Journal of Industrial Engineering
Scientific Journal on Research and Application of Industrial System
Volume 11 No 01 - March 2026
http://e-journal.president.ac.id/presunivojs/in.php/journalofIndustrialEngineerin
ISSN 2527-4139 (*online*) – ISSN 2503-3670 (*print*)

Development of an Artificial Neural Network Model for Predicting Speeding Behaviour: A Case Study from Indonesia

Sekar Sakti[1], Lintang Maulida Sekar Bawono[1], Fitri Trapsilawati[1]
[1)] Department of Mechanical and Industrial Engineering, Faculty of Engineering, Universitas Gadjah Mada
Sleman, Daerah Istimewa Yogyakarta - Indonesia 55281
Email: sekarsakti@ugm.ac.id, lintangmaulida@mail.ugm.ac.id, fitri.trapsilawati@ugm.ac.id

ABSTRACT

Traffic accidents are the third leading cause of death in Indonesia, with speeding behavior being the predominant human factor responsible for most fatal outcomes. Early detection of drivers' propensity to speed is therefore essential for effective prevention strategies. This study develops an Artificial Neural Network (ANN) model to predict the tendency of drivers to speed on intercity roads using a labeled questionnaire dataset comprising 14 input variables. The dataset was divided into training, validation, and testing subsets, where the validation set was used for hyperparameter tuning, while the testing set was used for final evaluation on unseen data. The model was trained using the Adam optimizer with a binary cross-entropy loss function. The optimal configuration consists of a single hidden layer with 12 neurons using ReLU activation, a Sigmoid output layer, 750 training epochs, and a learning rate of 0.03. The final model achieved an accuracy of 86.67% and a Cohen's kappa value of 0.7339, which indicates strong predictive reliability. These findings demonstrate the model's potential as a valuable tool for relevant stakeholders to identify high-risk drivers and design targeted interventions. As a result, the model can be used to proactively reduce speeding-related traffic accidents and improve road safety on intercity routes.

Keywords: Predictive Modeling, Artificial Neural Network, Risk Prediction, Speeding Behaviour, Intercity Roads, Traffic Safety, Transportation, KNIME Analytics.

ABSTRAK

Kecelakaan lalu lintas merupakan penyebab kematian tertinggi ketiga di Indonesia, dimana faktor manusia yakni perilaku mengebut merupakan faktor utama penyebab kecelakaan fatal. Oleh karena itu, deteksi dini kecenderungan pengemudi untuk mengebut sangat penting sebagai upaya pencegahan yang efektif. Studi ini mengembangkan model Jaringan Saraf Tiruan untuk memprediksi kecenderungan pengemudi untuk mengebut di rute antar kota menggunakan dataset kuesioner berlabel yang terdiri dari 14 variabel input psikologis dan demografis. Dataset dibagi menjadi subset pelatihan, validasi, dan pengujian, di mana set validasi digunakan untuk penentuan *hyperparameter*, sedangkan set pengujian digunakan untuk evaluasi akhir pada data yang belum pernah dilihat sebelumnya. Model dilatih menggunakan Adam *optimizer* dengan *binary cross-entropy loss function*. Konfigurasi optimal terdiri dari satu *hidden layer* dengan 12 neuron menggunakan aktivasi ReLU, *output layer* dengan satu neuron menggunakan fungsi Sigmoid, 750 epoch pelatihan, dan *learning rate* 0.03. Model yang diusulkan mencapai tingkat akurasi sebesar 86.67% dan nilai Cohen's kappa sebesar 0.7339, yang menunjukkan reliabilitas prediksi yang kuat. Hasil penelitian ini menunjukkan bahwa model memiliki potensi sebagai alat bantu bagi para pemangku kepentingan untuk mengidentifikasi pengemudi berisiko tinggi dan merancang intervensi yang tepat sasaran. Dengan demikian, model dapat digunakan untuk secara proaktif mengurangi kecelakaan lalu lintas akibat kecepatan berlebih serta meningkatkan keselamatan jalan pada rute antar kota.

Kata Kunci: Pemodelan Prediktif, Jaringan Syaraf Tiruan, Prediksi Risiko, Perilaku Mengebut, Jalan Antarkota, Keselamatan Lalu Lintas, Transportasi, KNIME Analytics.

DOI: https://doi.org/10.33021/jie.v11i01.129

1. Introduction

Traffic accidents pose a significant public health challenge in Indonesia, and are among the leading causes of mortality. According to the World Health Organization (WHO), traffic accidents ranked the third most frequent cause of death in the country (Anshori, 2019). The number of vehicle accidents has increased from 2015 to 2019, as illustrated in Figure 1, with the average annual increase of 4.98%. Data from the Ministry of Transportation further indicated that traffic accidents account for approximately two to three fatalities per hour (Samudra & Wijaya, 2020). This substantial number of accident victims imposes serious economic and social consequences (see Figure 1). In 2020, the economic losses attributed to traffic accidents were estimated to account for approximately 3.1% of Indonesia's Gross Domestic Product, equivalent to IDR 47.8 trillion out of a total GDP of IDR 15,434 trillion (Ginanjar, 2021).

