Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs

Tugas Akhir

Oleh:

Vira Khairatul Hisan 4212231002

Program Studi Teknik Mekatronika
Jurusan Teknik Elektro
Politeknik Negeri Batam
2024

Surat Penyataan Keaslian Tugas Akhir

Saya yang bertandatangan dibawah ini menyatakan bahwa isi sebagian maupun keseluruhan Jurnal Ilmiah yang berjudul : “Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs” adalah **hasil karya bersama seperti terlampir dalam jurnal ilmiah tersebut nama-nama penulis yaitu : Heru Wijanarko, Vira Khairatul Hisan, Wahidil Fajri, Hamdani Arif, dan Andy Triwinarko, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang diakui sebagai karya sendiri.** Semua referensi yang dikutip atau dirujuk telah tertulis secara lengkap pada daftar Pustaka. Apabila ternyata pernyataan saya ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Batam, 3 Maret 2025

Vira Khairatul Hisan
NIM : 4212231002

Lembar Pengesahan

Tugas Akhir disusun untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana Terapan Teknik (S.Tr.T) di Politeknik Negeri Batam

Disusun oleh:

Vira Khairatul Hisan 4212231002

Disetujui oleh:

Heru Wijanarko, S.T., M.Sc
NIK : 110070

Kata Pengantar

Puji syukur senantiasa penulis panjatkan kehadirat Allah SWT atas segala limpahan rahmat, nikmat dan karunia-Nya sehingga penulis dan rekan lainnya dapat menyelesaikan tulisan ilmiah yang berjudul "***Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs***" yang telah dipaparkan pada konferensi ilmiah dan tulisan ini juga digunakan untuk memenuhi syarat kelulusan dari jurusan Teknik Elektro studi D4 Teknik Mekatronika, Politeknik Negeri Batam.

Pada kesempatan ini penulis menyampaikan terima kasih atas segala bantuan dan dukungannya yang telah diberikan selama proses pembuatan tulisan ilmiah ini kepada,

1. Heru Wijanarko, S.T., M.Sc. selaku dosen pembimbing yang telah banyak berkontribusi dalam membimbing, mendukung dan meluangkan waktu sehingga terselesaikan tulisan ilmiah ini.
2. Seluruh Dosen Teknik Mekatronika yang telah berjasa dalam memberikan ilmu pengetahuan yang bermanfaat bagi penulis selama proses menempuh Pendidikan di Politeknik Negeri Batam.
3. Dan semua pihak lain yang ikut serta memberikan bantuan dan dorongan dalam proses penulisan ilmiah ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan pada tulisan ilmiah ini, oleh karena itu penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak. Semoga tulisan ilmiah ini dapat bermanfaat khususnya bagi penulis laporan ini dan pada umumnya bagi para pembaca.

Batam, 3 Februari 2025

Penulis

Daftar Isi

2024 IEEE International Conference on Communication, Networks and Satellite

Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs

Heru Wijanarko
Department of Electrical Engineering
Politeknik Negeri Batam
Batam, Indonesia
wijanarko@polibatam.ac.id

Vira Khairatul Hisan
Department of Electrical Engineering
Politeknik Negeri Batam
Batam, Indonesia
virachairatul@gmail.com

Wahidil Fajri
Department of Electrical Engineering
Politeknik Negeri Batam
Batam, Indonesia
wahidil.fajri@gmail.com

Hamdani Arif
Department of Informatics Engineering
Politeknik Negeri Batam
Batam, Indonesia
hamdaniarif@polibatam.ac.id

Andy Triwinarko
Department of Informatics Engineering
Politeknik Negeri Batam
Batam, Indonesia
andy@polibatam.ac.id

***Abstract*—The study addresses the issue of predicting traffic density in Vehicular Ad-hoc Networks (VANETs), where dynamic and unexpected traffic patterns limit accurate forecasting. Recent models frequently encounter challenges with accuracy caused by overfitting or complications in handling real-time data. The study introduces a hybrid model that combines Random Forest with Simulated Annealing, optimising the model's parameters to mitigate overfitting and improve reliability. The research follows several steps: first, data from a VANETs dataset was collected and preprocessed, and then several standard machine learning models, like Linear Regression, Decision Trees, Random Forest, Support Vector Regression, and K-Nearest Neighbors, were tested. The Random Forest model showed the best performance metrics and was optimized using Simulated Annealing. The hybrid Simulated Annealing-Random Forest model significantly improved accuracy, outperforming traditional models.**

***Keywords*—machine learning optimization, random forest, simulated annealing, traffic density prediction, vehicular ad-hoc networks**

I. Introduction

Vehicular Ad-hoc Networks (VANETs) are increasingly crucial to intelligent transportation systems (ITS), and their development is rising. Recent research from Grand View Research estimates that the VANET market could grow to around $16.3 billion by 2028, with a compound annual growth rate of 15.5% from 2021 to 2028. This rapid growth underscores the importance of accurate and efficient traffic density prediction systems for real-time traffic management. These technologies enable automobiles to communicate with one another (V2V) and interact with road infrastructure (V2I). This action minimizes congestion, boosts road safety, and advances environmentally friendly transportation.

VANETs have implications for improving traffic management and boosting road safety. They are vital for the ITS to enable communication between vehicles and roadside devices. This promotes the development of an integrated traffic sys- tem that adapts to changing traffic conditions. Nevertheless, accurately predicting traffic density can be challenging due to the random and continuously changing conditions of traffic patterns. Thus, the gap this research addresses is the need for a more reliable and accurate method for predicting traffic density in the highly variable and complex settings of VANETs, where existing models either overfit or struggle with real-time data processing.

Various machine learning models have been employed to improve the accuracy of traffic density predictions. Linear Regression (LR), despite its simplicity and interpretability, is inadequate for capturing the non-linear and dynamic patterns that exist in real-world traffic data. Decision Trees (DT) exhibit greater adaptability to sophisticated traffic scenarios, yet they are at risk for overfitting, weakening their reliability for real-time predictions. Random Forest (RF) addresses this problem by integrating many decision trees, hence enhancing accuracy and generalizability [1]. Support Vector Regression (SVR) effectively manages high-dimensional data, whereas K-Nearest Neighbours (KNN) demonstrates ability in real-time adaption. Nonetheless, RF frequently demonstrates superior accuracy in forecasting urban congestion [2].

Numerous studies have examined hybrid models to improve prediction accuracy in VANETs, showing notable enhancements over traditional methods. Therefore, researchers could enhance performance by integrating RF with optimization methods, such as metaheuristic algorithms, thereby reducing computing complexity. These hybrid models are optimal for real-time traffic control according to their balance of accuracy and efficacy. Studies have predicted traffic accidents employing Bayesian Optimisation (BO) in association with Random Forest (RF). In addition, traffic flow predictions have been refined by Support Vector Regression (SVR) reinforced by Genetic Algorithm (GA) and RF [3], [4].

