Jurnal Matematika Thales (JMT)
Vol. 07, No. 01 (2025), pp. 54–63.

Optimizing Linear Models with Deep Neural Networks: A Case Study on Gambling Data Across Indonesian Provinces

Astri Ayu Nastiti[1*], Abdurakhman[2]

[1,2]**Department of Mathematics, University of Gadjah Mada, Yogyakarta, Indonesia**
[1]astriayunastiti@mail.ugm.ac.id, [2]rachmanstat@ugm.ac.id

Abstract. Linear regression analysis is a common method that are free to vary and are subject to error. In this study we used hybrid of linear regression and its family to Deep Neural Network (DNN) to fill these gaps. In this paper analyze the phenomenon of gambling in Indonesia in 2018. Results show that the hybrid model is significantly superior to the single model, with the hybrid linear model reducing RMSE by 15.9% and MAPE by 16.2% compared to the single linear model. The hybrid ridge model showed small but consistent improvements in RMSE and MAPE. The most notable improvement was seen in the hybrid lasso model which reduced RMSE by 34.1% and MAPE by 47.1% over the single lasso model. The hybrid elastic net model also showed improved performance with a decrease in RMSE by 16.9% and MAPE by 18.3%. In conclusion, the integration of traditional regression methods with DNN in this hybrid model offers a significant improvement in prediction accuracy, making it a more effective and efficient tool in the analysis of gambling phenomena.

Keywords: Linear regression, ridge regression, LASSO regression, elastic net regression, Deep Neural Network, Hybrid Model.

1. INTRODUCTION

Gambling is a social phenomenon that can have a negative impact on society, both in economic and social terms. Several studies reveal the effects of gambling including divorce [1], increased anxiety levels [2], affecting not only adults but also children [3]. Therefore, understanding the characteristics of the community and the factors that contribute to the incidence rate of gambling is an important step towards formulating effective policies to address it.

*Corresponding author
2020 Mathematics Subject Classification: 62J05, 62J07, 92B20
Received: 25-06-2025, accepted: 07-08-2025.

This paper uses population data by province in Indonesia to analyze the relationship between various socio-economic factors and the incidence of gambling. Given the relatively small amount of data, linear regression was chosen as the main analysis method. Linear regression is an appropriate choice because it is simple and effective in handling small datasets without requiring complex training and testing processes as in machine learning techniques.

In regression analysis, multiple linear regression is often used to understand the relationship between independent variables and dependent variables. However, to overcome the problem of multicollinearity and to improve the accuracy of the model, extensions of simple linear regression such as Ridge, Lasso, and Elastic Net regression have been introduced [4].

Ridge regression addresses multicollinearity by adding an L_2 penalty to the coefficients, thus preventing the coefficients from becoming too large [5]. Lasso regression introduces the L_1 penalty, which not only addresses multicollinearity but also performs feature selection by reducing some coefficients to zero, thus simplifying the model (Vidaurre et al., 2013). Elastic Net Regression is a combination of L_1 and L_2 penalties, which allows handling multicollinearity while performing feature selection, providing more flexibility in building robust and interpretable predictive models [6]. These three methods are known as an "extended linear regression family" that offers more advanced solutions for complex data analysis and potentially better predictive performance.

As is well known, linear regression and extended linear regression approaches can give results that differ from the actual data [7]. In some cases, simple linear regression models may fail to capture the complexity of the data patterns, especially when multicollinearity is present or when the relationship between variables is not strictly linear. Furthermore, to overcome the weaknesses of these traditional regression approaches, deep-linear regression hybrid artificial neural networks are used. This hybrid approach combines the analytical power of linear regression with the capability of artificial neural networks to recognize non-linear patterns and complex interactions in the data.

The purpose of this research is to explore and maximize the potential of linear regression and its families (ridge, lasso, and elastic net) in producing accurate predictions by combining them with deep learning networks. This research aims to identify the advantages of the hybrid approach in reducing prediction error compared to the use of a single model, as well as to develop predictive models that are more robust and efficient in handling data complexity. Thus, this research hopes to make a significant contribution to more accurate and reliable predictive modeling through the integration of traditional regression methods and deep learning technology.

