

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DAY – 18

Date: Jul 16, 2025

What is Sentiment Analysis?

Sentiment analysis is the process of classifying whether a block of text is positive, negative, or neutral. The goal that Sentiment mining tries to gain is to be analysed people's opinions in a way that can help businesses expand. It focuses not only on polarity (positive, negative & neutral) but also on emotions (happy, sad, angry, etc.). It uses various Natural Language Processing algorithms such as Rule-based, Automatic, and Hybrid.

Why is Sentiment Analysis Important?

- **Customer Feedback Analysis:** Businesses can analyze customer reviews, comments, and feedback to understand the sentiment behind them helping in identifying areas for improvement and addressing customer concerns, ultimately enhancing customer satisfaction.
- **Brand Reputation Management:** Sentiment analysis allows businesses to monitor their brand reputation in real-time.
By tracking mentions and sentiments on social media, review platforms, and other online channels, companies can respond promptly to both positive and negative sentiments, mitigating potential damage to their brand.
- **Product Development and Innovation:** Understanding customer sentiment helps identify features and aspects of their products or services that are well-received or need improvement. This information is invaluable for product development and innovation, enabling companies to align their offerings with customer preferences.
- **Competitor Analysis:** Sentiment Analysis can be used to compare the sentiment around a company's products or services with those of competitors.
Businesses identify their strengths and weaknesses relative to competitors, allowing for strategic decision-making.

How does Sentiment Analysis work?

Sentiment Analysis in NLP, is used to determine the sentiment expressed in a piece of text, such as a review, comment, or social media post. The goal is to identify whether the expressed sentiment is positive, negative, or neutral.

Preprocessing

Starting with collecting the text data that needs to be analysed for sentiment like customer reviews, social media posts, news articles, or any other form of textual content. The collected text is pre-processed to clean and standardize the data with various tasks:

- Removing irrelevant information (e.g., HTML tags, special characters).
- Tokenization: Breaking the text into individual words or tokens.
- Removing stop words (common words like "and," "the," etc. that don't contribute much to sentiment).
- Stemming or Lemmatization: Reducing words to their root form.

Analysis

Text is converted for analysis using techniques like bag-of-words or word embeddings (e.g., Word2Vec, GloVe). Models are then trained with labeled datasets, associating text with sentiments (positive, negative, or neutral).

After training and validation, the model predicts sentiment on new data, assigning labels based on learned patterns.

What are the Approaches to Sentiment Analysis?

There are three main approaches used:

Rule-based: Over here, the lexicon method, tokenization, and parsing come in the rule-based. The approach is that counts the number of positive and negative words in the given dataset. If the number of positive words is greater than the number of negative words then the sentiment is positive else vice-versa.

Machine Learning: This approach works on the machine learning technique. Firstly, the datasets are trained and predictive analysis is done. The next process is the extraction of words from the text is done. This text extraction can be done using different techniques such as Naive Bayes, Support Vector machines, hidden Markov model, and conditional random fields like this machine learning techniques are used.

Neural Network: In the last few years neural networks have evolved at a very rate. It involves using artificial neural networks, which are inspired by the structure of the human brain, to classify text into positive, negative, or neutral sentiments. It has Recurrent neural networks, Long short-term memory, Gated recurrent unit, etc to process sequential data like text.

Hybrid Approach: It is the combination of two or more approaches i.e. rule-based and Machine Learning approaches. The surplus is that the accuracy is high compared to the other two approaches.