Abstract

The main problem of an online forum user who often comment on a thread or comment on the topic but what is not useful or does not relate at all to the content of the thread or topic . Comments or postings like this is known as junk post . Lots of losses due to junk post , such as adding the topic page resulting in another user 's hard to read the whole topic and add time to read the forums , change the direction of the conversation so that it can evolve other junk post from another user , which is an important staple topic could be obscured by junk post so there are a few things missing , and a variety of other problems .

For that reason, this research aims to classify a post or comment, including junk or not, so expect the future to reduce junk posts on internet forums . Methods that will be used for this research is to use a K - Nearest Neighbor method for text classification . In its application , was also weighted with the tf - idf post and to calculate the similarity values of an existing post with a thread or topic related to using Pearson Correlation Distance.

The results of this study indicate that *f1-measure* obtained for first test scenario was 70.71% with k = 93, for second test scenario was 70.00% with the k = 61, and for third test scenario was 71.72% with k = 41 . In each scenario, the resulting *f1-measure* is better when not using a stop word removal . This is because post in the forum are not standardized and very varied so it is less suitable when the stop word removal used in this study.

Key word : *text classification, junk post, k-nearest neighbor, pearson correlation distance, tf-idf.*