Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (22-31)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Hybrid Logistic Super Newton Model for Predicting Small Sample Size Data

Nurmalitasari[1*], Zalizah Awang Long[2], and Nurchim[3]

[1]Information System, Faculty of Computer Science, Duta Bangsa University Surakarta
[2]Malaysian Institute Information Technology, Kuala Lumpur University
[3]Informatics Engineering, Faculty of Computer Science, Duta Bangsa University Surakarta
[1,3]Jl. Bhayangkara No.55, Surakarta, Central Java, Indonesia
[2]Jl.Sultan Ismail, Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

Article:
Accepted: January 15, 2025
Revised: September 02, 2024
Issued: April 30, 2025

***Correspondence Address:**
nurmalitasari@udb.ac.id

ABSTRACT

Logistic regression is a model commonly used for predicting data with large sample sizes. However, in real-world scenarios, many cases involve small datasets that need to be addressed using logistic regression. The aim of this research is to develop a hybrid logistic regression model to address issues with small sample sizes by combining the Newton Raphson and Super Cubic methods. This hybrid model is applied to predict student dropout at Universitas Duta Bangsa Surakarta. The performance of the hybrid model is evaluated using two main metrics: the convergence of the parameter approximation to measure the precision of parameter estimation, and the ROC curve to assess prediction accuracy. Experimental results show that the Hybrid Logistic Super Newton model outperforms the logistic regression Newton Raphson model, requiring only three iterations to converge, thus improving computational efficiency. Moreover, this model achieves higher accuracy, with an AUC of 0.8833. These findings suggest that the developed model has the potential to be applied in various fields, such as healthcare, finance, and others, offering an effective solution for accurate, real-time predictive analytics. Further research could focus on optimizing the model's computational efficiency and exploring its application in other domains with small dataset challenges, such as healthcare and finance.

Keywords : *logistic regression; newton raphson; super cubic; small sample size data; prediction model.*

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (22-31)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

1. INTRODUCTION

Logistic regression is a non-linear model based on independent variables with categorical or numerical attributes that is used to model binary dependent variables. Besides that, each independent variable is non-linear. Several assumptions need to be considered while using logistic regression. This includes the binary dependent variable, observations that are independent of each other and a large sample size [1]. In real-world scenarios, it is common to encounter cases with small datasets that need to be addressed using logistic regression. According to Hosmer et al. [1] and Abu Zohair [2], in the prediction model, small sample problems can make the data-driven model unable to achieve better performance and accuracy. Fong et al. [3] proposed fine-tuning the prediction model parameters as a potential solution to this issue.

Research on estimating logistic regression parameters typically uses Maximum Likelihood [1], [4], [5]. Maximum Likelihood measures are parameter estimation methods commonly applied to linear models, whereas logistic regression is inherently a non-linear model. When this parameter estimation method is applied to a non-linear model, the solution system becomes challenging to derive algebraically [6]. To address this challenge and obtain an accurate solution, iterative methods are commonly employed. The Newton Raphson method is an efficient iterative technique for finding the roots of non-linear equations, requiring relatively few iterations [7]–[11]. It has been widely used in research for logistic regression parameter estimation due to its ability to converge quickly to an optimal solution [12]. However, the Newton Raphson method faces a significant drawback: it relies on accurately identifying the differential of the calculated function, which can be difficult in complex scenarios. This limitation can lead to challenges in estimating parameter values and determining root values effectively [13], [14].

The super cubic proposed by Darvishi & Barati [15] is a method used to solve non-linear equations based on the Adomian decomposition (AD) method. The AD method has the advantage of being able to solve complex differential problems [15]–[21]. To overcome the difficulties in Newton Raphson, this study will modify Newton Raphson with Super Cubic to determine the parameters of a Logistic Regression model. The combination of newton raphson with super cubic is expected to have high performance for estimating the parameters of the logistic regression model.

The objective of this research is to develop a new hybrid model by combining the basic Logistic Regression model with Newton Raphson with Super Cubic methods. This hybrid model will be applied to dropout data from students of Universitas Duta Bangsa Surakarta, which has the characteristics of a small dataset, to evaluate its performance. The accuracy of the proposed model will be assessed based on two criteria: the convergence of the approximation to evaluate the precision of parameter estimation [22], and the ROC curve to assess the accuracy of the prediction model [23], [24]. In the results, the levels of precision of the new model and the previous model (Regresi Logistic Newton Raphson) will be compared.

2. METHODS

The hybrid model is developed by combining logistic regression, Newton Raphson, and the super cubic method. Analytical evidence is presented in theoretical and numerical analysis. The prediction model development process is illustrated in Pseudocode 1 and 2. The stages of the Hybrid Logistic Super Newton model can be seen in Figure 1.