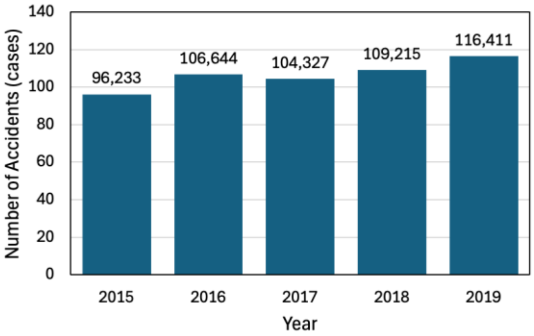

Figure 1. Number of Vehicle Accidents in 2015-2019 (Central Bureau of Statistics, 2019)

According to the Indonesian National Police, 61% of traffic accidents in Indonesia are caused by human factors. Infrastructure and environmental factors contribute to 30% of accidents, while vehicle-related factors account for the remaining 9%. Among the human factors, both driver ability and behavior play significant roles. One specific factor is speeding behavior, which accounts for 6.9% of traffic accidents, following driver negligence during turning or overtaking (27.4%) and failure to maintain a safe distance (16.1%). Although these statistics primarily reflect accident frequency rather than severity, multiple studies have demonstrated that speeding is strongly associated with fatalities (LTSA, 2001; Atombo *et al.*, 2016). Therefore, speeding behavior warrants particular attention in traffic safety research and policy development.

In Indonesia, vehicle speed limits are regulated under Government Regulation No. 79 of 2013 on Road Traffic and Transportation Networks (Article 23, Paragraph 4), and further supported by the Minister of Transportation Regulation on Procedures for Determining Vehicle Speed Limits (Article 3, Paragraph 4). These regulations stipulate that the maximum speed limits are as follows:

a. 100 km/h on expressways (with a minimum of 60 km/h),
b. 80 km/h on intercity roads,
c. 50 km/h in urban areas, and
d. 30 km/h in residential zones.

The limits above serve as a legal benchmark for safe driving behavior; any vehicle operating beyond these thresholds is considered to be exceeding the speed limit, or engaging in speeding.

The Indonesian government has undertaken various initiatives to prevent and address speeding behavior, including public awareness campaigns and safety promotion programs (Ridwan, 2018; Anshori, 2019; Liputan6.com, 2021). Despite these efforts, speeding remains prevalent, particularly on intercity roads. These conditions suggest that conventional enforcement and awareness strategies alone may be insufficient and should be complemented with more proactive and data-driven approaches.

Previous studies have examined psychological and behavioral determinants of speeding behavior. Research grounded in the Theory of Planned Behavior has demonstrated that perceived behavioral control, subjective norms, attitudes, and behavioral intention significantly influence drivers' likelihood of speeding (Atombo *et al.*, 2016; Boisson *et al.*, 2019). Other studies have also identified contextual and demographic factors, including sanction promptness, driver gender, driving license ownership, and average daily distance traveled, as relevant contributors (Bawono & Trapsilawati, 2021). While these studies provide important explanatory insights into why drivers engage in risky behavior, most of them focused on identifying influencing factors rather than developing predictive models capable of identifying high-risk drivers before violations occur. A notable gap therefore exists in research on predicting the likelihood of drivers to speed. A relevant study was conducted by Cheng *et al.* (2019), who utilized speed recorder data from the Wujiang Public Security Bureau to predict speeding behavior. However, such approach primarily detect violations after they occur, which limits their usefulness for early intervention. Early detection of speeding behavior is critical, as it could enable timely interventions to reduce the incidence of speed-related traffic accidents (Bawono & Trapsilawati, 2022). In this context, predictive modeling using behavioral data offers a practical approach for developing decision-support tools that can assist transportation authorities and related stakeholders in designing effective interventions.

This study fills the gap by developing a predictive model using an Artificial Neural Network (ANN) to identify drivers' propensity to engage in speeding behavior on intercity roads (i.e., regular roads) in Indonesia. ANN is selected due to its capability to capture complex and non-linear relationships among multiple behavioral, psychological, and demographic variables without relying on restrictive statistical assumptions (Karanika-Murray & Cox, 2010). These characteristics make ANN particularly suitable for modeling human driving behavior, where interactions among influencing factors are often multidimensional.

The contributions of this study are threefold. First, it introduces a predictive modeling approach for speeding behavior using a theory-driven behavioral dataset rather than relying solely on vehicle-based speed measurements. Second, it applies an ANN model to the Indonesian intercity road context, which remains underrepresented in predictive traffic safety research. Third, this study demonstrates that the proposed model achieves strong predictive performance and highlights its potential as a practical decision-support tool for traffic safety stakeholders seeking to reduce speed-related accidents through proactive and targeted interventions.

2. Methodology

This section outlines the methodological framework for developing our ANN model, as seen in Figure 2. The modeling process consists of three main stages: data preprocessing, ANN architecture design and training, and performance evaluation (see). The proposed ANN model was implemented using KNIME Analytics, an open-source data analytics platform that enables modular workflow construction for preprocessing, network configuration, training, and evaluation (KNIME, n.d.). KNIME integrates Keras-based neural network components within a structured visual workflow environment. Although KNIME provides a visual interface for workflow design, the underlying ANN training process follows standard backpropagation and gradient-based optimization principles. Therefore, the use of KNIME serves as an implementation platform without altering the theoretical structure of the ANN model.