Furthermore, hybrid model techniques, such as RF optimization utilizing Simulated Annealing (SA), have succeeded in various domains. One study [5] employed this approach for predicting student performance, enhancing accuracy by optimizing the number of variables and tree sizes. Another investigation [6] utilized the SA-RF model to estimate groundwater potential, surpassing the conventional RF model by decreasing prediction error by 9%. Building on the promising results of previous research, which highlighted the potential of hybrid machine learning models [7], [8], this study proposes integrating SA and RF to enhance prediction accuracy. To address the complex, dynamic nature of VANET traffic prediction, which conventional models struggle to obtain accurately, this study leverages a hybrid SA and RF approach. SA offers an effective method to optimize the hyperparameters of RF, reducing overfitting and improving model generalization, particularly in dynamic environments like VANETs. This hybrid approach aims to provide a more robust and reliable real-time traffic density prediction solution.

II. Related Work

A. Traffic Density Prediction in VANETs

Several techniques are applied to predict traffic density, including different machine learning models with edges and drawbacks for use in VANETs. LR: One of the simplest methods, LR is frequently used because it is easy to establish. However, the main drawback is that it assumes a linear effect between input features and traffic volume, which may not always be accurate in practical traffic data. Hence, it generally produces unsatisfactory predictions since realistic VANET traffic data exhibit non-linear and complex patterns, especially during highly dynamic traffic situations [9].

DT offers a more adaptable approach than LR, as they partition data into branches based on feature values, handling non-linear relationships more effectively. But DT are at risk for to overfitting, especially for deeper and more complicated trees, which inhibits their generalization and constrains their reliability for real-time traffic predictions in VANETs [2]. However, RF can be computationally intensive, making it less suitable for large-scale real-time VANET systems, and its complexity limits interpretability [7].

SVR and KNN are widely used machine learning techniques for traffic density prediction. SVR excels at managing high-dimensional data and modelling non-linear relationships. Still, its high computational cost, particularly with the large datasets typical in VANET systems, limits its practicality for real- time applications [10]. Similarly, KNN is straightforward to implement and works well for real-time predictions, but it faces scalability issues because it retains and processes the entire dataset for each prediction. Moreover, KNN's sensitivity to noise and challenges in hyperparameter selection can lead to inconsistent outcomes [11].

While each machine learning approach has advantages, its limitations emphasize the need for more efficient, scalable, and interpretable models in VANETs. Given the complexity and dynamic conditions of traffic, hybrid models such as RF, combined with optimization techniques like SA, are emerg- ing as promising solutions to address the shortcomings of existing methods. These models seek to balance computational efficiency with predictive accuracy, making them particularly suitable for real-time traffic management.

B. Simulated Annealing in Optimization

SA is recognized as a metaheuristic optimization method commonly engaged in complex and computationally demanding problems because of its capacity to overcome local minima and identify near-optimal solutions. In VANETs, SA has enhanced network parameters and protocols for better performance. The research in [12] provided a ”slow heat”-based SA appeal in VANETs to boost packet delivery ratios and minimize end-to-end delay in routing protocols. By grad- ually cooling the system, this approach mimics the physical annealing process, enabling the algorithm to make increasingly optimized routing decisions in the highly dynamic and unpredictable environment of VANETs. The progressive cooling mechanism ensures more effective convergence, resulting in improved network performance, particularly regarding reliability and speed. The study showed that the SA-based routing algorithm significantly outperformed traditional protocols like Geographic Source Routing (GSR) on main metrics, such as delay and packet loss, offering a more robust solution for real-time vehicular communication.

Over the past years, some research has been working on combining SA with machine-learning techniques such as RF to enhance the prediction model by improving accuracy and efficiency in VANETs. By combining many decision trees, RF can effectively handle non-linear data and reduce overfitting. However, it often performs poorly in feature extraction and fine-tuning of hyperparameters—two factors that limit the predictability of RF models in volatile situations like VANETs. SA has slowly but surely been employed to tune RF settings to patch these gaps. For example, determining the number of trees in a model, the depth limitation on the tree, and the importance of its features are all essential tasks RF needs to accomplish before being ready to predict traffic density.

C. Hybrid Approaches in Machine Learning

Hybrid machine learning approaches, especially the combination of SA and RF, have become popular because they balance predictive accuracy and computational efficiency. Simulated Annealing, an optimization technique inspired by the annealing process in metallurgy, excels in hyperparameter tuning and feature selection, helping to navigate complex solution spaces. While highly accurate, RF often struggles with non-linear relationships and can overfit if not properly tuned. By combining SA with RF, researchers can fine-tune the model to handle complex traffic patterns in VANETs, significantly improving accuracy and reducing computational overhead.

Although neither SA nor RF is ideal when used alone, their combination allows researchers to leverage their strengths and overcome limitations like small datasets and complex hyperparameter optimization. SA helps explore a diverse parameter space and avoids local minima, enhancing RF's robustness in predicting variable traffic patterns. Moreover, machine learning models like RF offer better interpretability, which is essential for improving traffic predictions in ITSs.

A hybrid SA-RF model was introduced by [5] to predict student performance. SA optimized RF's parameters through binary encoding of feature numbers, tree size, and decision weights, improving generalization ability. Similarly, [13] applied the SA algorithm to improve RF classification accuracy by optimizing global and local search processes. The success of these hybrid models across various domains suggests their suitability for traffic density prediction in VANETs, where they offer improved accuracy and efficiency, which is ideal for real-time applications in dynamic environments.

In [14], it was shown that the hybrid method of integrat- ing SA with RF enhances model performance by mitigating RF's weaknesses in parameter optimization. SA's global optimization ability allows for more efficient parameter space exploration compared to grid search or random search, overcoming local minima and improving model generalization. This integration enables the RF algorithm to maintain its predictive effectiveness while SA promotes overall efficiency and accuracy through the optimization of the feature selection process. The research indicated that the SA-RF hybrid model substantially surpassed conventional RF methods regarding prediction accuracy and computing efficiency, rendering it an optimal solution for traffic flow prediction in VANET systems.

