BAB I

PENDAHULUAN

Semantic similarity has an important role in the field of linguistics, especially those related to the similarity of words meaning. Semantic similarity between words is the search for similarities between two words or more. In terms of the similarity of words meaning, two words may differ syntactically but have the same meaning. For example, Me and I have the same meaning. Calculating the similarity of words meaning has been widely represented in the field of linguistics with basic rules as a result of the reasoning of human thought. This calculation can also be done through the field of computer science, namely the study of Natural Language Processing and Text Mining based on the field of linguistics.

Natural Language Processing is one of the fields of science of Artificial Intelligence that deals with the interaction between computers and natural human language. Computers need to process the human language that is received first so that the intentions of humans can be understood and then provide the appropriate response. For example, a computer must know the value of how many similarities between Me and I.

The main technique for calculating the similarity and relevance of words meaning using the Word2Vec model from word embeddings. Word2Vec is a model used to represent words into vectors. Then, the similarity value can be generated using the Cosine Similarity formula of the word vector values produced by the Word2Vec model. In the construction of the Word2Vec model called the training process, there are several features that used to produce the Word2Vec model including windows size and vector dimensions configuration. Some previous study mostly used Windows size and vector dimensions configuration to produce the Word2Vec model. In this study several Windows size and vector dimensions configuration were used to compare the similarity values of each configuration of the resulting Word2Vec model. The configuration of the Word2Vec model that produces the best similarity values will be the result of this study. It is important to know the best configuration of the Word2Vec model to find the best value for similarity word meanings.

.