Pseudocode 1: Logistic Super Newton Model

```
INPUT: Pre-processed data training, Pre-processed data testing
OUTPUT: Best Prediction
ALGORITHM:
        Set i ← 1
        While i ≤ length of data training do
            Set φ(x)_i ← Training data instance
            m ← length of independent variable
            l(α) ← Compute log-likelihood (x_im)
            W(α) ← Compute Gradient (l(α))
            Z(α) ← Compute Hessian (W(α))
            α ← Compute Estimation Parameter (W(α), Z(α))
            i ← i + 1
        END WHILE
(Call Pseudocode 2)
```

Pseudocode 2: Parameter estimation

```
INPUT: α_0 = initial guess of α; ε = tolerance error; t = iteration
OUTPUT: α_(t+1)
ALGORITHM:
Set t ← 0
While ‖α_(t+1) - α_t‖ > ε do
        β_t ← Compute Newton Raphson's (α_t)
        γ_t ← Compute Super Cubic (β_t, α_t)
        α_(t+1) ← Compute Super Newton (γ_t)
        t ← t + 1
END WHILE
```

```
        Set i ← 1
        While i ≤ length of testing data do
            φ(x)_i ← Compute sigmoid (y_i)
            If φ(x)_i ≥ 0.5 then
                y_i ← 1
            Else
                y_i ← 0
            i ← i + 1
        END WHILE
```

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (22-31)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

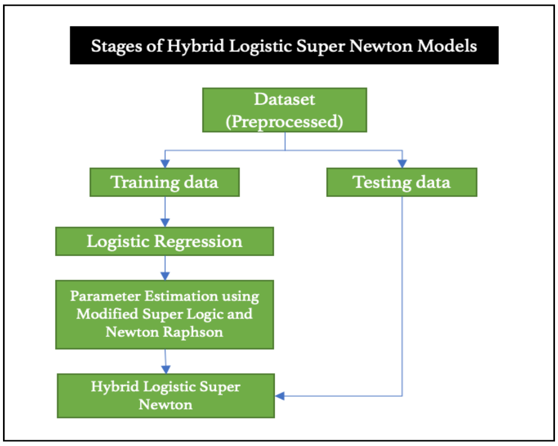

Figure 1. *Stages of hybrid logistic super newton model*

The data used in this research consists of 329 students, including those who graduated and those who dropped out from Universitas Duta Bangsa Surakarta in 2023, with variables based on previous findings by Nurmalitasari [1], as shown in Table 1. MATLAB was chosen for implementation due to its efficiency in numerical computing, particularly for matrix operations and iterative methods, although the approach can also be adapted to other platforms such as Python or R. Data pre-processing included mean and mode imputation to handle missing values and outlier management using the interquartile range (IQR) method to ensure the quality and reliability of the dataset for accurate modelling.

Table 1. *The variables used in the research*

Variable	Variable Construct Value
Dropout Indicator	0 = Dropout 1 = Graduate
Cumulative Grade Point Average (CGPA)	1 = 0.00 - 1.99 2 = 2.00 - 2.75 3 = 2.76 - 3.50 4 = 3.51 - 4.00
Individual Income	1 = No income 2 = Less than IDR 1,000,000 3 = IDR 1,000,001 - IDR 2,000,000 4 = IDR 2,000,001 - IDR 5,000,000 5 = IDR 5,000,001 - IDR 20,000,000 6 = More than IDR 20,000,000
Parental Income	1 = No income 2 = Less than IDR 1,000,000 3 = IDR 1,000,001 - IDR 2,000,000 4 = IDR 2,000,001 - IDR 5,000,000 5 = IDR 5,000,001 - IDR 20,000,000 6 = More than IDR 20,000,000

Table 1 continued…

Variable	Variable Construct Value
Marital Status	1 = Single 2 = Married 3 = Divorced
Student Employment Status	1 = Studying while working 2 = Studying without working
Interest in Study Program	1 = Very interested in the study program 2 = Interested in the study program 3 = Neutral about the study program 4 = Not interested in the study program 5 = Very disinterested in the study program
Relationship with lecturers/supervisors	1 = Very poor relationship 2 = Poor relationship 3 = Neutral relationship 4 = Good relationship 5 = Very good relationship
Satisfaction with lecturer quality	1 = Very dissatisfied 2 = Dissatisfied 3 = Neutral 4 = Satisfied 5 = Very satisfied

2.1. Logistic Regression

Logistic regression with parameter $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_M)^T$ with $x_{i0} = 1$ is provided by

$$\varphi_i(x) = \frac{e^{\sum_{m=0}^{M} \alpha_m x_{im}}}{1 + e^{\sum_{m=0}^{M} \alpha_m x_{im}}} \quad (1)$$

The logit transformation of $\varphi_i(x)$ is given by

$$logit(\varphi_i(x)) = ln\left(\frac{\varphi_i(x)}{1 - \varphi_i(x)}\right) \quad (2)$$

