

ENSEMBLE VOTING CLASSIFIER FOR PREDICTION OF TRAFFIC USING SUMO TRANSEL SCENARIOS

Mr. K. Vinoth M.E CSE
Assistant Professor
Department of computer
Science and Engineering
KSR institute for engineering and technology
Tiruchengode
kvinothcse83@gmail.com

Deepan G
Department of computer
science and Engineering
KSR institute for engineering and technology
Tiruchengode
deepangovindaraj2002@gmail.com

Prabhu M
Department of computer
science and Engineering
KSR institute for engineering and technology
Tiruchengode
mprabhu28320@gmail.com

Manojkumar M
Department of computer
science and Engineering
KSR institute for engineering and technology
Tiruchengode
manojkumar22042002@gmail.com

Ruthresan V
Department of computer
science and Engineering
KSR institute for engineering and technology
Tiruchengode
ruthreshvelu2002@gmail.com

ABSTRACT

One of the most significant issues in India is congestion in traffic, which is especially prevalent in some of the nation's major cities. For example, packed roadways serve as a stark reminder of the lodge's awfulness. Since streets are frequently free for the taking, there is barely any incentive for cars to use them correctly, resulting in to traffic jams whenever interest exceeds the ability to pay for it. Both street estimating and the privatization of interstates have been suggested as approaches to reduce traffic through financial disincentives and rewards. Infrequent parkway events, such as an accident or roadwork, can also cause blockage and lower the street's capability below normal levels. While congestion is a possibility for every means of transportation, the majority of the systems focused on automobile obstruction on open streets. Techniques for image analysis have been utilized extensively in traffic framework supervision and control. This paper suggests an alternative methodology, an algorithm, that would help distribute the traffic equitably while controlling the signal by utilising HERE maps API in order to get rid of the excess and unsuitability of these picture processing frameworks.

Keywords—Traffic control, traffic clog, congestion, api, traffic signal algorithm, machine learning.

1.

INTRODUCTION

India took great pride in being the second-largest street organisation in the world. The astonishing 5.4 million millimeters still left in the Indian street networks! As a result, it forms a massive final recommendation for the Indian Government to provide immaculate streets at each stage. Passing through the Indian lanes is undoubtedly a problem that no Indian, whether traditional or modern, may like to confront. An unbelievable traffic jam happens as demand approaches a street's capacity (or the crossroads along the street). A traffic jam or (colloquially) a traffic growl up occurs when all moving cars come to a complete stop for an extended period of time. Drivers who encounter traffic obstructions may become perplexed and engage in street rage. In numbers, a clog is normally considered of as the number of vehicles that pass through a particular location within a specific period of time or a stream. Some of the common traffic issues include: Bad street quality due to excessive traffic-The absurd clogging of urban streets due to heavily used private vehicles results in the deterioration of the nature of the streets. This frequently results in constant traffic issues. The sheer enormity of traffic difficulties also gives birth to additional health-damaging problems including the air and sound pollution, particularly in urban areas. Hence, we suggest a solution that dynamically manages traffic based on a number

of key variables, including the time of day, the weather, the state of the roads, etc. The technique makes it possible to spread the area's traffic congestion fairly.

2. LITERATURE REVIEW

In [1] Tawfeek et al., "Deep Learning for Urban Traffic Control and Management: A Literature Analysis," 2021: An overview of deep learning-based methods for managing and controlling urban traffic is provided in this survey. The uses of various deep learning models, such as LSTM networks, in traffic prediction, congestion detection, and control are discussed by the authors.

In [2] By Khan et al. (2020), "A Survey of Traffic Prediction and Congestion Management Methods in Intelligent Transportation Systems": This study examines various methods for predicting traffic patterns and reducing congestion, including some that rely on machine learning. The authors explore LSTM networks' usage in traffic prediction and point out their benefits over more conventional statistical techniques.

In [3] By Asif et al. (2021), "Microscopic Traffic Simulation: A Complete Assessment of SUMO" This report gives a general overview of SUMO and its uses in simulating traffic. The writers go over the many SUMO parts, such as network construction, vehicle routing, and traffic flow modelling, utilised to forecast crop productivity, such as the Means algorithm to forecast the atmospheric pollution factor.

In [4] By Deep learning techniques have been used by Hyun-Kyo Lim et al. (2019) to categorise network traffic. Several datasets based on packets were produced for this reason through the preprocessing of network traffic. Convolutional neural networks (CNN) and residual networks (ResNet) were employed to categorise network traffic. Five deep learning models were trained using these methods. Finally, the performance of packet-based datasets in the classification of network traffic was examined using the f1 score of the CNN and ResNet deep learning models. The results revealed how effective deep learning models were in categorising network traffic.