2. LITERATURE REVIEW

2.1. **Linear Regression.** Linear regression is an equation model that explains the correlation of one response variable (Y) with two or more predictor variables ($X_1, X_2, \ldots, X_p$). In addition, it is used to determine the direction of the relationship between the response variable and the predictor variables. The relationship between the response variable and the predictor variables is expressed as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

with:

$Y = n \times 1$ vector of dependent variables,

$X = n \times (p-1)$ matrix of independent variables,

$\boldsymbol{\varepsilon} = n \times 1$ vector of independent normal random variables with expectation $E(\boldsymbol{\varepsilon}) = 0$ and variance-covariance matrix $\sigma^2(\varepsilon) = \sigma^2 I$.

According to Firdaus (2004) [8], the least squares method or also called the Ordinary Least Square (OLS) method is one of the most popular methods in estimating linear regression models that produce the minimum number of squared errors. This method was first used by Carl Friedrich Gauss in the calculation of astronomical problems. The practical advantages of this method increased after the development of electronic computers, the formulation of calculation techniques in matrix notation, and the connection of the least squares concept to statistics.

Definition 2.1. *Let $p \geq 1$ be a real number. The p-norm of vector $x = (x_1, x_2, \ldots, x_n)$ is*

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{(1/p)}. \tag{2}$$

For $p = 1$, we get the taxicab/manhattan norm, for $p = 2$ we get the Euclidean norm, and as p approaches ∞ the p-norm approaches the infinity norm [9].

From Equation (1) is obtained

$$\begin{aligned}
\boldsymbol{\varepsilon} &= \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \\
\mathbf{S}(\boldsymbol{\beta})_{OLS} &= \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2 \\
&= \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} \\
&= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= (\mathbf{Y}^\top - \mathbf{X}^\top \boldsymbol{\beta}^\top)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{Y}^\top \mathbf{Y} - (\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y})^\top - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{Y}^\top \mathbf{Y} - 2(\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y}) + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}
\end{aligned} \tag{3}$$

Next, a partial derivative of $\boldsymbol{\beta}$ is performed to obtain the minimum value of the equation:

$$\begin{aligned}\frac{\partial(\mathbf{S}(\boldsymbol{\beta})_{OLS})}{\partial(\boldsymbol{\beta})} &= \frac{\partial\left(\mathbf{Y}^\top\mathbf{Y} - 2(\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y}) + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}\right)}{\partial(\boldsymbol{\beta})} \\ &= 0 - 2\mathbf{X}^\top\mathbf{Y} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + (\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X})^\top \\ &= 2\mathbf{X}^\top\mathbf{Y} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}\end{aligned} \tag{4}$$

and then equating it to zero, we get:

$$\begin{aligned}0 &= 2\mathbf{X}^\top\mathbf{Y} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\ 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} &= 2\mathbf{X}^\top\mathbf{Y} \\ \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{Y} \\ \hat{\boldsymbol{\beta}}_{OLS} &= (\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{Y})\end{aligned} \tag{5}$$

Therefore, Equation (5) as the solution of the OLS method.

2.2. **Ridge Regression.** Ridge regression is the result of the least squares method with the addition of a bias value c to the correlation matrix and the variables are transformed using the centering and scaling method, the selection of the bias constant c is a very instrumental thing in Ridge regression [10]. The penalty in ridge with the following constraints:

$$\|\boldsymbol{\beta}\|_2 \leq t, t > 0$$

Loss function ridge regression is as follow:

$$\begin{aligned}\mathbf{S}(\boldsymbol{\beta})_R &= \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2 + c\|\boldsymbol{\beta}\|_2 \\ &= \boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon} + c\boldsymbol{\beta}^\top\boldsymbol{\beta} \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + c\boldsymbol{\beta}^\top\boldsymbol{\beta} \\ &= (\mathbf{Y}^\top - \mathbf{X}^\top\boldsymbol{\beta}^\top)(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + c\boldsymbol{\beta}^\top\boldsymbol{\beta} \\ &= \mathbf{Y}^\top\mathbf{Y} - \mathbf{X}^\top\boldsymbol{\beta}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + c\boldsymbol{\beta}^\top\boldsymbol{\beta} \\ &= \mathbf{Y}^\top\mathbf{Y} - 2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + c\boldsymbol{\beta}^\top\boldsymbol{\beta}\end{aligned} \tag{6}$$