III. Proposed Methodology

The proposed phase is illustrated in Fig. 1. The flowchart illustrates the structured logic employed for predicting traffic density through several machine learning models. The procedure begins with data acquisition from Data Port-IEEE, followed by preprocessing that encompasses data partitioning, normalization, and the management of absent data. Predictions are generated utilizing many models, including Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN). The models are assessed utilizing performance indicators, including MSE, RMSE, MAE, MAPE, and R^2. If additional optimization is demanded, the optimal model is maintained in its original state or adjusted by Simulated Annealing (SA). Prior to optimization, the model conducts re-evaluation to figure out if the original or enhanced version exhibits superior performance, leading to a final recommendation.

A. Data Collection and Processing

This study utilizes the dataset from the IEEE DataPort repository [15] to analyze traffic data for Vehicular Ad-hoc Networks (VANETs). The dataset contains essential factors, including the Number of Current Vehicles (target variable), Average Speed, Number of Joining Vehicles, Joining Frequency, and Bloom Filter (designated for feature variables). The data is processed employing the Orange Data Mining platform, where preprocessing encompasses addressing missing data, standardizing variables, and partitioning the dataset into training and testing sets. The Number of Current Vehicles serves as the aim for predictive modeling, whereas the other variables function as features for predicting traffic density. After preprocessing, the data is input into machine learning models, including LR, DT, RF, SVR, and KNN.

B. Baseline Machine Learning Models

This study evaluates the efficacy of several baseline machine learning models in predicting traffic density utilizing the dataset refined in the preceding phase. The models chosen for comparison include LR, DT, RF, SVR, and KNN. These models encompass many predictive strategies, from basic linear models to more complex ensemble and kernel-based methods, offering insights into the effectiveness of each model in predicting traffic density within VANETs. This study did not consider deep learning models due to their high computational cost and the need for larger datasets, making them unsuitable for real-time VANET applications.

- Linear Regression (LR):
  This is the most basic model employed for predicting traffic density. LR suggests a linear correlation between the dependent variable (Number of Current Vehicles) and the independent variables (e.g., Average Speed, Number of Joining Vehicles). Although easily interpretable, linear regression frequently encounters difficulties with non-linear data, a common problem in traffic prediction.
- Decision Tree (DT):
  DT is a non-linear model that separates data into decision nodes according to feature values. It performs in non-linear interactions but is vulnerable to overfitting, particularly in complex settings such as VANETs, where traffic conditions fluctuate fast.
- Random Forest (RF):
  RF is an ensemble technique that employs several decision trees to enhance predictive accuracy and mitigate overfitting. It is recognized for its efficiency in managing high-dimensional data, rendering it appropriate for traffic density prediction, where numerous features affect the result.
- Support Vector Regression (SVR):
  SVR employs kernel functions to address non-linear correlations within the data. It is highly successful at modeling sophisticated traffic patterns; yet, it can be computationally intensive, particularly when managing enormous datasets.
- K-Nearest Neighbors (KNN):
  KNN is a simple, non-parametric technique that predicts traffic density by analyzing nearest data points. Although it is efficient for real-time predictions, its efficacy might reduce with large and high-dimensional datasets because of the necessity to calculate distances for each prediction.

Every model will be assessed according to established performance metrics. This comparison will facilitate the identification of the most appropriate model for traffic density prediction in VANETs, optimizing accuracy, computational efficiency, and generalizability to novel data.

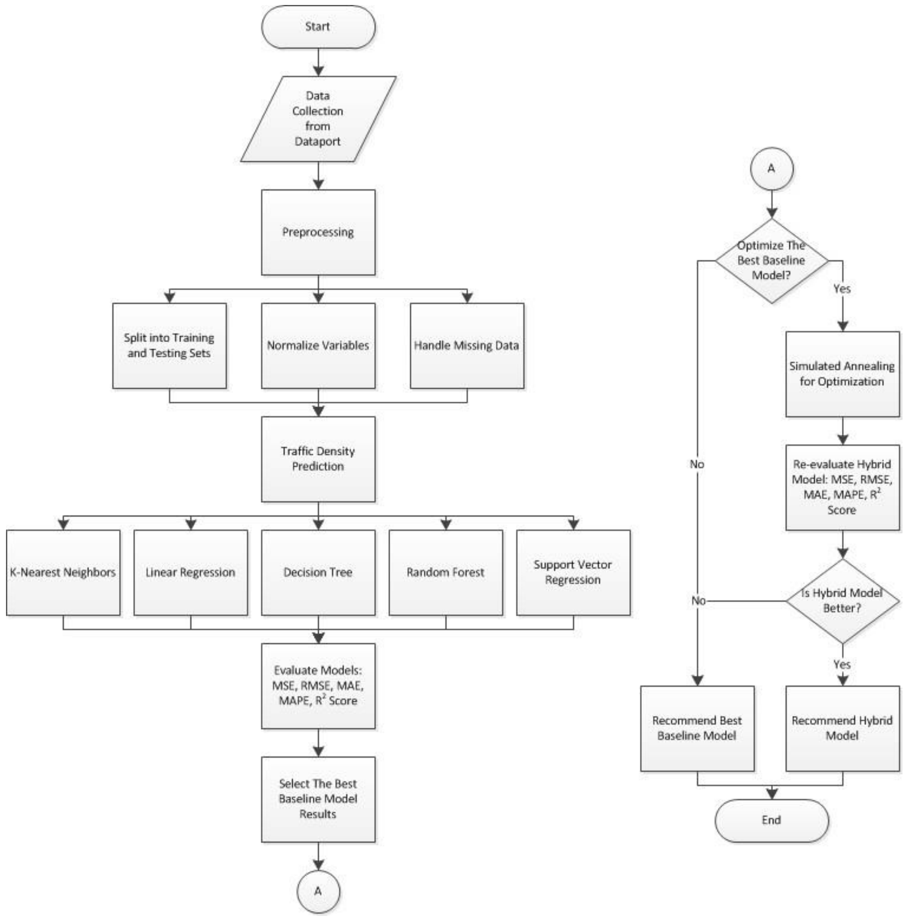


Fig. 1: Flowchart of the proposed research.

C. *Hybrid Model*

This section introduces the Simulated Annealing (SA) method for optimizing the parameters of a previously identified best-performing machine learning model. SA is a probabilistic optimization technique inspired by the annealing process in metallurgy, where materials are heated and slowly cooled to minimize defects and stabilize. Similarly, SA explores the solution space to find the global minimum of an optimization problem, refining the solution to avoid being trapped in local minima.

SA works by generating new solutions and accepting them based on changes in the objective function and a probability that decreases as the system cools. Early in the process, the algorithm accepts less optimal solutions, allowing it to explore more solution space and reducing the chance of getting stuck in local minima. In this case, the objective function is the error metric (e.g., Mean Squared Error, MSE) of the baseline model, and SA seeks to minimize this error by optimizing the model's hyperparameters (e.g., number of trees in RF, regularization parameters in SVR).