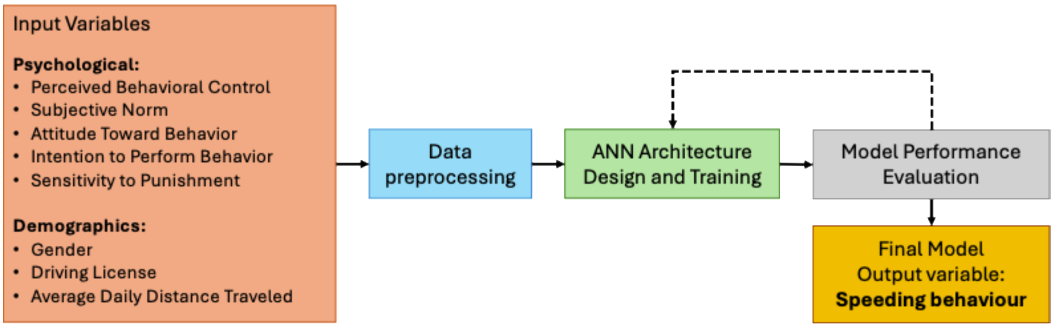

Figure 2. Model development framework

2.1 Data Source and Variable Definition

This study adopts the questionnaire data by Bawono & Trapsilawati (2021). They implemented Theory of Planned Behaviour (TPB), Classical Deterrence Theory (CDT), and Sensitivity to Punishment and Sensitivity to Reward (SPSRQ) to look for factors that influence speeding behaviour. Speeding behaviour was divided into three, namely exceeding the speed of vehicles on intercity roads, in urban areas, and in residential zones. By implementing the Driver Behaviour Questionnaire (DBQ), speeding behaviour was categorized into three: mistake, deliberate violation, unintended violation. The study identified factors that significantly affect speeding behaviour, namely Perceived Behavioural Control (X1, X2, and X3), Subjective Norm (X4, X5, and X6), Attitude (X7 and X8), Intention (X9 and X10), Swiftness (X11), Gender (X12), Driving License "SIM" (X13) ownership, and the average daily distance traveled (X14). Details of the variables are provided in the Appendix Tabel A1. This study uses these factors as input variables (X) in the model training process. Meanwhile, speeding behaviour (i.e., exceeding the speed limit of 80 km/hour on intercity roads) is used as the response variable (Y). This selection is based on its R-squared value, which is the highest among all behavioral types analyzed (R^2 = 0.408), compared to Mistake violation (R^2 = 0.192), Deliberate violation (R^2 = 0.331), and Unintended violation (R^2 = 0.117) (Bawono & Trapsilawati, 2021). In other words, the significant factors identified effectively explain speeding behavior, and this aligns with government regulations. Furthermore, speeding on intercity roads exhibited the highest coefficient (0.844), followed by speeding in urban areas (0.830), while speeding in residential areas was excluded due to its outer loading value being below 0.7. This suggests that intercity roads should be a primary focus for traffic enforcement and policy interventions. In our context, speeding behavior on intercity roads is selected as the most appropriate target for the predictive model, as it demonstrates the strongest statistical relationship with the explanatory variables.

2.2 Data Preprocessing

Prior to model development, the dataset was examined for completeness and internal consistency. There were no substantial missing values identified; therefore, data imputation procedures were not required. The dataset was first partitioned into two subsets, namely 70% for training and 30% for validation and testing (Chandwani, 2016). The "Partitioning" node in KNIME Analytics used for this process is illustrated in Figure 3 (left). This step ensures independent model evaluation and prevents information leakage between training and testing data. All input variables were normalized using Min–Max scaling to ensure comparable feature magnitudes and enhance training stability. The "Normalizer" node applied for this step is shown in Figure 3 (right). Normalization prevents variables with larger numerical ranges from dominating the learning process.

Because the dataset exhibited class imbalance between speeding and non-speeding observations, the Synthetic Minority Oversampling Technique (SMOTE) was applied prior to the data partition (Fernández *et al.*, 2018). This technique generates synthetic samples of the minority class to reduce model bias toward the majority class and improve sensitivity in identifying high-risk drivers (see Figure 3).

Figure 3. Partitioning node (left) and Normalizer node (right) (KNIME, n.d.)

2.3 Artificial Neural Network Architecture and Training

2.3.1 Model Selection Rationale

An ANN based on a Multilayer Perceptron (MLP) architecture was selected due to its ability to capture complex, non-linear relationships among multiple variables. Speeding behavior is shaped by interacting behavioral constructs that may not follow linear patterns. Unlike conventional statistical models, ANN does not require restrictive distributional assumptions and is well suited for modeling multidimensional human behavior patterns. Compared with other algorithms such as logistic regression or decision trees, ANN offers greater flexibility in identifying latent interactions among variables (Wanyonyi *et al.*, 2025; Beucher *et al.*, 2019). Therefore, ANN is well suited for predicting behavioral risk.