Next, a partial derivative of $\boldsymbol{\beta}$ is performed to obtain the minimum value of the equation:

$$\begin{aligned}\frac{\partial(\mathbf{S}(\boldsymbol{\beta})_R)}{\partial(\boldsymbol{\beta})} &= \frac{\partial\left(\mathbf{Y}^\top\mathbf{Y} - 2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + c\boldsymbol{\beta}^\top\boldsymbol{\beta}\right)}{\partial(\boldsymbol{\beta})} \\ &= 0 - 2\mathbf{X}^\top\mathbf{Y} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + (\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X})^\top + 2c\boldsymbol{\beta} \\ &= -2\mathbf{X}^\top\mathbf{Y} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + 2c\boldsymbol{\beta} \\ &= -2\mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + 2c\boldsymbol{\beta}\end{aligned} \tag{7}$$

and then equating it to zero, we get:

$$\begin{aligned} 0 &= -2\mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + 2c\boldsymbol{\beta} \\ \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + c\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{Y} \\ (\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + c\mathbf{I})\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{Y} \\ \hat{\boldsymbol{\beta}}_R &= (\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + c\mathbf{I})^{-1}(\mathbf{X}^\top\mathbf{Y}) \end{aligned} \tag{8}$$

Therefore, Equation (8) as the solution of the ridge regression.

2.3. **LASSO Regression.** LASSO (Regression Least Absolute Shringkage and Selection Operator) is one of the shrinkage methods to overcome multicollinearity problems. The LASSO method is a method introduced by Tibshirani in 1996 [11] after the LAR (Least Angle Regression) method introduced by Effron in 2004 by changing the penalty in Ridge regression in L_1 regularization. This regularization is used to reduce overfitting by adding L_1 and L_2 penalty factors where L_1 regularization is called LASSO regression which uses L_1 penalty, an approach that penalizes the absolute size of the coefficients. Whereas L_2 regularization is called Ridge regression which uses an L_2 penalty, which is an approach that penalizes the squared size of the coefficients. LASSO aims to improve the estimation of simple linear regression. The penalty in LASSO with the following constraints:

$$\|\boldsymbol{\beta}\|_1 \leq t, t > 0$$

. The value of t above is a quantity that checks the amount of shrinkage in the LASSO coefficient estimates where $t \geq 0$. If the estimator $\hat{\boldsymbol{\beta}}$ is a least squares estimator and $t_0 = \|\boldsymbol{\beta}\|_1$, then values of $t < t_0$ will lead to solving classical regression with OLS estimators that shrink towards zero, and allow some coefficients to also shrink exactly towards zero. Loss function for LASSO regression is as follow:

$$\mathbf{S}(\boldsymbol{\beta})_{LASSO} = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \boldsymbol{\alpha}\|\boldsymbol{\beta}\|_1 \tag{9}$$

Unlike OLS estimation of linear regression in the equation (4-5) and ridge regression in the equation (7-8), LASSO regression cannot find direct results for the beta derivative so that one way that can be done is to find the iteration value that minimizes the loss function. The coefficient estimates in LASSO regression are written as follows [11]:

$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg\min_{\boldsymbol{\beta}}(\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \boldsymbol{\alpha}\|\boldsymbol{\beta}\|_1) \tag{10}$$

2.4. **Elastic Net Regression.** Elastic net is a penalty regression method similar to ridge regression and LASSO that can overcome the problem of multicollinearity assumption 2005 [12]. Elastic net combines the penalty between Ridge regression and LASSO. Elastic net can overcome the problem of high correlation and has the properties of variable selection and shrinkage of the estimation coefficient. Zou and Hastie (2005) [12] introduced the Elastic Net penalty as follows:

$$(1 - \lambda)\|\boldsymbol{\beta}\|_2 + \lambda\|\boldsymbol{\beta}\|_1 \leq t, t > 0$$