The acceptance probability of a new solution is given by the (1):

$$P(\Delta E) = \exp\left(-\frac{\Delta E}{T}\right) \tag{1}$$

Where:

- ΔE is the change in the objective function (i.e., error or MSE) between the current and new solutions.
- *T* is the temperature, which decreases over time according to a cooling schedule.
- *P*(ΔE) is the probability of accepting a worse solution; as the temperature decreases, the probability of accepting worse solutions also decreases.

The SA process begins by raising the temperature, facilitating extensive exploration of possible solutions. As the temperature reduces, the algorithm exhibits more selectivity, concentrating on enhancing the optimal solution identified to the present. This procedure optimizes the hyperparameters of the baseline machine learning model to improve its accuracy in predicting traffic density. The SA optimization reconciles the aim of a global minimum with the refinement of the model for optimal performance, offering a highly effective prediction model for VANET traffic density.

D. *Performance Metrics*

A variety of performance metrics were used to evaluate the effectiveness of the machine learning models and the hybrid SA-optimized model in predicting traffic density. These metrics comprehensively assess the model's predictive accu-

Algorithm 1 Simulated Annealing Optimization for Vehicle Density Prediction

Input:

Dataset D (Vehicle data: *Average Speed, Number of Joining Vehicles, Joining Frequency, Bloom Filter*)

Parameter space P (Ranges for $n_{\text{estimators}}$, max_depth, min_samples_split, min_samples_leaf)

Initial temperature T_0, Cooling rate α, Min temperature $T_{\min}$, Max iterations $N_{\max}$

Output:

Optimal parameters P_{best}, Minimum MSE MSE_{best}

Initialization:

Randomly select initial parameters P_{current} within the parameter space P

Set initial temperature $T \leftarrow T_0$

Evaluate initial MSE $\text{MSE}_{\text{current}} \leftarrow f(P_{\text{current}})$

Set $P_{\text{best}} \leftarrow P_{\text{current}}$, $\text{MSE}_{\text{best}} \leftarrow \text{MSE}_{\text{current}}$

SA Optimization (Equation 1)

Return: Optimal parameters P_{best}, Minimum MSE MSE_{best}

racy, robustness, and reliability in real-world scenarios. The performance metrics considered are,

- Mean Squared Error (MSE)
  MSE is a widely used performance metric in regression models, representing the average squared differences between predicted and actual values. Because of the squaring of the errors, the MSE emphasizes larger errors over smaller ones, making it particularly sensitive to outliers. It aids in comprehending the average error magnitude despite being in squared units. The equation for the MSE is delineated in (2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

- Root Mean Squared Error (RMSE)
  RMSE measures the square root of the average squared differences between the predicted and actual values. It indicates how spread out the residuals (errors) are, which helps identify large deviations from the actual traffic density. It is defined as equation (3).

$$\sqrt{\frac{1}{n}\sum}$$

is less sensitive to outliers than RMSE. The formula is shown in (4).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (4)$$

- Mean Absolute Percentage Error (MAPE)
  MAPE quantifies prediction accuracy as a percentage, facilitating comprehension of the error magnitude in relation to the actual values. The subsequent (5) is computed for MAPE.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i} \quad (5)$$

  MAPE is especially advantageous for analyzing performance in practical scenarios, as it presents errors in percentage format, easing understanding.
- R^2 Score (Coefficient of Determination)
  The R^2 score indicates the proportion of variance in the dependent variable (traffic density) that can be predicted from the independent variables. It spans from 0 to 1, with 1 signifying an optimal fit and 0 indicating that the model does not account for any variability in the data. It is computed as (6).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

  Where $\bar{y}$ is the mean of the actual values.

These metrics will offer a comprehensive assessment of the model's performance, verifying that it is both accurate and reliable for traffic density prediction. MSE, RMSE, and MAE provide insight into error size, whereas R² emphasizes the model's capacity to clarify data variability. Finally, MAPE offers a practical, interpretable metric for real-world application.

IV. Results and Discussion

A. Comparative Analysis

The result obtained from the simulation on Orange Data Mining shows that the distribution of five models varies. Shown in Fig. 2 is the combined visualization comparing the predictions of all five models (LR, DT, RF, SVR, and KNN) against the actual traffic density values. Each model's predictions are represented by a different color, with the ideal prediction line in black.

Moreover, the result of five model predictions is presented in Table I. Based on the comparison of the metrics table, the RF Regression and KNN Regression models stand out due to

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (3)$$

Where y_i is the actual value, $\hat{y}_i$, and n is the number of observations.

- Mean Absolute Error (MAE)
  MAE calculates the average magnitude of the errors in a set of predictions without considering their direction (i.e., whether the predictions are overestimations or underestimations). It provides a linear measure of accuracy and

TABLE I: Prediction Model Metrics

Model	Performance Metrics				
	MSE	***RMSE***	***MAE***	***MAPE***	***R^2 Score***
LR	669.747	25.879	17.431	1.896	0.498
DT	615.056	24.800	14.138	0.973	0.539
RF	**361.284**	**19.007**	**10.836**	0.813	**0.729**
SVR	1133.898	33.673	25.843	3.904	0.151
KNN	364.743	19.098	11.140	**0.573**	0.727

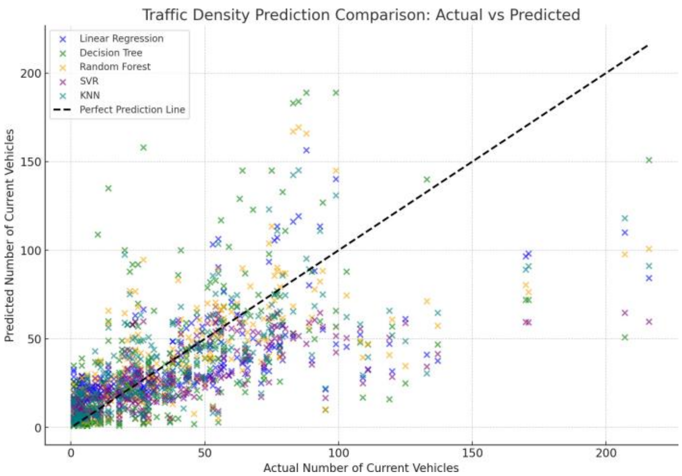


Fig. 2: Traffic density prediction comparison: Actual vs Predicted (Linear Regression, Decision Tree, Random Forest, Support Vector Regression, and K-Nearest Neighbors).

