In my research, I employed a combination of unsupervised and supervised machine learning techniques to analyze the sentiment of Twitter data. Here's a comprehensive overview of my approach:

1. **LDA for Topic Modeling**: I began by using Latent Dirichlet Allocation (LDA), an unsupervised learning method, for topic modeling. I preprocessed the Twitter data by cleaning the text and preparing it through tokenization and lemmatization. I then transformed this preprocessed data into a Bag-of-Words model, which was essential for the LDA model to process. The LDA model(10 topics), implemented in Python using the Gensim library, was trained on this dataset to identify the main themes or topics in the tweets. I configured the LDA model to extract five topics, as indicated by the **num\_topics** parameter.
2. **Predictive Modeling for Sentiment Analysis**: After topic modeling, I shifted to sentiment analysis using a supervised learning approach. There is a large dataset with 1.6 million entries with pre-labeled sentiments (0 = negative, 2 = neutral, 4 = positive), which was treated with the same process in step 1. After it was fitted with topics, I set up a supervised machine learning model, **LightGBM**. Eventually, four variants got the averaged scores in the among the 4 \* 1000 data.

In summary, my research involved a comprehensive approach using Python scripts for topic modeling, sentiment prediction, and distribution analysis of Twitter data. I employed LDA for extracting themes from the tweets, followed by a GBM or a similar predictive model for sentiment analysis, and concluded with statistical methods to analyze the sentiment distribution.