Equations (1) and (2) can be formulated

$$logit(\varphi_i(x)) = X^T \alpha$$

X is a matrix where each row corresponds to an observation, and each column corresponds to a predictor variable (including the intercept). For i-th observation, x_i is the row vector representing the values of the independent variables. The transpose X^T changes the rows of X into columns. If a sample of S-independent observation $\{(y_i, x_i)\}_{i=1,2,\dots,S} \in [\{0,1\} \times R^{M+1}]^S$ where y_i a represents a binary outcome, and x_i is the number of independent variables for the i^{th} subject. The likelihood function $L(\alpha)$is defined as

$$\boldsymbol{L(\alpha) = \prod_{i=1}^{S} \varphi_i(x)^{y_i} \left(1 - \varphi_i(x)\right)^{s_i - y_i}} \quad (3)$$

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

Equation (1) is substituted into Equation (3), it forms

$$\boldsymbol{L(\alpha)=\prod_{i=1}^{S} e^{y_i\sum_{m=0}^{M}\alpha_m x_{im}}\left(1+e^{\sum_{m=0}^{M}\alpha_m x_{im}}\right)^{-s_i}} \quad (4)$$

The function of log-likelihood $(l(\alpha))$ derived from equation (4) can be expressed as follows:

$$\boldsymbol{l(\alpha)=\sum_{i=1}^{S}\left[\left(y_i\sum_{m=0}^{M}\alpha_m x_{im}\right)-s_i\log\left(1+e^{\sum_{m=0}^{M}\alpha_m x_{im}}\right)\right]}. \quad (5)$$

Equation (5) is also called the cos function. The critical point of the log-likelihood function can be determined by taking its derivative with respect to the parameter and setting it equal to zero.

$$\boldsymbol{\frac{\partial l(\alpha)}{\partial \alpha_m}=\sum_{i=1}^{S}[(y_i x_{im})-s_i\varphi_i(x)x_{im}]=0} \quad (6)$$

Equation (6) can be rewritten in matrix form as follows:

$$\boldsymbol{W(\alpha)=l'^{(\alpha)}=X^T\alpha,}$$

where $l'(\alpha)$ is a column vector of length $M+1$ whose elements are $\frac{\partial l(\alpha)}{\partial \alpha_m}$, $m=0,1,\dots,M$, likewise ρ is a column vector of length S with elements $\rho_i=(s_1\varphi_1,s_2\varphi_2,\dots,s_S\varphi_S)^T$ and X^T is a $(M+1)\times S$ matrix. Parameter α can be found using Maximum Likelihood by setting each of the $M+1$ equations in (6) equal to zero and solving for each α_m. Equation (6) is also called the gradient function. The general from of matrix of second partial derivatives is:

$$\boldsymbol{\frac{\partial^2 l(\alpha)}{\partial \alpha_m \partial \alpha_{m'}}=-\sum_{i=1}^{N} s_i\varphi_i(x)x_{im}\left(1-\varphi_i(x)\right)x_{im'},} \quad (7)$$

with $m=0,1,2,\dots,M$ dan $m'=0,1,2,\dots,M$. Equation (7) can be expressed in term of matrix multiplication Z.

$$\boldsymbol{Z(\alpha)=-X^T DX.}$$

The Hessian matrix, denoted as $Z(\alpha)$, includes D, an $S\times S$ diagonal matrix. By setting equation (6) equal to zero, we obtain a system of $M+1$ non-linear equations with $M+1$ unknown variables. Solving this system yields a vector containing elements α_m. After confirming that the matrix of second-order partial derivatives is negative definite, ensuring that the solution represents a global maximum rather than a local one, it can conclude that this vector provides the parameter estimates that maximize the likelihood of the observed data.

2.2. Newton Raphson

The Newton Raphson method solves non-linear equations using a single starting point approach and approaches it by paying attention to the slope or gradient. Consider the system of non-linear equations $W(\alpha)=0$ for solving the system, if an initial guess $\alpha^{(0)}$ is computed from the observed data, then

$$\boldsymbol{\alpha_{t+1}=\alpha_t-\frac{W(\alpha_t)}{[Z(\alpha_t)]}} \quad (8)$$

Here $W'(\alpha_t)$ refers to the derivative $W(\alpha_t)$. Equation (8) is called the Newton Raphson formula for solving a system of nonlinear equations of the from $W(\alpha_t)=0$. An initial guess for the root is computed to solve such a system using (8). Assume that α_0 is that guess. The next value α_1 will be determined. This iterative process will continue until the desirable root is obtained; it is with $\|\alpha_{t+1}-\alpha_t\|<\varepsilon$ for some specific value ε, $t=0,1,\dots$. The Newton Raphson method has the advantage is faster convergence in determining the roots of the non-linear equation.

