Report - Sentiment Analysis on IMDB Movie Reviews

The sentiment analysis evaluation aimed to compare different approaches, including VADER sentiment analyzer and machine learning models, in predicting sentiment based on "User Review" data. The results obtained from these evaluations, along with a thorough analysis, shed light on the strengths and limitations of each approach.

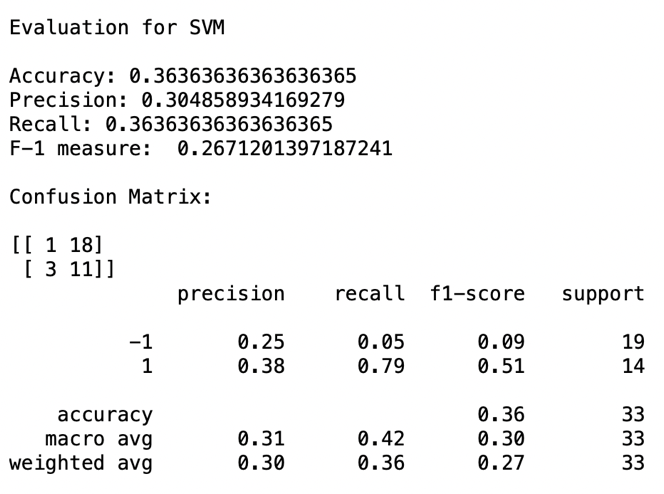
# VADER Sentiment Analysis

VADER, a rule-based approach, provides quick insights into sentiment analysis but struggles with subtle nuances in language and context. For instance, a correctly classified review expressing positive sentiment can be accurately identified by VADER. However, a misclassified review with negation and subtle language, such as "I consider myself a big fan of Marvel and Spider-Man, but something wasn't right about this film," showcases VADER's difficulty in capturing negative sentiment.

# Machine Learning Approaches

Machine learning models using the TF-IDF and count matrix approaches achieved varying results in sentiment prediction. The SVM model using TF-IDF achieved an accuracy of 36.36% and an F1 score of 0.2671. The Decision Tree model using TF-IDF achieved an accuracy of 45.45% and an F1 score of 0.355. The TF-IDF approach assigns weights to words based on their occurrence in a document and across the entire corpus. While it captures the importance of words, the provided results indicate that TF-IDF alone did not yield highly accurate sentiment predictions.

TF-IDF Evaluation

A screenshot of a computer screen

Description automatically generated

On the other hand, SVM model using the count matrix approach achieved an accuracy of 45.45% and an F1 score of 0.4289. The Decision Tree model using the count matrix approach achieved an accuracy of 45.45% and an F1 score of 0.355. The count matrix approach represents the frequency of words in the document. While it captures occurrence counts, the results demonstrate that it did not significantly improve performance.

BoW Scheme Evaluation

A screenshot of a computer program

Description automatically generatedA screenshot of a computer screen

Description automatically generated

In order to strengthen above evaluation, a few examples have been extracted to ponder over the result. A review, "I enjoyed it." expressing positive sentiment, was accurately classified by both VADER and ML models. The review "Wack Friday!" for the movie "Black Friday" was predicted as neutral by VADER, but the ML models classified it as positive. However, the review is actually negative. VADER's may have missed the negative connotation of "wack," while the ML models focused on other positive words, failing to capture the negative sentiment. This highlights the challenge of accurately classifying sentiments with informal language

Moreover, misclassifications can occur due to subtle use of positive and negative words or opinions, making sentiment prediction challenging. VADER misclassified the review "I'd rather gargle with hydrochloric acid than see a 'Baby Geniuses 3'!" as neutral. VADER struggles to capture negative intensity, possibly due to phrases like "rather" and "than." In another review "A decent production with no script," ML models predicted it as positive, while VADER classified it as negative. The presence of positive words influenced the ML models, ignoring the negation and "with no script." The models overlooked negation, whereas VADER considered it. Capturing nuanced sentiments is a challenge when positive/negative elements coexist.

The discussion highlights the limitations of VADER and machine learning models in sentiment analysis, leading to inconsistent results. VADER's reliance on predefined lexicons may overlook nuanced language, while machine learning models require sufficient and diverse training data to generalize accurately. Gathering more extensive training data is crucial for improved performance. In conclusion, sentiment analysis evaluations reveal language's impact on classification and the need for refined techniques. Enhancements should focus on better training data and advanced techniques for more accurate sentiment prediction in movie reviews.