# Abstract

The current technological developments can facilitate activities that must be performed by humans. With the a lot of information on the internet, so people will easily access and process information from anywhere. Such as checking the degree of similarity of documents that take a long time and a great power, it can be replaced by applying text mining. According to Indonesian Big Dictionary, the resemblance is almost the same or similar. In this final project will be calculated similarity of documents that abstract data thesis students of the Faculty of Informatics Telkom University.

Text Mining is a discussion that analyze and manage text into an information that can be processed for a specific purpose. In the Text Mining be known about preprocessing, consisting of folding case, tokenizing, filtering, and stemming. This is done before the tf-idf weighting and calculation cosine similarity. The method used is Cosine Similarity. Cosine Similarity is a method of calculating the distance between vectors A and B which produces x cosine angle between the two vectors. Cosine value of the angle between two vectors determine the similarity of two objects are compared where the smallest value is 0 and the largest value is 1. A value of 0 indicates that the two abstracts were compared are not similar at all and getting closer to the value of 1 means that the greater the degree of similarity.

Attaining a yield of 0.5729 at Pearson correlation coefficient calculation, it can be stated that the correlation between the calculations using TF-IDF and Cosine Similarity method with manual assessment has a positive linear correlation value because it is located between 0 and 1.

Keywords: *text mining, similarity, final project, Cosine Similarity, Telkom University*