The network traffic categorization approach is used to separate malicious from non-malicious traffic. This method aids in anticipating the fraudulent user activity. Three crucial phases are used in the proposed methodology to categorise the network. The data are first clustered according to how similar and dissimilar they are using the k-mean clustering algorithm. Redundancy and missing values are two issues that are removed in order to enhance the dataset provided in the form of input. In the second stage, the k-means clustering approach is used to determine the network's centre. Here, the arithmetic mean of the entire dataset is computed. Calculating the Euclidian distance from the centre allows for the separation of comparable and unlike sites. One cluster contains all related data points. Different clusters of data points have different characteristics. The final stage is applying an SVM classifier so that the data points can be divided into two groups. KNN classifier, which clusters the unclustered points, is used to increase the performance and accuracy of the classification procedure. Additionally, it determines the Euclidian distance and distinguishes between data that is similar and distinct.

SUPPORT VECTOR MACHINE

An SVM classifier, a very well predictive model, is used to do text categorization. By classifying the data into two groups, an input can be provided and an output can be produced. The SVM training technique is used in the mode for text corpora. The training sample here can be found in any of the two classes. The data is categorised using an N-dimensional hyperplane that is built. Two parallel hyper planes are created on either side of the hyper plane to divide the data. The hyper plane is utilised in this case to divide the data so that the distance between two hyper planes can be increased. A data set that can be linearly separated and for which a linear classification function can be produced is utilised to correspond to the division of the hyperplane $f(X)$. Two classes are divided by this hyperplane, which runs across the middle of them. The classification of a new data instance, X_n , is done as follows by determining the function's sign and then testing it:

If $f(X_n) > 0$, then X_n belongs to a positive class.

It is possible to generalise the classifier error for wider margins or distances. An algorithm is known to perform better with high dimensional feature sets. This technique also used the kernel approach to turn the non-linearly separable data into new linearly separable data. SVM is used to calculate regression analysis and carry out several calculations. Also, this system ranks the elements. SVM performs better for numerous qualities even when only specific cases are used to train. The training and testing phases of this algorithm, however, encounter problems with speed and size. Another problem with this technique is how difficult it is to choose the kernel function settings. closest neighbour in K

The KNN classifier is the most basic classification method used in many applications. It is also known as the non-parametric supervised learning algorithm because there are no assumptions made about the distribution of the underlying data. The samples are classified based on the nearby training samples included in the feature space. The feature vectors used in the training process are saved along with the labels of the training images. Knearest neighbours are labelled, and during classification the unmarked query point is discarded. The majority of individuals characterized the thing using these labels. The item is classified as belonging to the class of objects that, at $k=1$, is closest to being even. If there are just two classes, K is considered as an odd number. A tie in the values can also occur during multiclass categorization if k is an odd full number. The most crucial goal of KNN algorithms is to classify the samples based on the majority class of its nearest neighbour.

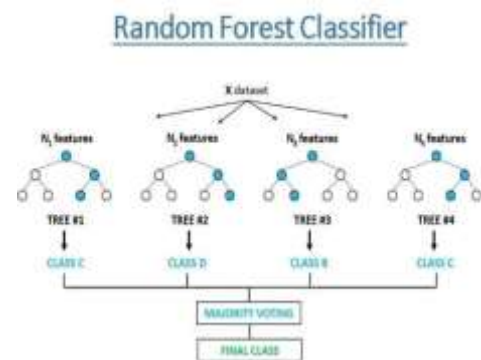
4.OVERVIEW OF THE EXISTING APPROACHES

Using SUMO, ensemble techniques like bagging and boosting have also been utilised to estimate traffic. In one study, researchers employed SUMO to forecast traffic flow on a highway network using a bagging-based ensemble classifier. To increase the predictability of the predictions, the ensemble classifier was trained on a variety of decision tree-based classifiers. The findings demonstrated that, in comparison to individual classifiers, the ensemble classifier was able to greatly increase the accuracy of forecasts of traffic flow. Overall, these methods show that utilising machine learning classifiers for traffic prediction

International Journal of Engineering Research & Technology (IJERT) using SUMO is beneficial. It is crucial to remember that the quality and representativeness of the training data, together with the complexity and resilience of the machine learning classifier employed, will all have an impact on how accurate the predictions are.