2.3.2 Model Architecture

The ANN developed in this study (see Table 1) consists of three main components: an input layer, a single hidden layer, and an output layer. The input layer contains 14 neurons representing the predictor variables (X1-X14) listed in Appendix Table A1, while the output layer represents the dependent variable (Y), namely drivers' tendency to engage in speeding behaviour. The overall ANN configuration and network architecture are presented in Table 1 and Figure 4, respectively. Implementation was carried out in KNIME using a sequence of Keras Layer nodes, as illustrated in Figure 5: (1) Keras Input Layer, (2) Keras Dense Layer for the hidden layer, and (3) Keras Dense Layer for the output layer. The input layer was configured according to the number of predictors specified in Table 1.

The hidden layer employs the Rectified Linear Unit (ReLU) activation function, which is widely used because of its computational efficiency and its ability to activate only neurons with a non-zero result (Goodfellow *et al.*, 2016). The number of neurons in this layer was varied based on the number of input variables (X_n); following the approach of Lima-Junior and Carpinetti (2019), testing configurations ranging from $X_n - 2$ to $X_n + 2$. The structure of the output layer depends on the classification task. For binary classification, a single neuron with a Sigmoid activation function was used to produce outputs between 0 and 1. For multi-class classification, the number of output neurons equals the number of target categories, and the Softmax activation function is applied (Doug, 2010). Accordingly, this study tested output configurations of either one neuron (Sigmoid) or five neurons (Softmax), depending on the model variant.

Table 1. ANN model configuration

Setting Model	Description / Options
Number of input layer units	14
Number of hidden layer units	X_n-2, X_n-1 X_n, X_n+1, X_n+2
Hidden layer activation function	ReLU
Number of output layer units	1, 5
Layer output activation function	Sigmoid, Softmax

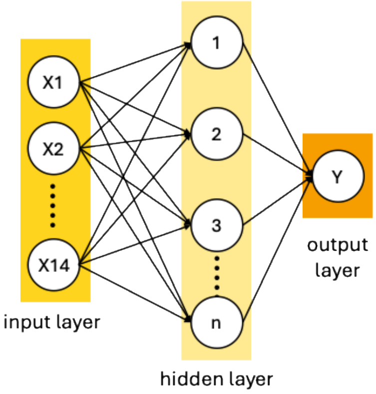

Figure 4. The proposed network architecture

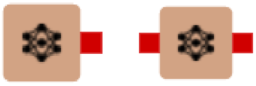

Figure 5. Keras input layer node (left) and Keras dense layer node (right) (KNIME, n.d.)

2.3.3 Training Procedure

In KNIME, the training is facilitated through the “Keras Network Learner” node (Figure 6, left). Within the “Keras Network Learner” node, key configuration parameters—including the number of epochs, learning rate, batch size, optimizer, and loss function—are specified. At the beginning of the experiment, the number of training epochs was set to 1,000, and the learning rate was initialized at 0.1. These are based on commonly recommended values for initial neural network development (Brownlee, 2019). These parameters were subsequently tuned to explore the possibility of achieving a more optimal configuration. The batch size was set to 32, a default size frequently cited in literature as effective for training stability and convergence (Bengio, 2012). The choice of loss function in this study follows the activation function used in the output layer. Since the problem addressed in this study is classification, a standard loss function of binary cross-entropy was employed when using the Sigmoid activation function, while categorical cross-entropy was applied in conjuction with the Softmax function.

For model training, the Adam optimization algorithm is employed for training due to its effectiveness and efficiency during the training (Doshi, 2019). Adam combines two extensions of stochastic gradient descent, namely Root Mean Square Propagation (RMSProp) and the Adaptive Gradient Algorithm (AdaGrad). Additionally, it is well-suited for handling scattered or sparse data (Giordani, 2020). After training, the model was saved using the "Keras Network Writer" node (Figure 6, right) to ensure reproducibility and enable later reuse without retraining.

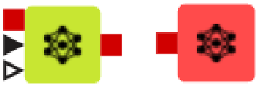

Figure 6. Keras Network Learner node (left) and Keras Network Writer node (right), (KNIME, n.d.)

2.4 Model Prediction and Evaluation

For prediction and evaluation, the saved model was first retrieved using the "Keras Network Reader" node (Figure 7, left). The "Keras Network Executor" node (Figure 7, right) with KNIME's default settings was then used to generate predictions on the testing dataset. The resulting outputs were passed to the "Container Output" node (Figure 8, left), while performance metrics were computed and displayed using the "Scorer" node (Figure 8, right).

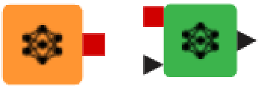

Figure 7. Keras Network Reader node (left) and and Keras Network Executor node (right) (KNIME, n.d.)

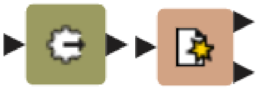

Figure 8. Container Output node (left), and Scorer node (right) (KNIME, n.d.)

