Paper Review: Detection of Sarcasm on Amazon Product Reviews using Machine Learning Algorithms under Sentiment Analysis

Summary: Amazon is a leading online shopping platform with 310 million active users and more than 350 million products. Here, people leave tons of reviews according to the product's quality and usage. And they do this in many ways, Sarcasm is one of them. Sarcasm is the bitter way of conveying the message. It can be even said as an acute or an expression of a state of ambivalence in conveying the message. The main idea of this project is to extract sarcasm from text reviews and classify them according to positive and negative reviews of the product using Sentiment Analysis.

Motivation: To analyze the effectiveness of Sentiment Analysis in opinion based texts and develop an algorithm that will detect and extract sarcasm in product reviews to understand the gist of the text.

Contribution: This study aims to advance identification of positive and negative reviews of people that mostly do by means of sarcasm in the product review section.

Methodology:

- Data Collection: Collecting and retrieving data from Amazon's database.
- Data Pre-Processing: Analyzing the data by tokenization, Lemmatization and Polarity Labeling.
- Feature Extraction: Term Frequency, TF-IDF, N-gram.
- Classification Used: Categorizing the data on observations. (SVM, Random Forest)

Conclusion: After going through different processes of methodologies in the dataset, sarcastic sentences are found after the classification and these sentences are labeled as sarcastic. After the sarcasm detection process, polarity is checked and labeled according to positive or negative.

Limitations:

- Contextual Understanding : Reviews posted by users might lack context or subtle cues that might make the program difficult to interpret.
- Data Imbalance : It can affect the program's ability to learn sarcasm patterns adequately, leading to biased predictions favoring the majority class (non-sarcastic reviews).

Synthesis:

- The model can be used in product development enhancement.
- The model can be used for more data sources other than texts.