2.3. Super Cubic

Darvishi & Barati [15] present super cubic convergence methods to solve systems of non-linear equations. The super cubic method is based on Adomian decomposition method Newton. The iteration scheme of the super cubic is as follows:

$$\boldsymbol{\alpha_{t+1}=\alpha_t-\frac{W(\alpha_t)}{Z(\alpha_t)}-\frac{W(\beta_t)}{Z(\beta_t)}} \quad (9)$$

with

$$\boldsymbol{\beta_t=\alpha_t-\frac{W(\alpha_t)}{Z(\alpha_t)}}$$

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (22-31)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

2.4. The Proposed Model

The Newton-Raphson method is an iterative method that converges quadratically. In this study, modifying the Newton Raphson algorithm with the super cubic proposed by Darvishi & Barati [15] is suggested to improve convergence in determining logistic regression parameters.

A three-step iterative method is employed to estimate logistic regression parameters from Newton Raphson's equation (8) and super cubic equation (9). Our proposed formula can be seen in equation (10).

$$\begin{cases} \boldsymbol{\beta_t = \alpha_t - \frac{W(\alpha_t)}{Z(\alpha_t)}} \\ \boldsymbol{\gamma_t = \alpha_t - \frac{W(\alpha_t)}{Z(\alpha_t)} - \frac{W(\beta_t)}{Z(\beta_t)}} \\ \boldsymbol{\alpha_{t+1} = \gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)} - \frac{W\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right)}{Z\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right)}} \end{cases} \quad (10)$$

Where $\alpha_t = (\alpha_{1t}, \alpha_{2t}, \dots, \alpha_{mt})^T$, $\beta_t = (\beta_{1t}, \beta_{2t}, \dots, \beta_{mt})^T$, $\gamma_t = (\gamma_{1t}, \gamma_{2t}, \dots, \gamma_{mt})^T$, $Z(\gamma_t)$ is Jacobian matrix. As a result,

$$\boldsymbol{Z(\alpha) = \begin{pmatrix} \frac{\partial W_1}{\partial \alpha_1}, \frac{\partial W_1}{\partial \alpha_2}, \dots, \frac{\partial W_1}{\partial \alpha_m} \\ \frac{\partial W_2}{\partial \alpha_1}, \frac{\partial W_2}{\partial \alpha_2}, \dots, \frac{\partial W_2}{\partial \alpha_m} \\ \dots \\ \frac{\partial W_m}{\partial \alpha_1}, \frac{\partial W_m}{\partial \alpha_2}, \dots, \frac{\partial W_m}{\partial \alpha_m} \end{pmatrix}}$$

This proves that the method defined by (10) has ninth-order convergence. This combination of the Newton Raphson and Super Cubic methods is used to estimate the parameters of logistic regression. The proposed new model is named the Logistic Super Newton model.

The evaluation of the proposed model's accuracy will be based on two factors: the convergence of the approximation to measure the precision of parameter estimation [22], and the ROC curve to evaluate the prediction model's accuracy [23], [24]. The results will compare the precision levels of the new model with those of the previous model (Newton Raphson Logistic Regression).

3. RESULTS AND DISCUSSION

3.1. Result

The first step in constructing the hybrid logistic Super Newton model is to theoretically prove that equation (10) has a local order of convergence of at least nine, as demonstrated by the following error equation.

Theorem 1. *The iterative method (10) has local order of convergence at least nine with the following error equation*

$$\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)} - \frac{W\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right)}{Z\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right)} = \theta + 16e_2^8\varepsilon_t^9 + o\left(\|\varepsilon_t^{10}\|\right) \quad (11)$$

Proof: let θ be a simple zero of Z. As Z is a sufficiently differentiable function, by expanding $W(\alpha_t)$ and $Z(\alpha_t)$ about θ, it is shown that

$$W(\alpha_t) = Z(\theta)\left[\varepsilon_t + e_2\varepsilon_t^2 + e_3\varepsilon_t^3 + e_4\varepsilon_t^4 + e_5\varepsilon_t^5 + e_6\varepsilon_t^6 + e_7\varepsilon_t^7 + e_8\varepsilon_t^8\right] + o(\|\varepsilon_t^9\|) \quad (12)$$

$$Z(\alpha_t) = Z(\theta)\left[1 + e_2\varepsilon_t^1 + e_3\varepsilon_t^2 + e_4\varepsilon_t^3 + e_5\varepsilon_t^4 + e_6\varepsilon_t^5 + e_7\varepsilon_t^6 + e_8\varepsilon_t^7\right] + o\left(\|\varepsilon_t^8\|\right) \quad (13)$$

Where $e_t = \left(1/_{t!}\right)\frac{W_t(\theta)}{Z(\theta)}$, $t = 2,3, \dots$ and $\varepsilon_t = \alpha_t - \theta$. The square brackets are polynomials in terms of ε_t. Equations 12 and 13 can be used to calculate the following equation.

