CLASSIFICATION ANALYSIS OF STROKE PATIENTS CONDITION BASED ON RISK FACTORS OF RSU HAJI SURABAYA PATIENT USING LOGISTICS REGRESSION AND SUPPORT VECTOR MACHINE (SVM)

Name : Dwi Ratnasari
NRP : 1308100097
Department : Statistics FMIPA – ITS
Supervisor : Santi Wulan Purnami, S.Si, M.Si, Ph.D

Abstract

Stroke is the third highest cause of death in the world after heart disease and cancer. In Indonesia stroke is ranked as first cause of death in all age groups with the percentage of 15.4%. Complications associated with the many events that occur in stroke patients then performed the grouping condition of stroke patients after treatment based on the diseases that accompany stroke with the methods using support vector machine (SVM) and logistic regression methods. Classification using binary logistic regression produces the maximum classification accuracy of 74.19% with 3 risk factors affecting the condition of patients when discharged from the hospital, namely age, TIA (transient ischemic attack) and stroke type suffered. According to the L1-Norm variable selection predictor variables that have the most impact on the final condition of the patient when discharged from the hospital is a type of stroke, TIA and gender. Classification with SVM produced the maximum value of the classification accuracy of 80% where the variable selection results showed that all patients had risk factors that affect the patient's condition when discharged from the hospital. It is also evident if the variable with the smallest coefficient of reduced lead classification accuracy also decreases.

Key words: Classification, Logistic Regression, Variable Selection, SVM, Stroke, Risk Factors, Patient Conditions.