Minimum Redundancy Maximum Relevance Implementation as Dimension Reduction Technique for Colon Cancer Classification with Random Forest

I.G.N.P.Vasu Geramona[1],Adiwijaya[2], Widi Astuti[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]ngurahgera@student.telkomuniversity.ac.id,
[2]adiwijaya@telkomuniversity.ac.id,
[3]widiwdu@telkomuniversity.ac.id.

Abstract

Cancer is a deadly disease. Quoting information from the Ministry of Health of the Republic of Indonesia in 2017 nine million people died from cancer. Therefore we need a method to detect cancer, one of which is by gene expression. Microarray is a technique of gene expression. Microarray itself has many features, many of these features are not always related to the problem being faced. So we need a dimension reduction technique to select features that correspond to the problem being faced.

In this final project a dimension reduction technique will be used using the Minimum Redundancy Maximum Relevance which will then be abbreviated as MRMR. The Classifier that will be used is Random Forest, where this technique creates several trees to classify data and then will vote for the most results. The MRMR equation used is FCD and FCQ because the data used is continuous. After the process done, the result from classify microarray data using FCQ is 83.87% and with FCD 61.29%

Keywords: microarray, gen expression, random forest, MRMR