



A REVIEW: DATAMINING USING DECISION TREE (DT)

Irfan abbas*

Departemen of Economic and businesses Digital, Muhammadiyah berau University

A. INTRODUCTION

Data mining is the computational process of discovering patterns in large data sets involving intersecting machine methods learning systems, statistics and databases. The common goal of The data mining process involves extracting information from a data set and turn it into an easy to understand structure for later use (Abellán et al., 2018). The task of supervised classification is data mining area. It is based on a series of confirmed observations by a set of characteristics or attributes associated with a known class label. A classifier obtains a set of rules for predicting the values of class variables from a subset of observations called the training set. for each new observation. To check the quality of this set of rules, A separate set of observations called the test set is used. The a measure of classification error is obtained during validation. Real-world datasets can contain noise. Noise can be defined as obscuring the relationship between the properties of an instance and its class. In particular, the situation in which a data set has incorrect class labels is called class noise or label noise. This situation is mainly due to flaws in data learning and/or test acquisition process that may misdiagnose disease (Abellán et al., 2018; Moral-García et al., 2022)(Mantas and Abellán, 2014) methods, human error in class labeling, etc. The presence of label noise in the data can make it more difficult to extract knowledge from the data. The study of the methods applied in this situation is an important task in the field of data mining (Abellán et al., 2018).

B. DECISION TREE

A Decision tree (DT) is a very useful tool for classification is a fantastic classification technique. It has an easy to understand structure. Additionally, developing the categorization model often takes little time, Its structure is simple and easy to interpret. A decision tree (DT), also known as a classification tree, is a type of of the classifier can be viewed as a set of rules in the form of a tree. An attribute variable is introduced in each node. Each page contains a class variable label or a set of probabilities for each class label. Hunt's work is the origin of decision trees, although it began to gain prominence with the publication of the ID3 algorithm proposed by Quinlan (1986). Then, Quinlan proposed algorithm C4.5 (Quinlan, 1993), which is an improvement of the previous ID3 with better results. (Abellán et al., 2018; Mantas et al., 2016; Mantas and Abellán, 2014; Moral-García et al., 2022)

C. ALGORITHM ID3

Many decision trees are created using the ID3 algorithm and its extension C4.5. The ID3 algorithm is improved by C4.5 in a number of ways, including how it handles continuous attributes, how it handles training data with missing attribute values, and how it prunes builtin



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

