

Abstract

Sexual Harrasment is behavior marked by harassing messages or comments, threats or obscene things, invitations to do pornographic actions, or all deviant behavior carried out in online media. The problem of sexual harassment on social media is an important thing that must be studied. Sentiment analysis can be used as a solution to identify social media regarding sexual harassment. The purpose of this final project is to classify speech data that leads to sexual harassment based on the quid pro quo class and hostile work environment from the #MeToo tweet data that has been summarized by the survey website theprofesorission.com using the Gaussian Naïve Bayes and KNN methods. The system is built through the preprocessing and oversampling stages. distribution of training data and test data using k-fold and evaluation by comparing the accuracy values. In system testing, the accuracy value of the Naïve Bayes classification model using K-Fold cross Validation is 90.7% and the KNN classification model using K-Fold cross validation reaches a value of 87.9%.

Keywords: sexual harassment, gaussian naïve bayes, sentiment analysis, quid pro quo, hostile work environment.