# Data Mining with Outlier

*by Mark Chatchai Wangwiwattana*

## Abstract

Outliers are common in data mining. Typically, outliers need to be removed before mining;otherwise, the result would be incorrect. However, removing outliers requires full understanding about data. In many cases, it is difficult to identify outliers such as dependent variables. Moreover, removing the outliers in real-time applications or data streaming is somewhat difficult. In order to address the issue, I present the statistical models that robust against outliers so call Robust Statistics. Finally, I introduce Random Sample Consensus (RANSAC) algorithm allowing to create a regression model with data with outliers which it is impossible for ordinary least square linear regression model. RANSAC is not only robust against outliers, but it also has high performance for real-time applications.