

The document is a research paper from the 2022 IEEE International Conference on Machine Learning and Applications (ICMLA), focusing on the application of semi-supervised machine learning to analyze Twitter data related to COVID-19 for the presence of Asian hate speech. Authors Caitlin Richardson, Sandeep Shah, and Xiaohong Yuan from the University of Tampa and North Carolina Agricultural and Technical State University present their findings on the effectiveness of semi-supervised learning compared to supervised learning in classifying tweets as hate, counter-hate, or neutral.

The study utilizes a dataset of tweets collected during the COVID-19 pandemic, which was expanded by the authors to include additional labeled data. They employ a count vectorizer as a feature extraction method and a Support Vector Machine (SVM) as a classifier. The paper explores two semi-supervised learning methods: self-training and co-training. In self-training, the model is initially trained on a small labeled dataset and then iteratively expands its training data by adding high-confidence predictions from the unlabeled data. Co-training involves training two classifiers on two independent sets of features and then using both to label unlabeled data. The researchers conducted experiments with varying amounts of labeled training data (15%, 20%, and 25%) and compared the performance of their semi-supervised models to a supervised model trained on 80% of the labeled data. They found that the semi-supervised models, particularly those trained on 20% and 25% of the labeled data, performed comparably to the supervised model, achieving high F1 scores and accuracy. This suggests that semi-supervised learning can be a cost-effective alternative to supervised learning, reducing the need for large amounts of labeled data.

The paper also discusses the challenges of classifying tweets, such as the difficulty in distinguishing between hate speech and offensive language. The researchers acknowledge that their model sometimes misclassifies tweets containing swear words, highlighting a limitation in the model's ability to understand context.

In conclusion, the study demonstrates that semi-supervised machine learning, particularly self-training, can be an effective approach for analyzing social media data for hate speech. The authors suggest future work could explore other machine learning algorithms, co-training methods, and strategies to address overfitting and bias in the models. The research is supported by various grants and acknowledges the contributions of multiple funding sources.