If $\lambda = 0$, the Elastic Net regression becomes a Ridge regression, while if $\lambda = 1$, the Elastic Net becomes a LASSO penalty. Elastic Net regularization has a shrinkage of the coefficient of correlated predictor variables like Ridge and LASSO. Loss function for Elastic Net regression is as follow:

$$\mathbf{S}(\boldsymbol{\beta})_{ELN} = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \boldsymbol{\gamma}\left((1-\lambda)\|\boldsymbol{\beta}\|_2 + \lambda\|\boldsymbol{\beta}\|_1\right) \tag{11}$$

The coefficient estimates in Elastic Net regression are written as follows:

$$\hat{\boldsymbol{\beta}}_{ELN} = \arg\min_{\boldsymbol{\beta}} \left(\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \boldsymbol{\gamma}\left((1-\lambda)\|\boldsymbol{\beta}\|_2 + \lambda\|\boldsymbol{\beta}\|_1\right)\right) \tag{12}$$

2.5. **Artificial Neural Network (ANN).** Artificial Neural Network (ANN) is an algorithm that has the same structure as the performance of the human brain in learning the pattern of data [13]. In an ANN cell, the weights and biases function as variables for the input data to be output. The weight and bias values are determined by the backpropagation method, which is an adjustment process so that the prediction results are close to the original value. This algorithm works by doing a back pass for each forward pass while adjusting the weights and biases. The process is assisted by an algorithm known as optimizers.

2.6. **Deep Neural Network (DNN).** Neural networks with multiple hidden layers are called deep neural networks (DNN) and the practice of training those networks are referred to as deep learning. Deep neural networks trained to adaptive to varied number of levels and nodes at each level, performance complex tasks, modeling the multiple outcomes.

3. MATERIAL AND METHOD

3.1. **Variable.** This paper uses data taken from BPS in 2018. The description of each variable can be explained below:

Y : Number of villages with gambling occurrences in the last year by province, 2018
X_1 : Open unemployment rate (TPT) by province in 2018
X_2 : Education completion rate by education level (SMA) and province, 2018
X_3 : Average monthly expenditure per capita on food in urban and rural areas by province (IDR), 2018
X_4 : Average monthly non-food expenditure per capita in urban and rural areas by province (IDR), 2018
X_5 : Average hourly wage of workers by province (IDR/hour)

The number of observations is 35 (35 provinces) with the dependent variable (Y) being the number of gambling cases in each province. This study uses log transformation to the dependent variable with the statistics descriptive of the variables is shown in **Table 1** below:

Variables	Minimum	Maximum	Mean
Y	35.0	1947.0	376.8
$ln(Y)$	3.555	7.574	5.422
X_1	1.400	8.470	4.803
X_2	29.56	83.48	61.19
X_3	402922	847847	565941
X_4	301832	1191310	576476
X_5	11359	25987	15911

TABLE 1. Statistics Descriptive

3.2. **Method.** This paper uses natural logarithm (ln) transformation on the dependent variable ($y^* = ln(y)$) to increase the R-square value of the regression model. After the transformation, modeling is done using several regression methods such as linear, ridge, lasso, and elastic net. From each model, the error is calculated using the following formula:

$$\varepsilon = y^* - \hat{y^*}_{\text{single}} \tag{13}$$

The error is then used as input to the Deep Neural Network (DNN) with the aim of filling the gap between the original value and the predicted value and produce the estimated error ($\hat{\varepsilon}$). The final prediction of this hybrid model is formulated as:

$$\hat{y^*}_{\text{final}} = \hat{y^*}_{\text{single}} + \hat{\varepsilon}. \tag{14}$$

To evaluate the performance of the model, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) is used as a comparison metric with the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^* - \hat{y^*}_{\text{final}}\right)^2} \tag{15}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^* - \hat{y^*}_{\text{final}}}{y_i^*}\right| \times 100\% \tag{16}$$

4. RESULTS AND DISCUSSION

4.1. **Variable Selection.** In this subsection, the first step is the selection of independent variables using the backward method in multiple linear regression. This process starts by including all five independent variables along with their intercepts into the initial model. Next, the variables are selected one by one, gradually removing the variables that have the least influence on the model. This process continues until only those variables remain that have a significant contribution to the dependent variable. The final results of this selection show that the three selected independent variables are X_2, X_3, and X_4.

