# ABSTRACT

*Plagiarism is the act of taking part or all of people's ideas in the form of documents or texts without attaching the sources of information retrieval. Therefore plagiarism detection is necessary to reduce plagiarism and keep the originality of people's work. This research aims to detect the similarity of text documents using the Word2vec method and TF-IDF extraction fiture to determine the difference in values. The document used for comparison of this text is containing of 116 Indonesian abstracts. From the result, when stemming is applied the result was on average 5%, which is higher when stemming isn't applied. Produces a similarity value over 50% for documents with a high level of similarity. Meanwhile for documents with a low level of similarity or not plagiarism produces a similarity value under 30%. The step of preprocessing is consisting of folding cases, tokenizing, removeal stopwords, and stemming. After the preprocessing process, the next step is weighting TF-IDF and Word2vec. Than the next step was the similarity value uses Cosine Similarity to get percentage of similarity value. Based on the results of the experiment, Word2vec results the similarity value higher by an average of 28% compared to the TF-IDF weighting value.*

*Keyword: Cosine Similarity, Document, plagiarism, preprocessing*, TF-IDF, *Word2vec*