

ABSTRACT

The development of the film in Indonesia escalated, ranging in terms of storyline, the content of the film and the actor. In this research the authors want to predict attendance film Indonesia in twitter. Twitter is a social media used to search for data through audience comments tweet-tweet users. And used to explore data that aims to research, from people who talked about film in twitter can make a trend for prediction. Through a prediction of expected production house producer and strategize in order for films that will be sold over the counter to appear. On previous research conducted using the data of Hollywood movies, However, this research uses data on film Indonesia. In this research do two approaches through user and tweet. The approach features using user data, i.e. follower official, production house, Director, producer, and lead artist 3 in the film, While the features of the data i.e. tweet, retweet, like, number of Tweets containing the URL. The purpose of this research is to apply pre-processing process to form a more suitable input data were incorporated into the process of naïve bayes classifier algorithms. The test results using the naïve bayes classifier algorithm is of 78.79%. To improve the performasi and prevent overfitting from the system built, can be done by adding a feature subset selection on naïve bayes classifier which have an average accuracy of 77.96%.

Keywords : Twitter, *Crawling Data*, *Pre-processing*, *Naive Bayes Classifier*, *Measuring performance*, Film.