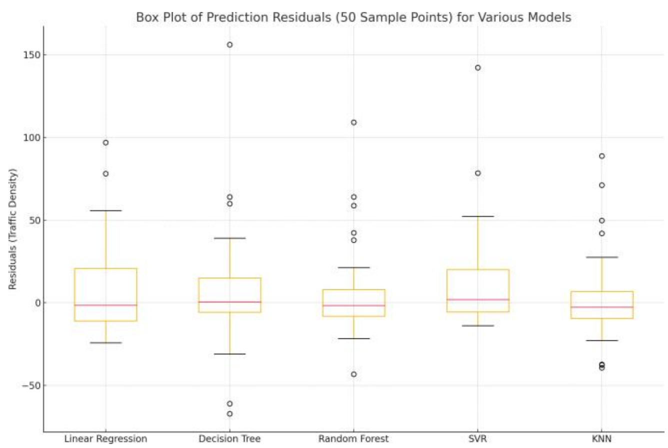


Fig. 3: Box plot of prediction residuals for five prediction models.

their lower MSE, RMSE, and MAE values. Specifically, RF has the lowest MSE (361.284), RMSE (19.007), and MAE (10.836), signifying its outstanding results in the prediction assignment. KNN achieves an MSE of 364.743, an RMSE of 19.098, and an MAE of 11.140. KNN exhibits the optimal MAPE of 0.573, providing it an acceptable alternative. RF exhibits the greatest R^2 Score (0.729), signifying the better ability to compensate for variance in the data. This implies that RF and KNN are the most effective models in the present case. A comprehensive review of the box plot of prediction residuals depicted in Fig. 3 reveals that RF has outstanding results, characterized by a minimal Interquartile Range (IQR). A small IQR indicates that most residuals (errors) are around the median. The median corresponds to zero, signifying that its predictions are aligned with the actual numbers. RF exhibits fewer outliers than alternative models, indicating its ability to manage extreme cases. RF is the superior model due to its reduced and more consistent residuals.

TABLE II: Prediction Model Metrics

Parameters	Result
Best Solution (Row Index)	1235
Best Score (Number of Vehicles)	25
Corresponding Data Row:	
Average Speed	14.215
Number of Joining Vehicles	17
Joining Frequency	0.519
Bloom Filter	3.891
Number of Current Vehicles	25

B. Feature Importance and Optimization

Upon obtaining or recommending that RF offers the optimal results, the subsequent stage will involve executing the optimization process by using the SA algorithm. Using (1) and the SA algorithm, the optimization process starts with an initial temperature of 1000, which decreases by 5% per iteration (cooling rate of 0.95) until it is below one or after 100 iterations. The parameter space includes hyperparameters such as number of trees (50-200), tree depth (5-50), minimum samples to split (2-10), and minimum samples per leaf (1-4). The algorithm explores these hyperparameters within the specified range, optimizing the model for better performance while gradually refining the solution as the temperature decreases. The optimal solution's result is at row 1235 in the dataset, as outlined in Table II. This solution presents a balanced system that maintains a vehicle count of 25 while obtaining a moderate flow rate, illustrating the effectiveness of the SA algorithm in optimizing traffic dynamics within the specified parameters.

The reassessment of hybrid prediction model metrics (Table III) reveals that the SA-RF model substantially improves error reduction. The decreased MSE and RMSE demonstrate that the predictions from SA-RF are consistently closer to the actual values than those from standard RF. The reduction in MAE indicates that the mean magnitude of prediction errors has dropped for the SA-RF model. Although the SA-RF model improves most error metrics, the MAPE is slightly worse than the baseline RF model (0.822 vs. 0.813). This suggests that the SA-RF model might still be struggling with percentage- based errors, which could be caused by small values in the target variable disproportionately affecting the MAPE. The R^2 value for SA-RF is improved (0.771 compared to 0.729), indicating that the SA-RF model handles the more significant variance in the target factor, implying a more acceptable match overall. The improvement in R^2, along with the decreases in error measures (MSE, RMSE, and MAE), signifies that the

TABLE III: Re-evaluate hybrid prediction model metrics.

Metric	RF	SA-RF	Improvement
MSE	361.284	305.772	SA-RF improved
RMSE	19.007	17.486	SA-RF improved
MAE	10.836	10.235	SA-RF improved
MAPE	0.813	0.822	RF slightly better
R^2	0.729	0.771	SA-RF improved

SA-RF model offers better predictions.

The hybrid SA-RF model surpasses conventional methods by taking advantage of the strengths of both techniques. RF is proficient in handling complex, high-dimensional, and non-linear traffic data; nonetheless, it frequently encounters challenges in appropriate hyperparameter selection, potentially resulting in overfitting or underfitting. SA cautiously examines the hyperparameter space and optimizes parameters such as the number of trees and tree depth, assisting the RF model in evading poor solutions and enhancing generalisation to unexpected data.

Furthermore, SA improves the RF model by facilitating a more comprehensive search of the solution space, employing a probabilistic method to prevent local minima, and optimizing the model's accuracy. This combination enhances predictive accuracy while optimizing computing efficiency, making the hybrid model more appropriate for real-time applications in dynamic contexts such as VANETs. The result is a model that more effectively captures the non-linear correlations in traffic data, offering enhanced prediction performance.

C. Limitation

Current research on traffic density prediction in VANETs using a hybrid model of RF and SA shows promising results but faces several challenges. While RF-SA is computationally lighter than deep learning, it still faces scalability challenges in large, dynamic networks, potentially causing delays. Additionally, unoptimized RF models may overfit complex traffic data, limiting their generalizability in unpredictable scenarios [2]. Even though the hybrid approach improves accuracy and reduces errors, its increased complexity poses challenges for real-time applications due to the high computational demands and dynamic nature of VANETs [1].

V. Conclusion

This research concludes that a hybrid model integrating RF and SA has been successfully implemented for traffic density prediction in VANETs. The research illustrates that this hybrid methodology markedly enhances predictive accuracy through hyperparameter optimization and overfitting mitigation, making it preferable to conventional models such as LR and standalone RF. Nonetheless, issues persist, especially with computational complexity and real-time application, due to the highly dynamic characteristics of VANET systems.

Future studies could enhance the model's efficiency and accuracy by including advanced optimization approaches or real-time adaption mechanisms. Furthermore, integrating additional lightweight machine learning techniques, investigating various metaheuristic algorithms, or refined feature selection techniques may provide alternate answers to the constraints identified in real-time predictions. Evaluating this hybrid model in extensive, real-world VANET settings would provide a more comprehensive insight into its practical applications and potential for expansion.