$$\gamma_t = \theta + 2e_2^2\varepsilon_t^2 + (7e_2e_3 - 9e_2^3)\varepsilon_t^4 + (6e_3^2 - 44e_3e_2^2 + 10e_2e_4 + 30e_2^4)\varepsilon_t^5 + (17e_3e_4 - 62e_4e_2^2 + 188e_3e_2^3 - 88e_2^5 - 70e_2e_3^2)\varepsilon_t^6 + o(\|\varepsilon_t^7\|) \quad (14)$$

$$\frac{W(\gamma_t)}{Z(\gamma_t)} = 2e_2^2\varepsilon_t^3 + (7e_2e_3 - 9e_2^3)\varepsilon_t^4 + (6e_3^2 - 44e_3e_2^2 + 10e_2e_4 + 30e_2^4)\varepsilon_t^5 + (17e_3e_4 - 92e_2^5 - 62e_4e_2^2 + 188e_3e_2^3 - 70e_2e_3^2)\varepsilon_t^6 + o(\|\varepsilon_t^7\|) \quad (15)$$

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (22-31)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

$$\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)} = \theta + 4e_2^5\varepsilon_t^6 + (28e_3e_2^4 - 36e_2^6)\varepsilon_t^7 + o(\|\varepsilon_t^8\|) \qquad (16)$$

With (16), it can be seen that

$$W\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right) = 4e_2^5\varepsilon_t^6 + (28e_3e_2^4 - 36e_2^6)\varepsilon_t^7 + o(\|\varepsilon_t^8\|) \qquad (17)$$

From (14), (16), and (17) it is concluded that

$$\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)} - \frac{W\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right)}{Z\left(\gamma_t - \frac{W(\gamma_t)}{Z(\gamma_t)}\right)} = \theta + 16e_2^8\varepsilon_t^9 + o(\|\varepsilon_t^{10}\|)$$

This shows the ninth-order convergence of the method. Hence, the Proof is completed.

The next step is to validate the new model through numerical analysis. This validation is conducted using Pseudocode 1.a and 1.b, applied to a small dataset of student dropouts from Universitas Duta Bangsa Surakarta. Using an initial guess of $\alpha^0 = -1e-5$, with $Itermax = 150000$ and $\varepsilon = 1e-5$, using MATLAB version R2021a software, as shown in Figure 2.

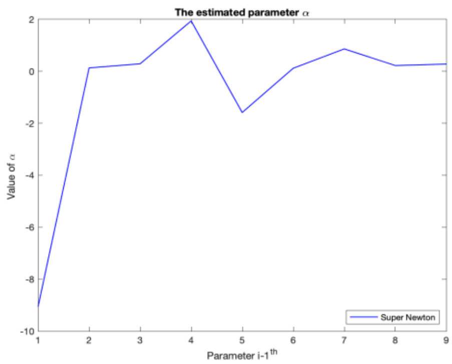

Figure 2. *The parameter estimation results*

The parameter values from Figure 1 are used to make predictions, and the results are presented in Figure 3 below.

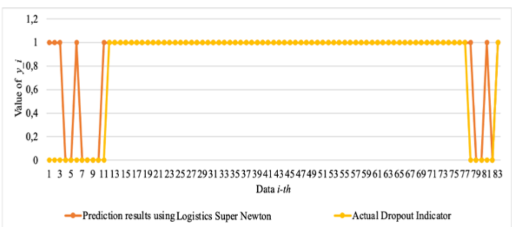

Figure 3. *Prediction results using hybrid logistics super newton*

Figure 3 illustrates the results of dropout predictions using the Logistics Super Newton model compared to actual data testing. From the Figure 3, it can be seen that there are seven different prediction results from 83 testing data. This means that 91.57% of the Hybrid Logistics Super Newton model prediction results are correct. This can be interpreted that the Logistic Super Newton model can be used to predict dropout students in a small sample size [25].

Assessing the accuracy of the prediction model is crucial because the accuracy of the model will determine the quality of predictions produced. There are two benchmarks for determining the prediction accuracy of the Hybrid Logistic Super Newton model. First, the convergence of the approximation is used to evaluate the precision of the parameter estimation. Using initial guess $\alpha^0 = -1e-5$, with $\text{Itermax} = 150000$ and error value $\varepsilon = 1e-5$. The results of the analysis can be seen in Figure 4.

