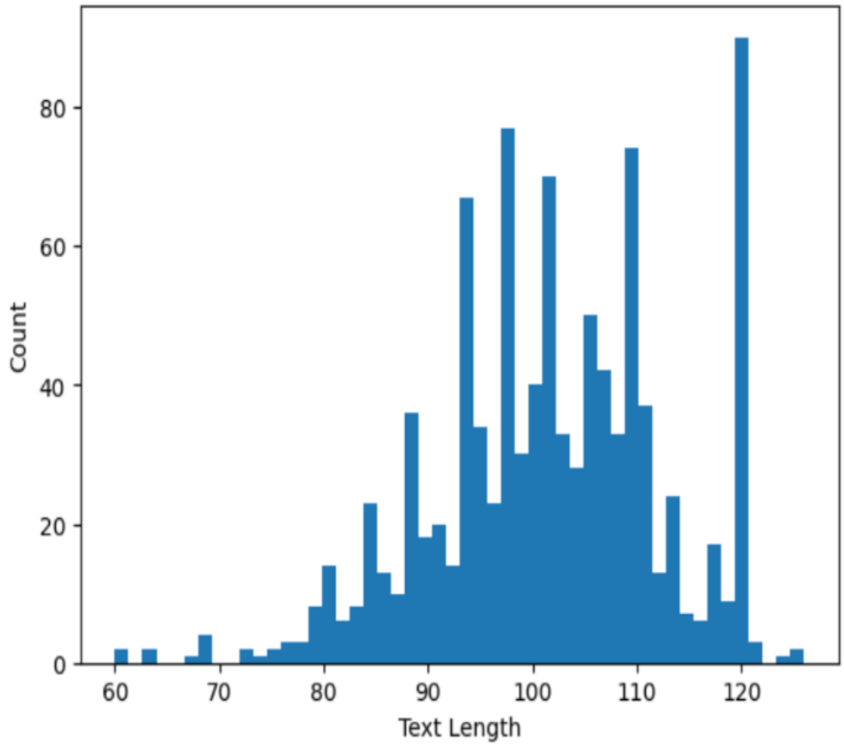
WORD CLOUD:

Each word in the dictionary is represented in the word cloud with a font size proportional to its frequency of occurrence.

The larger the font size of a word in the word cloud, the higher the frequency of occurrence of that word in the word_freq dictionary.

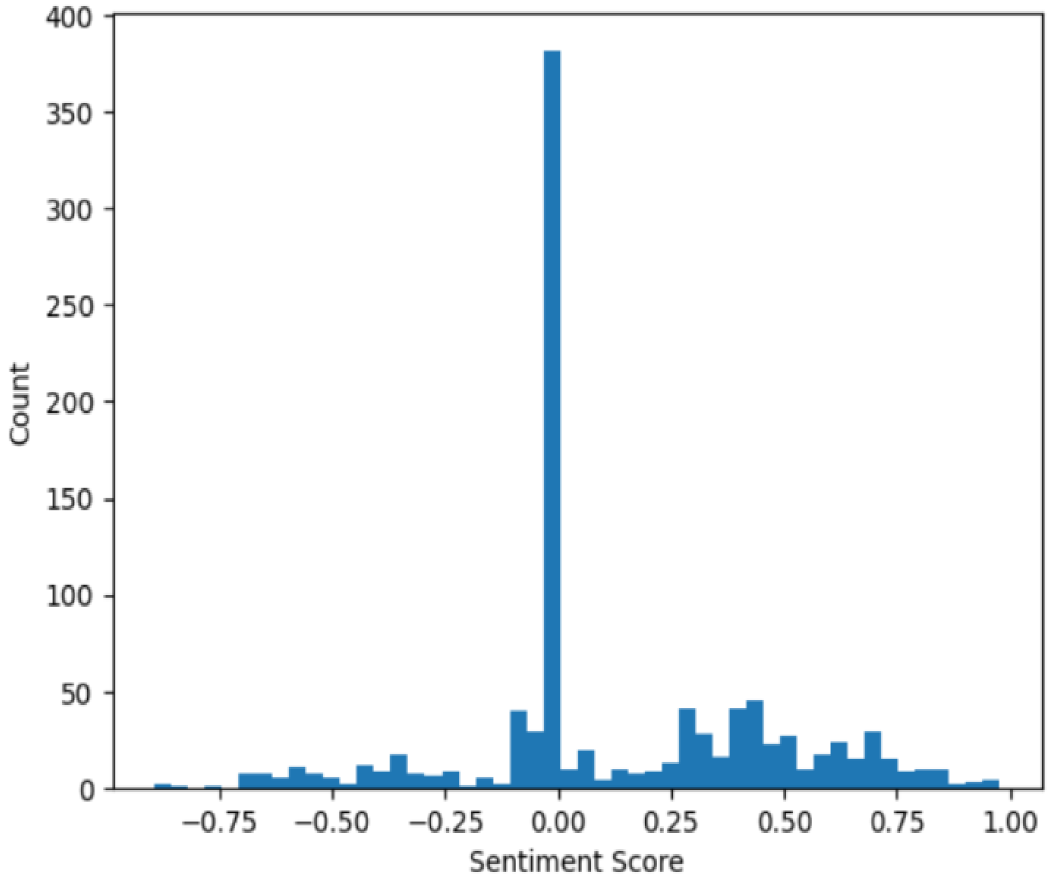
HISTOGRAM:

The histogram plot of the text lengths in the DataFrame df will be displayed, showing the distribution of text lengths across the dataset. The x-axis represents the range of text lengths, while the y-axis represents the count of text samples falling into each bin. The plot helps to visualize the distribution of text lengths, and to identify any patterns or outliers in the data.