Variables in Model	Description
intercept,X_1, X_2, X_3, X_4, X_5	$\beta_1, \beta_2, \beta_3, \beta_5$ not significant
	β_0, β_4 significant
X_1, X_2, X_3, X_4, X_5	β_1, β_5 not significant
	$\beta_2, \beta_3, \beta_4$ significant
X_2, X_3, X_4, X_5	$\beta_2, \beta_3, \beta_4$ not significant
	β_5 significant
X_2, X_3, X_4	$\beta_2, \beta_3, \beta_4$ significant

TABLE 2. Variable Selection

4.2. **Hybrid Model.** In this subsection, we performed modeling using several single models, namely linear, ridge, lasso, and elastic net models. Each of these models produces an error which is then used as input for the Deep Neural Network (DNN). To improve the performance of the DNN, we performed hyperparameter tuning using the grid search method with a range of 1:50 for each layer on three different layers. **Table 3** shows the best results of the number of neurons in each layer obtained from the tuning process with each RMSE.

Model	Layer 1	Layer 2	Layer 3	RMSE
Linear	9	19	1	0.901931
Ridge	20	5	14	0.749255
Lasso	18	13	18	0.611645
Elastic Net	17	7	6	0.764042

TABLE 3. Best of Hyperparameter for DNN

The results for various combination are given in **Table 4**. Based on this table, it can be seen that the RMSE by the hybrid model is smaller than that of single models such as linear, ridge, lasso, and elastic net models. This shows that the combination of several regression models with the use of Deep Neural Network (DNN) as the final stage of modeling is able to provide more accurate predictions.

Model	RMSE	MAPE	RMSE Hybrid	MAPE Hybrid
Linear	1.03137	15.928%	0.86786	13.347%
Ridge	0.81773	13.144%	0.81446	13.091%
Lasso	0.86023	13.611%	0.56669	7.2017%
Elastic Net	0.86377	13.730%	0.71794	11.219%

TABLE 4. Model Performance

This hybrid approach strengthens the prediction results by utilizing the advantages of each regression model and DNN shown by **Figure 1**, thus capturing complex patterns that a single model may not be able to identify effectively.