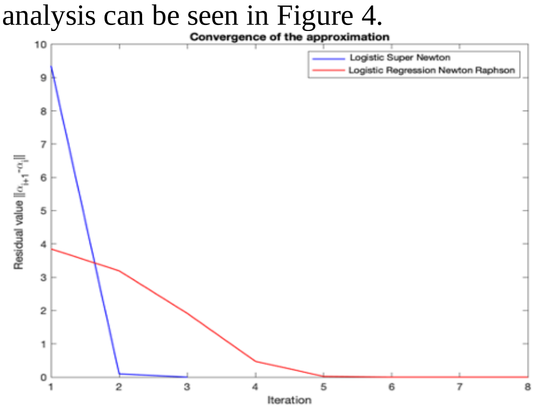

Figure 4. *Convergence of the approximation of the logistic super newton versus logistic regression newton-raphson*

According to Figure 4, the Hybrid Logistic Super Newton model has a higher convergence rate than the logistic regression newton Raphson model. This is evident from the number of parameter-finding iterations. The Hybrid Logistic Super Newton model, requires only three iterations to determine the actual parameters, whereas the Logistic Regression Newton Raphson model required eight iterations. This suggests that the propose model has a higher convergence rate than its predecessor.

Second, the Receiver Operating Characteristics (ROC) curve is used to assess the accuracy of the prediction model.

Jurnal Teknik Informatika Vol. 18 No. 1, April 2025 (22-31)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

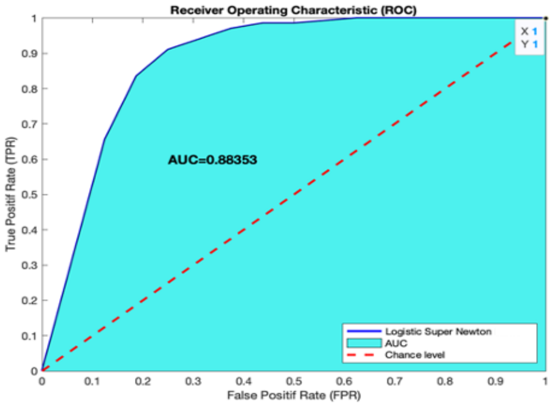

Figure 5. *ROC of the logistic super newton model*

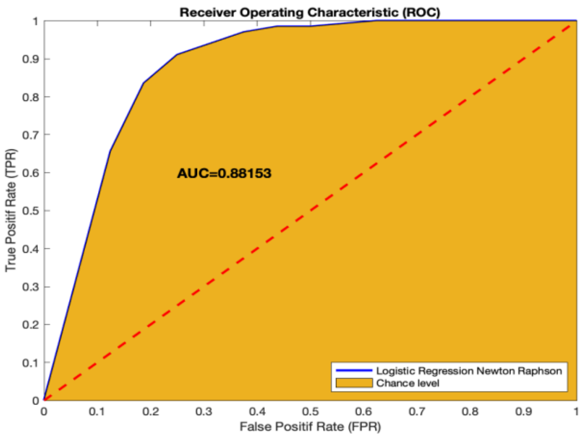

Figure 6. *ROC of the logistic regression newton raphson*

The ROC curves in Figures 5 and 6 compare the performance of two models: the Hybrid Logistic Super Newton and the Logistic Regression Newton Raphson. The Hybrid Logistic Super Newton model achieves an Area Under the Curve (AUC) of 0.8833, demonstrating slightly better predictive accuracy than the Logistic Regression Newton Raphson model, which has an AUC of 0.8815. Both curves are well above the diagonal reference line, indicating strong classification performance for both models. The slightly better performance of the Hybrid Logistic Super Newton model can be attributed to the integration of the Super Cubic method with the Newton-Raphson approach. This study leverages the strengths of the Super Cubic method to address the limitations of the Newton-Raphson method, particularly its reliance on accurately determining the differential of the calculated function, which can be challenging for non-linear models. By incorporating the Super Cubic method into the Newton-Raphson framework, the hybrid model achieves more precise and stable parameter estimation, even for small datasets. This hybrid approach enhances the efficiency and accuracy of the iterative process, enabling the model to better capture the relationships between predictors and outcomes. As a result, the Hybrid Logistic Super Newton model demonstrates improved classification performance and predictive accuracy, making it a robust solution for logistic regression models with small sample sizes.

3.2. Discussion

The experimental results demonstrate the superiority of the new model, the Hybrid Logistic Super Newton, compared to the older model, the Logistic Regression Newton Raphson. The faster convergence of the Hybrid Logistic Super Newton model indicates its efficiency in solving nonlinear equations with fewer iterations. In numerical methods, a faster convergence rate is a key indicator of an effective algorithm, as it reduces computational time without compromising accuracy [26]–[29]. This characteristic is particularly crucial for real-time prediction systems, where fast processing is essential to provide reliable and timely results [30]–[32].