Model performance was evaluated using multiple complementary metrics to ensure comprehensive assessment. Accuracy was calculated to measure the proportion of correctly classified instances. However, because accuracy alone may be insufficient in imbalanced datasets, Cohen's kappa was also computed. It provides a more robust indication of classification reliability by accounting for the probability of random agreement (Bland, 2015). In addition, precision, recall (sensitivity), and F1-score (F-measure) were examined to evaluate the model's ability to distinguish high-risk drivers from non-risk drivers. These metrics are particularly important in behavioral risk prediction, where both false positives and false negatives carry practical implications. Performance evaluation was conducted using an independent testing dataset, and comparative analysis across hyperparameter configurations was performed to determine the most robust model structure.

3. Results and Discussion

3.1 Preliminary Evaluation

At this section, we focused on comparing the performance of the Sigmoid and Softmax activation functions. For the model using the Sigmoid activation function, the oversampled dataset contained 494 instances, of which 345 were allocated for training and the remaining 149 for validation and testing. These 149 instances were further divided into 74 for validation and 75 for testing. In contrast, the model employing the Softmax activation function used a dataset of 420 instances, with 294 assigned for training and the remaining 126 split equally between validation and testing (63 instances each). As previously described, the experiment was conducted using the initial configuration of the proposed ANN model: 1000 epochs, a learning rate of 0.1, and a batch size of 32. These baseline settings were used to perform experiments by varying the number of hidden layer units and activation functions.

Tables 2 and 3 summarize the results for Sigmoid and Softmax, respectively. When using the Sigmoid function with one output unit, the model achieved its highest accuracy of 0.8 and Cohen's kappa of 0.601205 with 12 hidden units. Interestingly, using 14 hidden layer units also resulted in the same accuracy (0.8) and a slightly higher Cohen's kappa (0.602052). However, the difference in performance was negligible and not statistically significant. Therefore, 12 hidden units were favored due to the principle of model simplicity, where fewer parameters reduce computational burden and help avoid overfitting. In contrast, using Softmax in a five-class output layer resulted in considerably lower accuracy values, ranging from 0.31746 to 0.44444. These findings indicate that the Sigmoid function is more appropriate for binary classification tasks, such as predicting speeding behavior. In this case, the prediction output obtained is in the form of "Speeding" or "Not speeding". Consequently, the Sigmoid activation function with 12 hidden layer units was selected for subsequent experiments. The model architecture for predicting speeding behavior on intercity roads is depicted in Figure 9. This features 12 units of hidden layer utilizing the ReLU activation function, and a single unit in the output layer, where the Sigmoid activation function is applied.

Table 2. Preliminary experiment results with Sigmoid activation function

Number of Output Layer Units	Number of Hidden Layer Units	Accuracy	Cohen's kappa
1	12	0.8	0.601205
1	13	0.76	0.522293
1	14	0.8	0.602052
1	15	0.74667	0.495932
1	16	0.76	0.522293

Table 3. Preliminary experiment results with Softmax activation function

Number of Output Layer Units	Number of Hidden Layer Units	Accuracy	Cohen's kappa
5	12	0.412698	0.265364
5	13	0.31746	0.166974
5	14	0.39682	0.236850
5	15	0.44444	0.283858
5	16	0.42857	0.287688

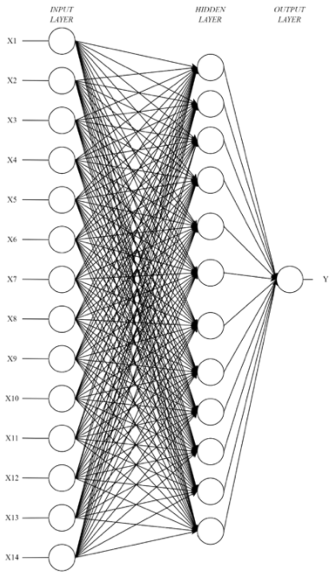

Figure 9. Final architecture of ANN to predict speeding behaviour

3.2 Parameters Tuning and Final Model Selection

To identify the optimal training configuration, parameter tuning was conducted by varying the number of epochs (1000, 750, 500, 300) and learning rates (0.1, 0.03, 0.001). As depicted in Table 4, the results demonstrated that the highest accuracy (0.86667) and Cohen's kappa (0.733854) were achieved using 750 epochs and a learning rate of 0.03. Based on these observations, a follow-up experiment was conducted by increasing the epoch hyperparameter to 875 for the same model. However, as shown in Table 5, this adjustment resulted in lower performance (accuracy 0.813334, Cohen’s kappa 0.627660). This implies that further increasing epochs would lead to overfitting rather than improvement.

A similar experiment was conducted using 14 hidden layer units. As shown in Tables 6 and 7, the best configuration for this model was also achieved using 750 epochs and a learning rate of 0.03, with an accuracy of 0.86667 and Cohen’s kappa of 0.733286. While the 14-unit model showed identical performance in terms of accuracy, it is important to highlight that the 12-unit model achieved slightly better kappa. This emphasizes that the 12-unit model is preferable due to its lower complexity, which reduces the risk of overfitting and training time without sacrificing performance. Based on these considerations, the configuration with 12 hidden layer units, 750 epochs, and a learning rate of 0.03 was selected as the final model.

