Abstract

Classification is the process to find a set model or function that can describe and distinguish data classes or concepts, in order to use these models to predict the class of an object where the class is unknown. One of the techniques or methods used in building a data classification model is Bayesian Network (BN). BN consists of two main parts, namely to build the structure of DAG (directed acyclic graph) and calculate the CPT (conditional probability table). Currently learning methods have been developed that allows to build BN structures directly from the database. These methods include search and scoring method (scored based) and dependency analysis (constraint based).

In this final project used learning algorithm with dependency analysis approach that is the Three Phase Dependency Analysis $\tilde{O}$ (TPDA $\tilde{O}$). This algorithm can construct a Bayesian Network structure from the known sequence of data attributes. TPDA $\tilde{O}$ build structures DAG (directed acyclic graph) by identifying / analyzing relationships parole (conditional independence test) or also called CI tests between attributes, where CI is a constraint in building the BN structure. From DAG structure that has been established by TPDA $\tilde{O}$ can be determined a data classification.

The purpose of this final project is to analyze the accuracy of the BN which was built using TPDA $\tilde{O}$. From the test results in cases of chest clinic and the primary tumor there is a change resulting accuracy on each data. Changes in accuracy are possible under the influence of threshold and data used. The average accuracy obtained in the case of chest clinic was 86.2% with 900 training data and testing records 100 records, whereas in the case of the primary tumor was 79.17%.

Keywords: Bayesian Network (BN), construction BN, TPDA $\tilde{O}$, classification, Conditional Independence Test.