The ROC curves in Figures 4 and 5 further reinforce the superior performance of the Logistic Super Newton model. With an AUC of 0.8833, the new model slightly outperforms the Logistic Regression Newton Raphson model, which has an AUC of 0.8815. Both models exhibit strong classification capabilities, as evidenced by their curves being significantly above the diagonal reference line. However, the higher AUC of the Logistic Super Newton model highlights its better predictive accuracy, solidifying its role as a more robust and effective logistic regression model. This improvement is particularly beneficial for datasets requiring high precision in classification tasks [24]. One of the main advantages of using ROC curves in this analysis is their ability to evaluate model performance at all possible prediction thresholds rather than relying solely on average accuracy. This comprehensive evaluation method enables a deeper understanding of the trade-offs between sensitivity and specificity for each model.

The findings indicate that the Hybrid Logistic Super Newton approach not only enhances prediction accuracy but also improves system performance by enabling faster analysis and more efficient delivery of predictive information. These results strongly support the practical implementation of the Hybrid Logistic Super Newton model in predictive systems, where both speed and accuracy are critically important.

While the Hybrid Logistic Super Newton model demonstrates clear advantages in terms of convergence speed and predictive accuracy, there are certain limitations that must be addressed. One significant challenge lies in scaling the model to larger datasets, where the computational demands may increase substantially due to the iterative nature of the Newton-Raphson and Super Cubic methods. While the current implementation is efficient for small datasets, the complexity of the hybrid approach may require optimization or parallel computing techniques to maintain its efficiency in larger datasets or high-dimensional data scenarios. Additionally, adapting the model to other domains with vastly different data characteristics, such as unbalanced datasets or those with categorical variables, may necessitate further modifications to ensure its robustness and generalizability. Despite these challenges, the broader implications of the model are promising, as it provides a pathway for developing real-time predictive systems in fields such as healthcare, finance, and education. Future research should focus on refining the model's computational framework, exploring domain-specific adaptations, and investigating its integration with advanced pre-processing techniques or alternative optimization algorithms to enhance its applicability across a wider range of predictive problems.

CONCLUSION

This research successfully developed a new hybrid model that combines the Logistic Regression model with the Newton-Raphson and Super Cubic methods, applied to dropout data from students at Universitas Duta Bangsa Surakarta. The experimental results show that the Hybrid Logistic Super Newton model outperforms the Logistic Regression Newton Raphson model in terms of both convergence and prediction accuracy. Specifically, the hybrid model requires only three iterations to find accurate parameters, compared to eight iterations required by the previous model, demonstrating its higher computational efficiency in solving non-linear equations. Moreover, in terms of predictive accuracy, the hybrid model achieves an AUC of 0.8833, which is slightly higher than the AUC of the previous model (0.8815), indicating its improved ability to classify outcomes accurately.

The Hybrid Logistic Super Newton model proves to be an effective solution for handling small sample datasets, addressing the challenges commonly faced by traditional logistic regression models in such scenarios. This model has the potential to be applied in various fields requiring accurate and efficient predictions, such as risk prediction systems in education, healthcare, and finance. Future research should focus on refining the hybrid approach to further enhance its computational efficiency and adaptability to small datasets. Exploring improvements in data processing techniques and model algorithms may also pave the way for more advanced and reliable predictive solutions tailored for small sample scenarios.

REFERENCES

[1] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression Analysis.*, Third Edit. John Wiley & Sons, Inc., 2013.

[2] L. M. Abu Zohair, 'Prediction of Student's performance by modelling small dataset size', *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, 2019.

[3] S. J. Fong, G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, 'Finding an accurate early forecasting model from small dataset: A Case of 2019-nCoV novel coronavirus outbreak', *arXiv*, vol. 6, pp. 132–140, 2020.

[4] J. C. Stoltzfus, 'Logistic regression: A brief primer', *Acad. Emerg. Med.*, vol. 18, no. 10, pp. 1099–1104, 2011.

[5] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Second Edi. Canada: John Wiley & Sons, Inc., 2000.

[6] A. Albert and J. A. Anderson, 'On the existence of maximum likelihood estimates in logistic regression models', *Biometrika*, vol. 71, no. 1, pp. 1–10, 1984.

[7] I. Azure, G. Aloliga, and L. Doabil, 'Comparative Study of Numerical Methods for Solving Non-linear Equations Using Manual Computation', *Math. Lett.*, vol. 5, no. 4, p. 41, 2019.

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i1.43929

[8] A. W. Nwry, Kareem, I. Hemin Muheddin, R. B. Brahim, and M. M. Sundws, 'Comparison Between Bisection, Newton and Secant Methods for determining the root of the Non-Linear equation using MATLAB', *Turkish J. Comput. Math. Educ.*, vol. 12, no. 14, pp. 1115–1122, 2021.

[9] G. Thakur and J. K. Saini, 'Comparative Study of Iterative Methods for Solving Non-Linear Equations', *J. Univ. Shanghai Sci. Technol.*, vol. 23, no. 7, pp. 858–866, 2021.

[10] S. Akram and Q. Ann, 'Newton-Raphson Method', *Financ. Simul. Model. Excel*, vol. 6, no. 7, pp. 183–185, 2015.