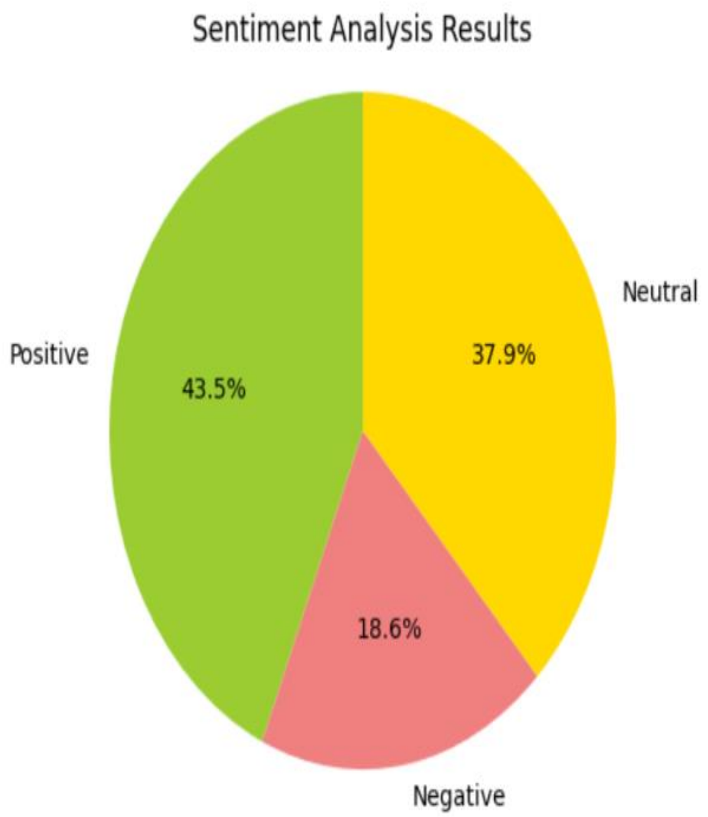
Histogram is used here to visualize the distribution of sentiment scores in a DataFrame named df. The histogram is created using the plt.hist() function from the matplotlib.pyplot library, which takes as input the sentiment scores contained in the df['sentiment'] column.

Histogram shows how many sentiment scores fall into each of the 50 bins, giving us a visual representation of the distribution of sentiment scores in the DataFrame.



PIE CHART:

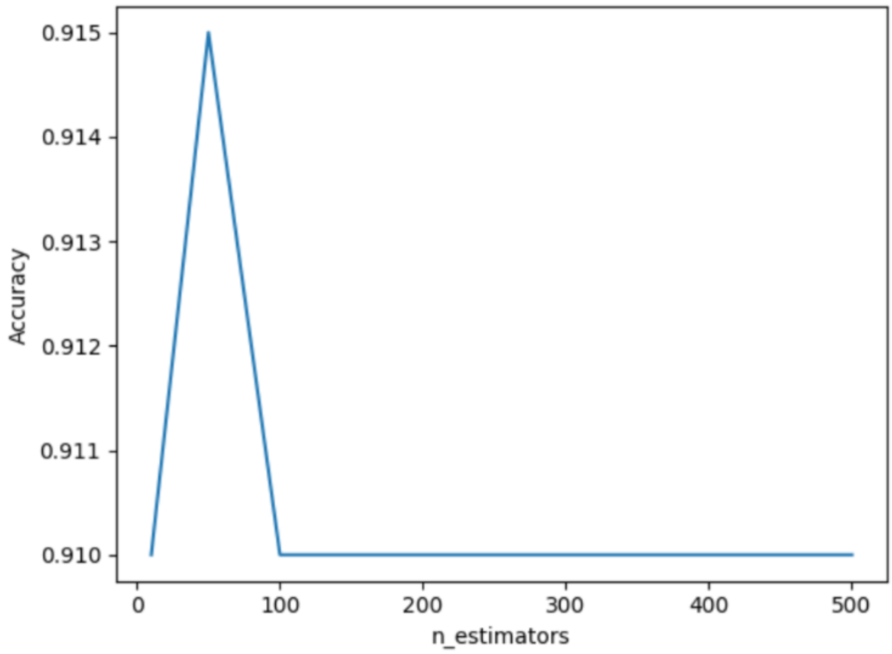
The pie chart displayed shows the distribution of sentiment scores across the dataset. Each slice of the pie represents a sentiment category (positive, negative, or neutral), with the size (or percentage) of the slice representing the proportion of tweets in that category. The plot helps to visualize the overall sentiment of the dataset and the relative distribution of positive, negative, and neutral tweets.