Table 4. The results of model experiments with 12 hidden layer units

Number	Number of Output Layer Units	Epoch	Learning Rate	Accuracy	Cohen's kappa
0	12	1000	0.1	0.8	0.601205
1		750	0.1	0.77334	0.548673
2		500	0.1	0.74667	0.493423
3		300	0.1	0.82667	0.653887
4		1000	0.03	0.84	0.680624
5		750	0.03	0.86667	0.733854
6		500	0.03	0.78667	0.575371
7		300	0.03	0.77334	0.547391
8		1000	0.001	0.85334	0.707343
9		750	0.001	0.82667	0.654133
10		500	0.001	0.81334	0.62766
11		300	0.001	0.77334	0.547712

Table 5. Experimental Results Model 12 Hidden Layer Units and 875 Epoch

	Speeding	No. Speeding
True Positives	27	34
False Positives	11	3
True Negatives	34	27
False Negatives	3	11
Precision	0.71053	0.91892
Sensitivity	0.9	0.75556
Specificity	0.75556	0.9
F-measure	0.79412	0.82927
Accuracy	0.813334	
Cohen's kappa	0.627660	

Table 6. Results of Model Experiments with 14 Units Hidden Layer

Number	Number of Hidden Layer Units	*Epoch*	*Learning Rate*	Accuracy	*Cohen's kappa*
0	14	1000	0.1	0.8	0.602052
1		750	0.1	0.77334	0.547391
2		500	0.1	0.76	0.521955
3		300	0.1	0.74667	0.49557
4		1000	0.03	0.76	0.521616
5		750	0.03	0.86667	0.733286
6		500	0.03	0.81334	0.627131
7		300	0.03	0.86667	0.733665
8		1000	0.001	0.81334	0.627395
9		750	0.001	0.84	0.680625
10		500	0.001	0.81334	0.62766
11		300	0.001	0.78667	0.5744681

Table 7. Experimental results model 14 hidden layer units and 875 epoch

	Speeding	**No Speeding**
True Positives	32	31
False Positives	6	6
True Negatives	31	32
False Negatives	6	6
Precision	0.84210	0.83784
Sensitivity	0.84210	0.83784
Specificity	0.83784	0.84210
F-measure	0.84210	0.83784
Accuracy	0.84	
Cohen's kappa	0.679943	

3.3 Discussion and Limitations

This study evaluates a predictive model designed to identify drivers' propensity to engage in speeding behavior. The model achieves an accuracy of 86.67%, which exceeds the commonly accepted benchmark of 70% for strong and realistic performance (Brownlee, 2019). Because overall accuracy can be misleading in the presence of class imbalance, Cohen's kappa was also used to assess agreement beyond chance. The model obtained a Cohen's kappa of 0.7339, which indicates substantial agreement and strong classification reliability. Together, the high accuracy and Cohen's kappa values confirm that the proposed ANN model delivers consistent and dependable performance rather than results driven by majority-class dominance.

When compared with existing literature, the model's performance is competitive and, in several cases, superior to previous neural network–based studies. Xiang *et al.* (2021), for example, reported 84% accuracy using a hybrid Cloud Model and Elman Neural Network to classify dangerous driving behavior, while Jamal & Umer (2020) achieved 77.5% accuracy in predicting traffic injury severity using ANN. The higher predictive performance observed in the present study suggests that integrating theory-driven behavioral variables with ANN modeling can effectively capture drivers' speeding tendencies. Compared with the study by Cheng *et al.* (2019), which employed a decision tree algorithm, the accuracy of our model is lower than their reported 92.18%. However, this difference should be interpreted cautiously. Their study focused on classifying specific levels of speeding (e.g., 10–20% versus 20–50% above the speed limit) using objective quantitative data obtained from speed cameras. In contrast, our study predicts whether a driver is likely to engage in speeding behavior based on qualitative data derived from questionnaires. Given the substantial differences in prediction focus and data sources, a direct comparison between the two models becomes less precise.

Despite the promising performance of our proposed model, several limitations should be considered when interpreting the findings. The dataset relies on self-reported questionnaire responses, which may introduce response biases such as social desirability effects, recall inaccuracies, or underreporting of risky behaviors. In addition, the dataset exhibited class imbalance between speeding and non-speeding observations. Although SMOTE was applied to improve minority-class representation, synthetic oversampling may not fully reflect real-world distributions. These factors may affect the model's external validity and generalizability.

3.4 Implications and Future Directions

The proposed ANN model demonstrates that psychological and demographic variables can be effectively utilized to predict drivers' propensity for speeding behavior on intercity roads. With high accuracy and Cohen's kappa, the model shows strong potential for identifying high-risk drivers using behavioral indicators. From a theoretical perspective, this study advances transportation safety research by illustrating how theory-driven constructs can be integrated with machine learning techniques to enable predictive modeling of risky driving behavior. From a practical standpoint, the findings offer several actionable applications for transportation authorities and policymakers. For example, the model could be incorporated into driver licensing or renewal systems to identify individuals with elevated speeding risk, this would enable targeted educational interventions or behavioral training prior to license issuance or renewal. Traffic authorities could also integrate the predictive framework into smart traffic management systems by combining behavioral risk profiling with enforcement databases. Additionally, insurance providers may utilize similar behavioral assessments to design incentive-based safety programs aimed at promoting safer driving practices.