Acknowledgment

This research is carried out as part of the final project for the Bachelor of Applied Mechatronics Engineering program at Politeknik Negeri Batam. This study was also endorsed by the Research and Community Service Center of Politeknik Negeri Batam. We extend our gratitude to all who contributed assistance and encouragement throughout the completion of this research.

References

[1] H. Wang, X. Wei, J. Yao, and Y. Zhang, "Traffic Flow Prediction Using Machine Learning Methods," in 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Dec. 2021, pp. 30–35.

[2] C. Chaoura, H. Lazar, and Z. Jarir, "Predictive System of Traffic Congestion based on Machine Learning," in 2022 9th International Conference on Wireless Networks and Mobile Communications (WINCOM), Oct. 2022, pp. 1–6.

[3] M. Yan and Y. Shen, "Traffic Accident Severity Prediction Based on Random Forest," Sustainability, vol. 14, no. 3, Art. no. 3, Jan. 2022.

[4] L. Zhang, N. R. Alharbe, G. Luo, Z. Yao, and Y. Li, "A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction," Tsinghua Sci. Technol., vol. 23, no. 4, pp. 479–492, Aug. 2018.

[5] S. Huang and J. Wei, "Student Performance Prediction in Mathematics Course Based on the Random Forest and Simulated Annealing," Sci. Program., vol. 2022, no. 1, p. 9340434, 2022.

[6] B. Choubin and O. Rahmati, "20 - Groundwater potential mapping using hybridization of simulated annealing and random forest," in Water Engineering Modeling and Mathematic Tools, P. Samui, H. Bonakdari, and R. Deo, Eds., Elsevier, 2021, pp. 391–403.

[7] S. S. Sepasgozar and S. Pierre, "Network Traffic Prediction Model Considering Road Traffic Parameters Using Artificial Intelligence Methods in VANET," IEEE Access, vol. 10, pp. 8227–8242, 2022.

[8] P. A. D. Amiri and S. Pierre, "An Ensemble-Based Machine Learning Model for Forecasting Network Traffic in VANET," IEEE Access, vol. 11, pp. 22855–22870, 2023.

[9] H. Wang, X. Wei, J. Yao, and Y. Zhang, "Traffic Flow Prediction Using Machine Learning Methods," presented at the 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), IEEE Computer Society, Dec. 2021, pp. 30–35.

[10] J. Tong, X. Gu, M. Zhang, J. Wan, and J. Wang, "Traffic Flow Prediction Based on Improved SVR for VANET," in 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Mar. 2021, pp. 402–405.

[11] F. LI et al., "Different Traffic Density Connectivity Probability Analysis in VANETs with Measured Data at 5.9 GHz," in 2018 16th International Conference on Intelligent Transportation Systems Telecommunications (ITST), Oct. 2018, pp. 1–7.

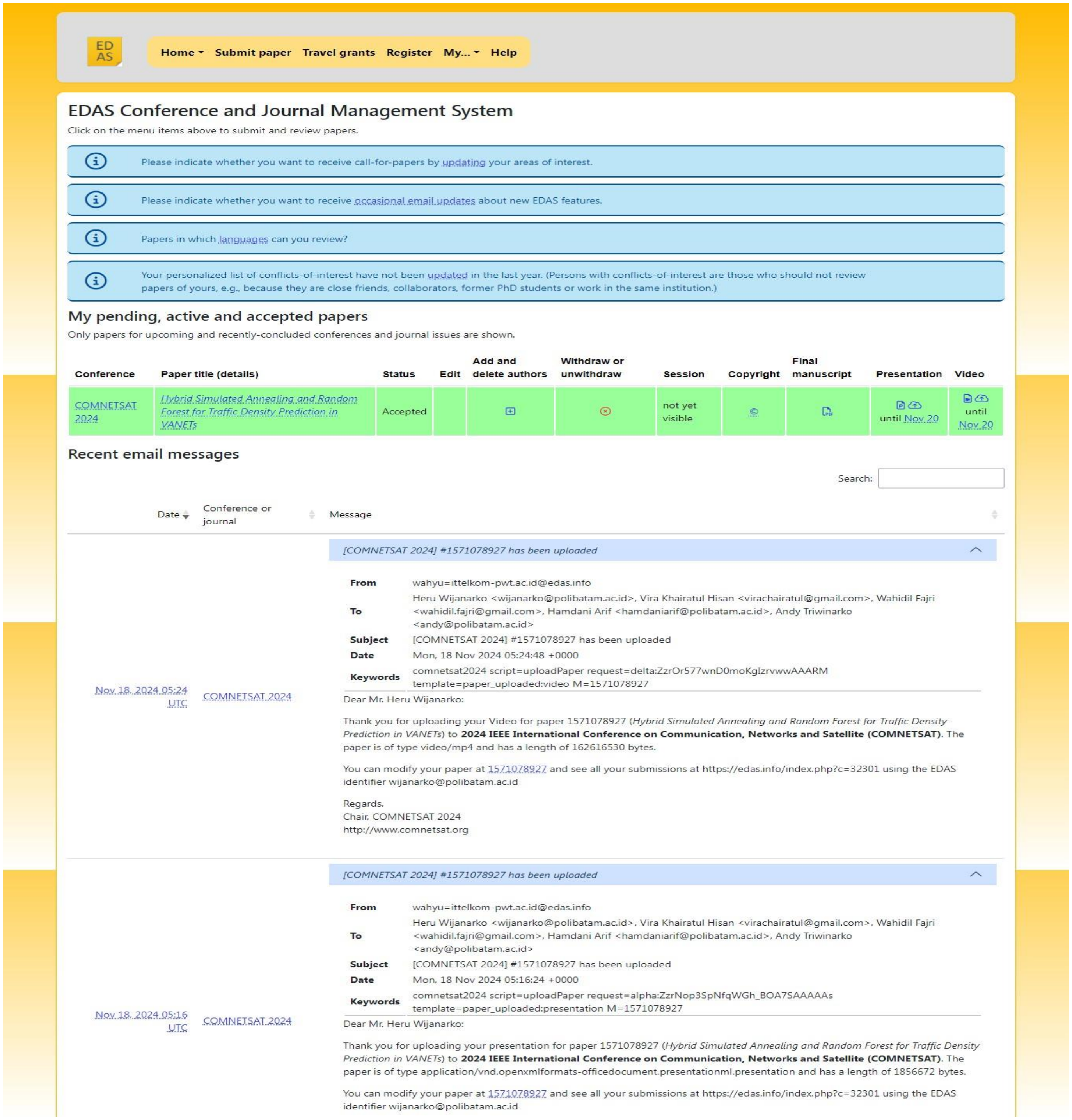
[12] P. K. Pagadala and N. M. Saravana Kumar, "Slow Heat Based Simulated Annealing and Quality Aware Routing in Vehicular Ad Hoc Network," Wirel. Pers. Commun., vol. 118, no. 4, pp. 3063–3073, Jun. 2021.

