

Abstract

Social media has made it easier for people to express their opinions and beliefs, but this freedom is often misused to commit online abuse, especially harassment of women or also known as misogyny. This research was conducted to select the best classification method between Support Vector Machine and Stochastic Gradient Descent methods by using the Support Vector Classification model and Stochastic Gradient Descent Classifier to classify misogyny data. Classification is performed on a dataset containing text with elements of misogyny from the social media Reddit. The evaluation metrics considered are the Accuracy, Precision, Recall, and F1-Score that are calculated from the Confusion Matrix. The results obtained show that SVC with Undersampling produces the best performance in identifying misogyny content with 82,84% Accuracy, 29% Precision, 71% Recall, and 41% F1-Score.

Keywords: misogyny detection, text classification, support vector machine, stochastic gradient descent