Future research should validate the model using larger and more representative samples to assess its generalizability. Comparative evaluations with alternative machine learning algorithms would further clarify the relative strengths of the proposed ANN. Additionally, extending the model to other risky driving behaviors, including unsafe overtaking, tailgating, aggressive driving, and distracted driving, may further enhance its applicability in transportation safety management.

4. Conclusions

This study demonstrates that an ANN model can effectively predict drivers' propensity to engage in speeding behavior on intercity roads. The proposed model indicates strong and reliable classification performance with accuracy of approximately 86.67% and Cohen's kappa 0.7339. Beyond methodological contribution, the model offers practical utility for transportation authorities and policymakers. By enabling early identification of high-risk drivers, the predictive model supports a shift from reactive enforcement toward proactive risk management to support safer transportation systems. The model could be integrated into driver licensing systems, traffic monitoring platforms, or decision-support tools to inform targeted education programs and optimized enforcement strategies. Although the present study focuses on speeding behavior, the modeling framework can be extended to other risky driving behaviors. Future research could also incorporate larger and more diverse datasets to further enhance robustness and real-world applicability.

References

1. Anshori, L. (2019, March 16). Ngeri, Kecelakaan Lalu Lintas Penyebab Kematian Nomor 3 di RI. Retrieved from https://oto.detik.com/motor/d-4470269/ngeri-kecelakaan-lalu-lintas-penyebab-kematian-nomor-3-di-ri
2. Atombo, C., Wu, C., Zhong, M., & Zhang, H. (2016). Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: Speeding and overtaking violations, *Transportation Research Part F: Traffic Psychology and Behaviour*, 43, 104–121. https://doi.org/10.1016/j.trf.2016.09.029
3. Bawono, L. M. S., & Trapsilawati, F. (2021). Pengembangan Model Perilaku Mengebut di Indonesia serta Rekomendasi Pencegahannya, *Seminar Nasional Teknik dan Manajemen Industri*, Universitas Maranatha, Bandung.
4. Bawono, L. M. S, & Trapsilawati, F. (2022). Pengembangan Model Artificial Neural Network Untuk Memprediksi Perilaku Mengebutdi Indonesia, Tesis, Departemen Teknik Mesin dan Industri, Universitas Gadjah Mada.
5. Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. *Neural networks: Tricks of the trade* (pp. 437-478), Berlin, Heidelberg: Springer Berlin Heidelberg.
6. Beucher, A., Møller, A. B., & Greve, M. H. (2019). Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark, *Geoderma*, 352, 351-359. https://doi.org/10.1016/j.geoderma.2017.11.004
7. Bland, M. (2015). *An introduction to medical statistics*, UK: Oxford University Press.
8. Boisson, C., al Maniri, A. A., Al-Azri, A. S., Hasselberg, M., & Laflamme, L. (2019). Determinants of speeding among new generations of car drivers from the Arabian Peninsula. An investigation based among Omani drivers using the theory of planned behaviour, *PLoS ONE*, 14(12), 1–14. https://doi.org/10.1371/journal.pone.0226441
9. Brownlee, J. (2019, August 6). How to Configure the Learning Rate When Training Deep Learning Neural Network. Retrieved from https://machinelearningmastery.com/learning-rate-for-deep-learning-neural-networks/
10. Central Bureau of Statistics. (2019). Jumlah Kecelakaan, Korban Mati, Luka Berat, Luka Ringan, dan Kerugian Materi. Retrieved from https://www.bps.go.id/indicator/17/513/1/jumlah-kecelakaan-korban-mati-luka-berat-luka-ringan-dan-kerugian-materi.html

