Rough Set Discretization: Equal Frequency Binning, Entropy/MDL and Semi Naives algorithms of Intrusion Detection System

Noor Suhana Sulaiman
TATi University College
24000 Kemaman, Terengganu, Malaysia
suhana@tatiuc.edu.my

Rohani Abu Bakar
Universiti Malaysia Pahang
Lebuhraya Tun Razak, 26300 Kuantan, Pahang, MALAYSIA
rohani@ump.edu.my

ABSTRACT: *Discretization of real value attributes is a vital task in data mining, particularly in the classification problem. Discretization part is also the crucial part resulting the good classification. Empirical results have shown that the quality of classification methods depends on the discretization algorithm in preprocessing step. Universally, discretization is a process of searching for partition of attribute domains into intervals and unifying the values over each interval. Significant discretization technique suit to the Intrusion Detection System (IDS) data need to determine in IDS framework, since IDS data consist of huge records that need to be examined in system. There are many Rough Set discretization technique that can be used, among of them are Semi Naives and Equal Frequency Binning.*

Keywords: Rough Set, Semi Naives, Equal Frequency Binning, Discretization, Intrusion Detection System