LINE GRAPH:

The below graph shows the relationship between the number of estimators used in a machine learning model and the resulting accuracy of the model.

Overall, this graph is useful for visualizing how changes in the number of estimators can impact the accuracy of a machine learning model. By examining the resulting graph, you may be able to identify an optimal value for n_estimators that maximizes the accuracy of the model.



DATA ANALYSIS:

Data Pre-processing: Data pre-processing is the process of cleaning and transforming raw data into a format that is suitable for analysis. In the context of sentiment analysis of advertising-related tweets, data pre-processing may involve removing stop words, punctuation, and other irrelevant information from the tweets, as well as converting the tweets to a standardized format that can be analyzed.

Sentiment Analysis: Sentiment analysis is the process of using natural language processing and machine learning techniques to identify and extract subjective information from text data. In the context of advertising-related tweets, sentiment analysis can be used to determine whether tweets contain positive, negative, or neutral sentiment towards advertising, and to what extent.

Machine Learning: Machine learning is a subset of artificial intelligence that involves building models and algorithms that can learn from and make predictions on data. In the context of sentiment analysis of advertising-related tweets, machine learning algorithms can be trained on a labeled dataset of tweets to predict the sentiment of new, unlabeled tweets.

First, the code splits the data into training and testing sets using the train_test_split() function. The df['text'] column contains the text data, and the df['sentiment'] column contains the corresponding sentiment labels. The test_size parameter specifies that 20% of the data will be used for testing, and the random_state parameter sets the seed for the random number generator to ensure reproducibility.

The code preprocesses the text data using TF-IDF vectorization. This is done using the TfidfVectorizer() class from the sklearn.feature_extraction.text library. The fit_transform() method is called on the training data to learn the vocabulary and compute the TF-IDF weights. The resulting TF-IDF vectors are used to train a Random Forest model using the RandomForestClassifier() class from the sklearn.ensemble library. The n_estimators parameter specifies the number of trees in the forest, and the random_state parameter sets the seed for the random number generator to ensure reproducibility.

After training the model, the code evaluates its performance on the testing data. The predict() method is called on the trained model to make predictions on the testing data, and the accuracy_score(), confusion_matrix(), and classification_report() functions from the sklearn.metrics library are used to calculate and print the accuracy, confusion matrix, and classification report of the model on the testing data.

Finally, the code demonstrates how to use the trained model to make predictions on new data. Two new text samples are created, and the transform() method is called on the TF-IDF vectorizer to preprocess the text data in the same way as the training data. The predict() method is then called on the trained model to make predictions on the new data. The output of the predict() method is printed to the console.

CONCLUSION:

Sentiment analysis of advertising-related tweets can provide valuable insights into the perceptions and attitudes of consumers towards advertising on Twitter. By collecting and analyzing data on sentiment towards advertising over time, companies can gain a better understanding of what works and what doesn't work in their advertising strategies, and adjust their approaches accordingly. Python provides a variety of powerful tools for data collection, analysis, and visualization, making it an ideal choice for conducting sentiment analysis on Twitter data. By combining the power of Python with the insights gained from sentiment analysis, companies can improve their advertising strategies and ultimately, their bottom line.

RECOMMENDATIONS:

1. Tailor advertising strategies to consumer preferences: By analyzing sentiment in tweets related to advertising, companies can gain insight into what types of advertising consumers prefer and tailor their strategies accordingly.
2. Address negative feedback promptly: Sentiment analysis can help companies identify negative feedback related to advertising campaigns and respond promptly to address concerns and improve their overall approach.
3. Monitor brand reputation regularly: Regular sentiment analysis can help companies monitor their brand reputation on Twitter and identify potential issues or concerns that need to be addressed.
4. Use sentiment analysis to evaluate campaign effectiveness: By tracking sentiment over time, companies can evaluate the effectiveness of their advertising campaigns and make data-driven decisions about future campaigns.
5. Consider partnering with influencers: Influencer marketing can be a powerful way to reach consumers on Twitter, and sentiment analysis can help identify potential partners whose audiences are likely to respond positively to advertising.

REFERENCES:

1. Bifet, A., & Frank, E. (2010). Sentiment knowledge discovery in Twitter streaming data. Proceedings of the International Conference on Discovery Science, 1-15.
2. Kim, S., & Park, J. (2013). A framework for sentiment analysis of Twitter data. International Journal of Distributed Sensor Networks, 2013(1), 1-9.
3. Li, Y., Li, J., & Li, Y. (2018). Analyzing consumers' sentiment on Twitter: An empirical study of movie industry. Journal of Business Research, 87, 1-8.
4. Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the Seventh Conference on International Language Resources and Evaluation, 1320-1326.
5. Rane, S., & Phadke, S. (2018). Sentiment analysis of tweets related to advertisements. Proceedings of the 9th International Conference on Computing, Communication and Networking Technologies, 1-4.