**Abstract**

**The use of popular social media such as Twitter, raises positive and negative content. One of the negative content that often appears is Hate Speech. The purpose of this study is to build a Hate Speech detection system by implementing feature expansion to minimize vocabulary mismatches. This study also uses several methods such as *TF-IDF* to extract features from a sentence. While the feature expansion uses the *GloVe* method and uses the *logistic regression*, *random forest* and *naive Bayes* classification method. Results in a fairly high accuracy of 86.44% in the *random forest* algorithm with the Top 5 feature in detecting Hate Speech so that it can be more accurate in detecting which tweets contain hate speech. and which ones do not contain Hate Speech.**

**Keywords: TF-IDF, GloVe, feature expansion, classification, Twitter, Hate Speech**