

**ESTIMASI MODEL REGRESI LOGISTIK NONPARAMETRIK  
ADITIF BERDASARKAN ESTIMATOR POLINOMIAL  
LOKAL**

**SKRIPSI**



MILIK  
PERPUSTAKAAN  
UNIVERSITAS AIRLANGGA  
SURABAYA

MPM. 54/10  
Ama  
e

**SITI AMANAH**

**DEPARTEMEN MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS AIRLANGGA  
SURABAYA  
2008**

Siti Amanah, 2008. **Estimasi Model Regresi Logistik Nonparametrik Aditif Berdasarkan Estimator Polinomial Lokal**. Skripsi ini dibawah bimbingan Nur Chamidah, S.Si, M.Si dan Ir. Elly Ana, M.Si, Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga.

## ABSTRAK

Regresi logistik adalah analisis regresi yang menggambarkan hubungan antara variabel respon  $Y$  dengan sekumpulan variabel prediktor  $X$ , dengan  $E(Y|X)$  berada dalam interval  $[0, 1]$ . Model regresi logistik dapat ditulis :

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \sum_{j=1}^p m_j(x_{ji}), \quad i = 1, 2, \dots, n$$

dengan  $\log\left(\frac{\mu_i}{1-\mu_i}\right)$  adalah fungsi logit.

Tujuan dari skripsi ini adalah untuk mengestimasi fungsi-fungsi dalam model regresi logistik nonparametrik aditif dengan pendekatan estimator polinomial dan menerapkannya pada data. Dalam skripsi ini digunakan algoritma *local scoring*. Berdasarkan pendekatan estimator polinomial lokal didapatkan bentuk estimasi model regresi logistik nonparametrik aditif berdasarkan estimator polinomial lokal sebagai berikut :

$$\begin{aligned} \log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) &= \sum_{j=1}^p \hat{m}_j(x_{ji}) \\ &= \sum_{j=1}^p X_j (X_j^T B X_j)^{-1} X_j^T B (Z - \sum_{k \neq j}^p m_k(X_k)) \end{aligned}$$

Dari hasil penerapan model regresi logistik nonparametrik aditif pada data jantung koroner diperoleh nilai  $MSE = 0.1112123$  dan model regresinya adalah :

$$\begin{aligned} \log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) &= \sum_{j=1}^3 \hat{m}_j(x_{ji}) \\ &= -13.93534625 + 0.07762493X_1 + 0.008716451X_2 + 0.2198283X_3 \end{aligned}$$

Berdasarkan model diatas, semakin bertambah usia, kolesterol, dan BMI (*Body Mass Index*) akan meningkatkan peluang terjadinya serangan jantung koroner. Dari model dapat dilihat pula bahwa variabel yang dominan adalah variabel BMI ( $X_3$ ). Hal ini dikarenakan variabel BMI memiliki nilai koefisien yang terbesar, yaitu 0.2198283.

**Kata kunci :** Regresi logistik, model aditif, estimator polinomial lokal, algoritma *local scoring*.

Siti Amanah, 2008. **Estimation of Logistic Regression Nonparametric Additive Models with Local Polynomial Estimator**. This *skripsi* under guidance of Nur Chamidah, S.Si and Ir.Elly Ana, M.Si, Mathematics Department. Faculty of Science and Technology. Airlangga University.

## ABSTRACT

Logistic regression is regression analysis to describe the relationship between respon variable  $Y$  and some predictor variables  $X$ , where  $E(Y|X)$  in  $[0, 1]$ . Logistic regression model can be shown as follows :

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \sum_{j=1}^p m_j(x_{ji}), \quad i = 1, 2, \dots, n$$

where  $\log\left(\frac{\mu_i}{1-\mu_i}\right)$  is a *logit function*.

The purpose of this *skripsi* is to estimate the function in logistic regression nonparametric additive model by approaching with local polynomial estimator and apply it on the coronneria heart disease data. In this *skripsi* is using local scoring algorithm. Based on local polynomial estimator is obtained estimation of the regression logistic nonparametric additive model as follows :

$$\begin{aligned} \log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) &= \sum_{j=1}^p \hat{m}_j(x_{ji}) \\ &= \sum_{j=1}^p X_j (X_j^T B X_j)^{-1} X_j^T B (Z - \sum_{k \neq j}^p m_k(X_k)) \end{aligned}$$

By applying the regression logistic nonparametric additive model on the coronneria heart disease data are obtained  $MSE$  value 0.1112123 and got the regression model

$$\begin{aligned} \log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) &= \sum_{j=1}^3 \hat{m}_j(x_{ji}) \\ &= -13.93534625 + 0.07762493X_1 + 0.008716451X_2 + 0.2198283X_3 \end{aligned}$$

Based on the model, if age, collesterol, and BMI (Body Mass Index) increase, probability of coronneria heart disease attack increase too. We can see that BMI ( $X_3$ ) is a dominant variable. It is because BMI variable has highest coefficient value, that is 0.2198283.

**Keyword :** Logistic regression, additive model, local polynomial estimator, and local scoring algorithm.