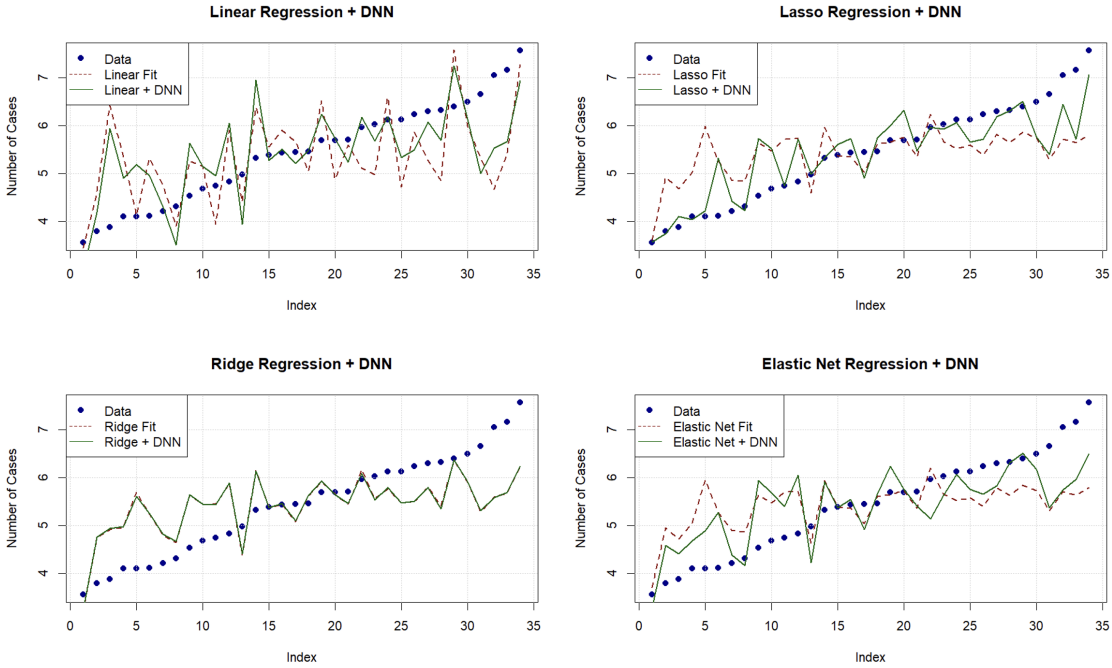

FIGURE 1. Comparison of Single Model vs Hybrid Model

5. CONCLUSION

The conclusion of this study shows that the use of linear regression and its extensions such as ridge, lasso, and elastic net can be maximized by combining it with a deep learning network. This hybrid approach is proven to produce smaller errors compared to a single model, indicating that this combination is able to provide more accurate and efficient predictions. Thus, hybrid models that integrate linear regression and deep learning networks offer a more robust solution in handling data complexity and improving predictive modeling performance.

REFERENCES

[1] D. Khoerunisa, I. Nurahmadi, J. A. Sari, S. Wianti, and Y. E. Y. Siregar, "Judi online sebagai faktor penyebab permasalahan perceraian di kabupaten bekasi:(studi kasus pada kecamatan cikarang utara kabupaten bekasi)," *Kultura: Jurnal Ilmu Hukum, Sosial, dan Humaniora*, vol. 2, no. 2, pp. 63–70, 2024. .

[2] L. Wahkidi, E. Puspitasari, and T. Tamrin, "Hubungan tingkat kecanduan dengan tingkat kecemasan pelaku judi online di wilayah kecamatan toroh," *Jurnal Ilmu Keperawatan Komunitas*, vol. 5, no. 2, pp. 68–76, 2022. .

[3] I. T. Jadidah, U. M. Lestari, K. A. S. Fatiha, R. Riyani, and Wulandari, "Analisis maraknya judi online di masyarakat," *Jurnal Ilmu Sosial dan Budaya Indonesia*, vol. 1, no. 1, pp. 20–27, 2023. .

[4] R. Walpole, R. Myers, S. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientists*. 1995. .

[5] J. Y. L. Chan, S. M. H. Leow, K. T. Bea, W. K. Cheng, S. W. Phoong, Z. W. Hong, and Y. L. Chen, "Mitigating the multicollinearity problem and its machine learning approach: a review," *Mathematics*, vol. 10, no. 8, p. 1283, 2022. .

[6] Z. Zhang, Z. Lai, Y. Xu, L. Shao, J. Wu, and G. S. Xie, "Discriminative elastic-net regularized linear regression," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1466–1481. .

[7] J. Ludbrook, "Linear regression analysis for comparing two measurers or methods of measurement: but which regression?," *Clinical and Experimental Pharmacology and Physiology*, vol. 37, no. 7, pp. 692–699, 2010. .

[8] M. Firdaus, *Ekonometrika Suatu Pendekatan Aplikatif*. 2004. .

[9] Weisstein and E. W., ""vector norm"," `mathworld.wolfram.com`.

[10] G. Azzahra, R. Herryanto, and F. Agustina, "Regresi ridge parsial untuk data yang mengandung masalah multikolinearitas," *Jurnal EurekaMatika*, vol. 8, no. 1, pp. 39–55, 2020. .

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of The Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996. .

[12] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net.," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320. .

[13] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 7, pp. 115–133, 1943. .