11. Chandwani, V. (2016, March 16). Is there an ideal ratio between a training set and validation set? Which trade-off would you suggest?. Retrieved from https:// researchgate.net/post/Is-there-an-ideal-ratio-between-a-training-set-and-validation-set-Which-trade-off-would-you-suggest/
12. Cheng, Z., Lu, J., Zu, Z., & Li, Y. (2019). Speeding Violation Type Prediciton Based on Decision Tree Method: A Case Study in Wujiang, China, *Journal of Advanced Transportation*, 8650845. https://doi.org/10.1155/2019/8650845
13. Doshi, S. (2019, Januari 13). Various Optimization Algorithms for Training Neural Networks. Retrieved from https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6
14. Doug. (2010, July 20). How to choose the number of hidden layers and nodes in a feedforward neural network?. Retrieved from https://stats.stackexchange.com/questions/181/how-to-choose-the-number-of-hidden-layers-and-nodes-in-a-feedforward-neural-netw
15. Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary, *Journal of artificial intelligence research*, 61, 863-905. https://doi.org/10.1613/jair.1.11192
16. Ginanjar, D. (2021, November 16). Kerugian Ekonomi Akibat Kecelakaan Mencapai Rp 47,8 Triliun. Retrieved from https://www.jawapos.com/oto-dan-tekno/otomotif/16/11/2021/kerugian-ekonomi-akibat-kecelakaan-mencapai-rp-478-triliun/
17. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (Vol, 1, pp. 23-24), Cambrigde, MA, USA: MIT Press.
18. Jamal, A., & Umer, W. (2020). Exploring the Injury Severity Risk Factors in Fatal Crashes with Neural Network, *Environmental Research and Health Public,* 17(20), 7466. https://doi.org/10.3390/ijerph17207466
19. Karanika-Murray, M., & Cox, T. (2010). The use of artificial neural networks and multiple linear regression in modelling work-health relationships: Translating theory into analytical practice, *European Journal of Work and Organizational Psychology*, *19*(4), 461-486. https://doi.org/10.1080/13594320902995916
20. Knime. (n.d.). Getting Started. Retrieved from https://www.knime.com/getting-started-guide#:~:text=In%20KNIME%20*Analytics*%20Platform%2C%20individual,creating%20visualizations%2C%20and%20so%20on
21. Land Transport Safety Authority (LTSA). (2001). Facts About Speed. Retrieved from https://www.ltsa.govt.nz/factsheets/33.html
22. Lima-Junior, F. R., & Carpinetti, L. C. R. (2019). Predicting Supply Chain Performance based on SCOR Metrics and Multilayer Perceptron Neural Networks, *International Journal of Production Economics,* 212, 19-38. https://doi.org/10.1016/j.ijpe.2019.02.001
23. Liputan6.com. (2021, April 29). Kebut-Kebutan Bukan Penyebab Utama, Ini Biang Keladi Kasus Kecelakaan di Indonesia. Retieved from https://www.liputan6.com/otomotif/read/4545221/kebut-kebutan-bukan-penyebab-utama-ini-biang-keladi-kasus-kecelakaan-di-indonesia
24. Ridwan, M. F. (2018, October 21). Kecelakaan Penyebab Kematian Terbesar Kedua di Indonesia. Retireved from https://www.republika.co.id/berita/nasional/umum/18/10/21/pgxsol354-kecelakaan-penyebab-kematian-terbesar-kedua-di-indonesia
25. Samudra, M. A., & Wijaya, I. (2020, March 6). Angka Kematian Kecelakaan Lalu Lintas di Indonesia Tinggi, Tiap Satu Jam Tiga Orang Tewas. Retreived from https://otomotifnet.gridoto.com/read/232051060/angka-kematian-kecelakaan-lalu-lintas-di-indonesia-tinggi-tiap-satu-jam-tiga-orang-tewas
26. Wanyonyi, M., Morris, Z. N., Musyoka, F. M., Makaa, D., & Kitavi, D. M. (2025). Balancing interpretability and accuracy: a comparative evaluation of logistic regression and machine learning approaches for coronary heart disease risk prediction in low-resource settings, *Authorea Preprints*. https://doi.org/10.36227/techrxiv.175393687.78366329/v1
27. Xiang, H., Zhu, J., Liang, G., & Shen, Y. (2021). Prediction of Dangerous Driving Behaviour Based on Vehicle Motion State and Passenger Feeling Using Cloud Model and Elman Neural Network, *Frontiers in neurorobotics*, 15, 641007. https://doi.org/10.3389/fnbot.2021.641007

Appendix

Table A1. Variables description

Variable	Label	Category	Data Type	Explanation
Y	Speeding Behaviour on Intercity Roads	Behaviour	Nominal or Ordinal	1 = Exceed the speed limit 0 = Does not exceed the speed limit Or 1 = Never exceed the speed limit 2 = Rarely exceed the speed limit 3 = Sometimes exceed the speed limit 4 = Often exceed the speed limit 5 = Always exceed the speed limit
X1	Since I am experienced in driving, it is easy for me to speed up	Behavioural control	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X2	The feeling of joy and feeling in control of my vehicle makes it easier for me to speed up	Behavioural control	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X3	Smooth and good road surface (no potholes and bumps) makes it easier for me to speed up	Behavioural control	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X4	I think my family agrees that I drive above the maximum vehicle speed limit	Subjective norm	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X5	I think my brother/sister agrees that I drive above the maximum vehicle speed limit	Subjective norm	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X6	I think my friends agree that I drive above the maximum vehicle speed limit	Subjective norm	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X7	Everyone must obey the maximum speed limit on the road for mutual safety	Attitude to behaviour	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X8	I believe that speeding (over the maximum speed limit) can result in an accident	Attitude to behaviour	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X9	How likely is it that you intend to ignore traffic signs on a good (no potholed) road with low (quiet) traffic flow?	Attitude to behaviour	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X10	How likely is it that you intend to exceed the maximum speed limit of vehicles on the road if you feel your vehicle is capable of doing so?	Intention to perform behaviour	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree

Variable	Label	Category	Data Type	Explanation
X11	The time between being caught speeding and receiving a penalty is very short	Swiftness	Ordinal	1 = Strongly disagree 2 = Disagree 3 = Neutral 4 = Agree 5 = Strongly agree
X12	Gender	Demographics	Nominal	0 = Male 1 = Female
X13	Driving License	Demographics	Nominal	1 = SIM A (cars) 2 = SIM B (large vehicles/heavy equipment vehicles) 3 = SIM C (motorcycles) 4 = SIM D (specially adapted vehicles for drivers with disabilities) 5 = More than one
X14	Vehicle Mileage	Demographics	Ordinal	1 = <11 km 2 = 11 km – 25 km 3 = 26 km – 55 km 4 = >55 km