[13] Z. Zhou, Y. Wang, X. He, and X. Zhang, "Optimization of Random Forests Algorithm Based on ReliefF-SA," IOP Conf. Ser. Mater. Sci. Eng., vol. 768, no. 7, p. 072065, Mar. 2020.

[14] P. K. Pagadala et al., "Slow Heat-Based Hybrid Simulated Annealing Algorithm in Vehicular Ad Hoc Network," Comput. Intell. Neurosci., vol. 2023, no. 1, p. 9918748, 2023.

[15] R. M. Das, "Data set for Vehicular Network." IEEE, Jan. 09, 2023. Accessed: Sep. 01, 2024. [Online]. Available: https://ieee-dataport.org/documents/data-set-vehicular-network

Lampiran



ED AS

Home ▾ Submit paper Travel grants Register My... ▾ Help

EDAS Conference and Journal Management System

Click on the menu items above to submit and review papers.

Please indicate whether you want to receive call-for-papers by updating your areas of interest.

Please indicate whether you want to receive occasional email updates about new EDAS features.

Papers in which languages can you review?

Your personalized list of conflicts-of-interest have not been updated in the last year. (Persons with conflicts-of-interest are those who should not review papers of yours, e.g., because they are close friends, collaborators, former PhD students or work in the same institution.)

My pending, active and accepted papers

Only papers for upcoming and recently-concluded conferences and journal issues are shown.

Conference	Paper title (details)	Status	Edit	Add and delete authors	Withdraw or unwithdraw	Session	Copyright	Final manuscript	Presentation	Video
COMNETSAT 2024	*Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs*	Accepted				not yet visible	©		until Nov 20	until Nov 20

Recent email messages

Search:

Date	Conference or journal	Message
Nov 18, 2024 05:24 UTC	COMNETSAT 2024	*[COMNETSAT 2024] #1571078927 has been uploaded*
Nov 18, 2024 05:16 UTC	COMNETSAT 2024	*[COMNETSAT 2024] #1571078927 has been uploaded*

[COMNETSAT 2024] #1571078927 has been uploaded

From wahyu=ittelkom-pwt.ac.id@edas.info

To Heru Wijanarko <wijanarko@polibatam.ac.id>, Vira Khairatul Hisan <virachairatul@gmail.com>, Wahidil Fajri <wahidil.fajri@gmail.com>, Hamdani Arif <hamdaniarif@polibatam.ac.id>, Andy Triwinarko <andy@polibatam.ac.id>

Subject [COMNETSAT 2024] #1571078927 has been uploaded

Date Mon, 18 Nov 2024 05:24:48 +0000

Keywords comnetsat2024 script=uploadPaper request=delta:ZzrOr577wnD0moKgIzrvwwAAARM template=paper_uploaded:video M=1571078927

Dear Mr. Heru Wijanarko:

Thank you for uploading your Video for paper 1571078927 (*Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs*) to **2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)**. The paper is of type video/mp4 and has a length of 162616530 bytes.

You can modify your paper at 1571078927 and see all your submissions at https://edas.info/index.php?c=32301 using the EDAS identifier wijanarko@polibatam.ac.id

Regards,
Chair, COMNETSAT 2024
http://www.comnetsat.org

[COMNETSAT 2024] #1571078927 has been uploaded

From wahyu=ittelkom-pwt.ac.id@edas.info

To Heru Wijanarko <wijanarko@polibatam.ac.id>, Vira Khairatul Hisan <virachairatul@gmail.com>, Wahidil Fajri <wahidil.fajri@gmail.com>, Hamdani Arif <hamdaniarif@polibatam.ac.id>, Andy Triwinarko <andy@polibatam.ac.id>

Subject [COMNETSAT 2024] #1571078927 has been uploaded

Date Mon, 18 Nov 2024 05:16:24 +0000

Keywords comnetsat2024 script=uploadPaper request=alpha:ZzrNop3SpNfqWGh_BOA7SAAAAAs template=paper_uploaded:presentation M=1571078927

Dear Mr. Heru Wijanarko:

Thank you for uploading your presentation for paper 1571078927 (*Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs*) to **2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)**. The paper is of type application/vnd.openxmlformats-officedocument.presentationml.presentation and has a length of 1856672 bytes.

You can modify your paper at 1571078927 and see all your submissions at https://edas.info/index.php?c=32301 using the EDAS identifier wijanarko@polibatam.ac.id

Regards,
Chair, COMNETSAT 2024
http://www.comnetsat.org

Nov 18, 2024 02:30 UTC — COMNETSAT 2024

[COMNETSAT 2024] Your paper #1571078927 ('Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs')

From	comnetsat=ittelkom-pwt.ac.id@edas.info
To	Heru Wijanarko <wijanarko@polibatam.ac.id>, Vira Khairatul Hisan <virachairatul@gmail.com>, Wahidil Fajri <wahidil.fajri@gmail.com>, Hamdani Arif <hamdaniarif@polibatam.ac.id>, Andy Triwinarko <andy@polibatam.ac.id>
Subject	[COMNETSAT 2024] Your paper #1571078927 ('Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs')
Date	Mon, 18 Nov 2024 02:30:21 +0000
Keywords	comnetsat2024 script=sendOneEmail request=alpha:ZzqmvCuBsT3WVrm8DYhW2QAAANY template= M=1571078927

Dear authors:

[Reminder]

We hope you are doing well. As part of the preparations for COMNETSAT 2024, we are conducting pre-conference activities to ensure smooth and high-quality presentation sessions during the event.

To facilitate this, we kindly request you to submit a recording of your presentation.

Pre-recorded Presentation Video Instruction:

Online attendees of the COMNETSAT 2024 are required to produce a video presentation of their work and submit it to EDAS. Participants who attend the event physically are not mandatory. The following are the prerequisites for the video presentation:

1. Please use the .pptx template provided by the committee (https://docs.google.com/presentation/d/1KMTVAiMw6oGRtxO-V89hX4gEfX95oU6UiVZ5CmqgzkQ/edit#slide=id.p).
2. The maximum duration of presentation is 10 minutes.
3. Display the slides and the face of the presenter, who is one (or some) of the authors of the paper (However, only one presenter allowed in the live discussion session).
4. Start with a short introduction of the presenter.
5. The maximum video size is 200MB.
6. The video must have sufficiently good resolution.
7. The video format is an MP4 file.
8. The video file name is the paper number. Ex: 1571012345.