[11] A. R. Srivastava, S. Sharma, and V. V Mehtre, 'Analysis of Real Roots by Newton Raphson Method and Secant Method for finding Mathematical Roots Evaluation', vol. 10, no. 11, pp. 48–54, 2022.

[12] C. Shalizi, 'Logistic Regression and Newton ' s Method', vol. 7, no. November, pp. 1–12, 2009.

[13] H. Altenbach and A. Öchsner, Eds., 'Newton-Raphson Method BT - Encyclopedia of Continuum Mechanics', Berlin, Heidelberg: Springer Berlin Heidelberg, 2020, p. 1810.

[14] L. Fang and L. Pang, 'Improved Newton-Raphson Methods for Solving Nonlinear Equations', *J. Adv. Math.*, vol. 13, no. 5, pp. 7403–7407, 2017.

[15] M. T. Darvishi and A. Barati, 'Super cubic iterative methods to solve systems of nonlinear equations', *Appl. Math. Comput.*, vol. 188, no. 2, pp. 1678–1685, 2007.

[16] D. A. Maturi and H. M. Malaikah, 'The Adomian Decomposition Method for Solving Nonlinear Partial Differential Equation Using Maple', *Adv. Pure Math.*, vol. 11, no. 06, pp. 595–603, 2021.

[17] M. Dehghan and M. Tatari, 'The use of convergence of highly nonlinear elasticity problems', *Comput. Methods Appl. Mech. Eng.*, vol. 337, pp. 110–127, 2018.

[27] S. V. Archontoulis and F. E. Miguez, 'Nonlinear regression models and applications in agricultural research', *Agron. J.*, vol. 107, no. 2, pp. 786–798, 2015.

[28] V. S. Ozsoy and H. Hasan Orkcu, Adomian decomposition method for solving problems in calculus of variations', *Math. Probl. Eng.*, vol. 2006, no. March 2005, pp. 1–12, 2006.

[18] W. Li and Y. Pang, 'Application of Adomian decomposition method to nonlinear systems', *Adv. Differ. Equations*, vol. 2020, no. 1, 2020.

[19] Oswaldo Gonzalez Gaxiola, 'Solution of nonlinear partial differential Equations by adomian decomposition method', *Stud. Eng. Exact Sci.*, vol. 3, no. 1, pp. 61–78, 2022.

[20] A. K. Jimoh and A. O. Oyedeji, 'On Adomian decomposition method for solving nonlinear ordinary differential equations of variable coefficients', *Open J. Math. Sci.*, vol. 4, no. 1, pp. 476–484, 2020.

[21] M. Javidi and A. Golbabai, 'Adomian Decomposition Method for Approximating', *Appl. Math. Sci.*, vol. 1, no. 5, pp. 219–225, 2007.

[22] L. Mochurad and N. Boyko, 'Solving systems of nonlinear equations on multi-core processors', *Adv. Intell. Syst. Comput.*, vol. 1080 AISC, pp. 90–106, 2020.

[23] N. R. Cook, 'Use and misuse of the receiver operating characteristic curve in risk prediction', *Circulation*, vol. 115, no. 7, pp. 928–935, 2007.

[24] N. A. Obuchowski and J. A. Bullen, 'Receiver operating characteristic (ROC) curves: Review of methods with applications in diagnostic medicine', *Phys. Med. Biol.*, vol. 63, no. 7, 2018.

[25] B. R. Cuji Chacha, W. L. Gavilanes López, V. X. Vicente Guerrero, and W. G. Villacis Villacis, 'Student Dropout Model Based on Logistic Regression', *Commun. Comput. Inf. Sci.*, vol. 1194 CCIS, pp. 321–333, 2020.

[26] Y. Mei, D. E. Hurtado, S. Pant, and A. Aggarwal, 'On improving the numerical

'Estimating the parameters of nonlinear regression models through particle swarm optimization', *Gazi Univ. J. Sci.*, vol. 29, no. 1, pp. 187–199, 2016.

[29] A. Ben-Israel, 'A Newton-Raphson method for the solution of systems of equations', *J. Math. Anal. Appl.*, vol. 15, no. 2, pp. 243–252, 1966.

[30] S. R. Kimbleton, 'A heuristic approach to

computer systems performance improvement, I: A fast performance prediction tool', *AFIPS Conf. Proc. - 1975 Natl. Comput. Conf. AFIPS 1975*, pp. 839–846, 1975.

[31] G. S. Tyson and T. M. Austin, 'Memory renaming: Fast, early and accurate processing of memory communication', *Int. J. Parallel Program.*, vol. 27, no. 5, pp. 357–380, 1999.

[32] I. H. Sarker, 'Machine Learning: Algorithms, Real-World Applications and Research Directions', *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021