You can record audio and video straight from your computer presentation and microphone using a variety of presentation software applications, such as Zoom and PowerPoint.

The presentation video is uploaded by authors via EDAS.

The session chairs will review all pre-recorded presentations to ensure that the content is high-quality and pertinent. The Chairs have the authority to reject the submission if the presenter is not an author or if the presentation does not accurately represent the accepted Final Manuscript.

The recorded presentation will be used during your 10-minute presentation session at the conference, allowing you to focus entirely on the 5-minute discussion session that follows.

Video Submission Deadline: November 20, 2024.

You can find the instruction in our website: https://comnetsat.org/submissions/. Please let us know if you encounter any issues with the recording or upload process. We appreciate your cooperation in making this event a success.

Regards,
Chair, COMNETSAT 2024
http://www.comnetsat.org

Nov 12, 2024 03:48 UTC — COMNETSAT 2024

Final Submission of Presentation and Video Materials for ComNetSat 2024

We kindly remind you to upload your presentation file and/or recorded video presentation to the EDAS system by November 20, 2024. Please note that this is a strict deadline. Failure to meet this deadline will result in the exclusion of your paper from the presentation schedule at ComNetSat 2024.

We appreciate your cooperation in ensuring a smooth conference experience for all participants.

Thank you for your attention to this matter. Should you have any questions or require assistance, feel free to reach out.

Regards,
{chairs}

[COMNETSAT 2024] #1571078927 has been uploaded

Home Register My... Help

Upload manuscript for #1571078927: Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs

Upload manuscript for #1571078927: *Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs*

The file has been uploaded.

Hide details

BibTEX

Authors	Heru Wijanarko, Vira Khairatul Hisan, Wahidil Fajri, Hamdani Arif and Andy Triwinarko (Politeknik Negeri Batam, Indonesia)
Paper title	*Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs* Only the chairs can edit
Conference and track	**2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)** - *Data Science & Artificial Intelligence*
Abstract	Only the chairs can edit The research addresses the challenge of traffic density prediction in Vehicular Ad-hoc Networks,...
Keywords	machine learning optimization; random forest; simulated annealing; traffic density prediction; vehicular ad-hoc networks Only the chairs can edit
Topics	Optimization Algorithms; Support Vector Machines (Only the chairs can edit)
Personal notes	
Roles	You are the creator, an author and a presenter for this paper. You have authored an accepted paper in this conference.
Status	Accepted
Copyright	IEEE; IEEE completed Oct 26
Registration	Heru Wijanarko has registered and paid for Authors 1 paper - Online:Aut-Student-NonMem Online
Presented	by Vira Khairatul Hisan, Wahidil Fajri and Heru Wijanarko (bio) : session to be announced

Review manuscript	Final manuscript	Presentation	Video

Review

The Research Novelty		Research Methods		Future Impact		The Clarity of Research Question		Scientific argument and discussion	
Good	3	Good	3	Adequate	2	Good	3	Good	3

Detailed comments

Please add setting parameters of SA, especially the parameters shown in eq. (1).

The Research Novelty		Research Methods		Future Impact		The Clarity of Research Question		Scientific argument and discussion	
Good	3	Good	3	Good	3	Adequate	2	Good	3

Detailed comments

1. Remove paragraph 5 in the Introduction, since it is identical to paragraph 4.
2. Please elaborate more about the motivation of your study. In section Introduction, paragraph 7, the authors mention "In view of these encouraging findings and further studies [7], [8] emphasizing the efficacy of hybrid machine learning models, we propose the integration of Simulated Annealing and Random Forest to enhance accuracy in traffic density prediction. This mixture could function as an effective solution to boost predictions in this field of study.", but it is not enough to reflect the motivation of the study.

The Research Novelty		Research Methods		Future Impact		The Clarity of Research Question		Scientific argument and discussion	
Adequate	2	Good	3	Good	3	Adequate	2	Adequate	2

Detailed comments

The abstract contains a summary of the research, including the background, problem formulation, proposed research, and results. However, the research gap is not explained.

In the introduction, it is stated that the proposed research integrates simulated annealing and random forest. How do simulated annealing and random forest improve accuracy?

In the design and methodology section, several aspects are explained, including data collection, baseline machine learning models, the hybrid model, and performance metrics. Why were approaches such as Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) chosen over deep learning approaches?

In the results and analysis section, it is shown that the performance of simulated annealing and random forest is superior to other methods. Add an analysis explaining why the proposed method performs better!

FORMULIR LOOGBOOK BIMBINGAN DAN PENGAJUAN
SEMINAR PROPOSAL/SIDANG TUGAS AKHIR*

Nama : Vira Khairatul Hisan
NIM : 4212231002
Pembimbing I : Heru Wijanarko, S.T.,M.Sc.
Judul : Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs

No	Hari/Tgl	Rincian Kegiatan	TTD Pembimbing 1
1	7/08/ 2024	Konsultasi mengenai review beberapa jurnal mengenai VANETs dan pengumpulan beberapa jurnal terkait.	
2	11/09/2024	Proses penulisan jurnal ilmiah terkait Simulated Annealing dan Baseline Machine Learning (LR,DT,RF,SVR,KNN)	
3	17/09/2024	Revisi Pembuatan jurnal ilmiah Hybrid Simulated Annealing and Random Forest for Traffic Density Prediction in VANETs	
4	27/09/2024	Finalisasi jurnal ilmiah dan proses submission ke conference ilmiah IEEE COMNETSAT	
5	07/10/2024	Jurnal ilmiah sudah di submit dan statusnya ACCEPTED	
6	10/10/2024	Pembuatan slide presentation untuk conference IEEE COMNETSAT 2024	
7	23/10/2024	Revisi jurnal ilmiah setelah di review oleh reviewer IEEE COMNETSAT 2024	
8	05/11/2024	Finalisasi pembuatan video presentasi untuk conference ilmiah IEEE COMNETSAT 2024	
9	16/11/2024	Submit video presentasi untuk conference ilmiah IEEE COMNETSAT 2024	
10	28/11/2024	Presentasi Jurnal ilmiah pada conference ilmiah IEEE COMNETSAT 2024	

Berdasarkan hasil bimbingan yang telah dilaksanakan selama 4 bulan dan telah disetujui oleh dosen pembimbing, maka dengan ini saya mengajukan diri sebagai peserta ~~Seminar Proposal~~ /Sidang Tugas Akhir*.

Batam, 28 November 2024
Peserta

Vira Khairatul Hisan
NIM: 4212231002

**Hapus yang tidak perlu.*
Jumlah bimbingan minimal 10 kali. Dalam satu minggu maksimal bimbingan yang dihitung adalah 2 kali.