RFC (Random Forest Classifier)

A random forest is an ensemble learning technique that brings together various decision trees to produce a model that is more reliable and accurate. A random subset of the features and a subset of the training data are used to construct each decision tree in the random forest. This randomization aids in lowering overfitting and enhancing the model's generalizability. The random forest classifier algorithm first passes each input data point through each of the forest's decision trees before making a prediction. Each tree forecasts a class label, and the final forecast is determined by adding up the votes from each tree's forecasts. The usage of random forest classifiers is widespread in a variety of fields, such as fraud detection, natural language processing, and picture classification. A Random Forest Classifier is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

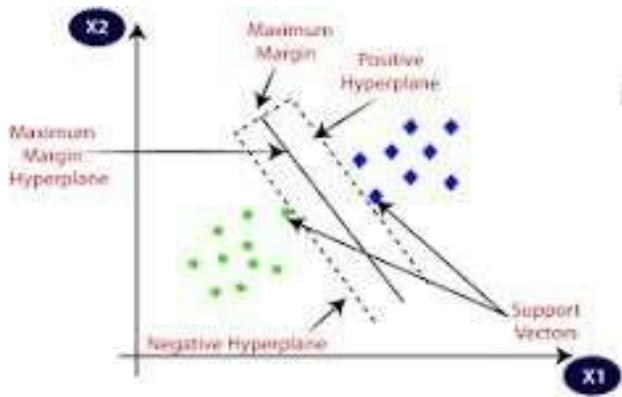


4.1.2 SVM (Support Vector Machine)

The objective of SVMs is to locate the hyperplane that divides the data points of various classes in a way that maximises the margin, or the distance between the hyperplane and the nearest data points of each class. The hyperplane is selected in a way that allows for good generalisation to fresh, untested data points. In order to move the data into a higher dimensional space where a hyperplane can more easily

separate the data points, SVMs require a kernel function. Radial basis function (RBF) kernels, polynomial kernels, and linear kernels are a few common kernel functions. Both linearly separable and non-linearly separable datasets can be handled by SVMs. SVMs can still be employed when the data cannot be separated linearly by adding a slack variable that permits some data points to be misclassified.

Support Vector Machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. The advantages of support vector machines are: Effective in high dimensional spaces. Still effective in cases where number of dimensions is greater than the number of samples.

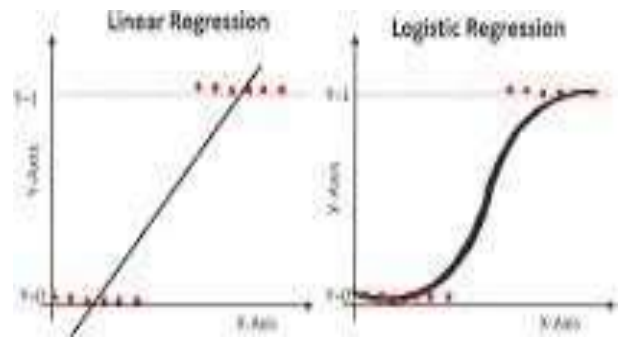


4.1.3. LR (Logistic Regression)

VR can be coupled in a variety of ways to produce cutting-edge applications. Using machine learning algorithms to provide more lifelike and engaging VR experiences is one example. Virtual characters that are responsive and intelligent may interact with users in more realistic and lifelike ways thanks to the application of machine learning (ML) technologies. This might make a VR experience more interesting and interactive. Another illustration is the use of machine learning to raise the effectiveness and performance of VR systems. For instance, ML algorithms can be applied to enhance the accuracy of tracking systems, decrease latency, and enhance picture and audio processing. This could contribute to giving people a more smooth and engaging VR experience.

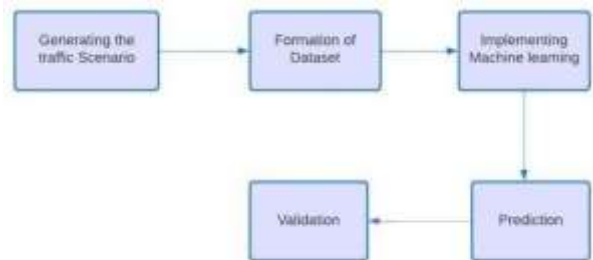
Logistic Regression is a special case of linear regression where the target variable is categorical in nature. It uses a log of odds as the dependent variable. Logistic Regression predicts the

probability of occurrence of a binary event utilizing a logistic function.



Search strategy

The searching is done by narrowing down to the basic concepts that are relevant for the scope of this review.



1. GENERATING THE TRAFFIC SCENARIO:

The initial stage in creating a traffic scenario using SUMO is defining the road network and traffic flow characteristics. The network editor in SUMO can be used to construct and change road networks, including the quantity of lanes, the posted speed limit, and the volume of traffic.

2.FORMATION OF DATASET:

The dataset should include a range of traffic flow parameters, such as traffic volume, speed, and congestion level, and should be typical of the traffic circumstances in the target area.

3.IMPLEMENTING MACHINE LEARNING:

To ensure accurate forecasts, it is essential to gather representative data and carefully preprocess it. Also, the accuracy of the model can be significantly impacted by choosing the proper model and hyperparameters.

4.PREDICTION:

Once trained and validated, a machine learning model can be used to anticipate new traffic scenarios. The model predicts the predicted output, such as the anticipated traffic

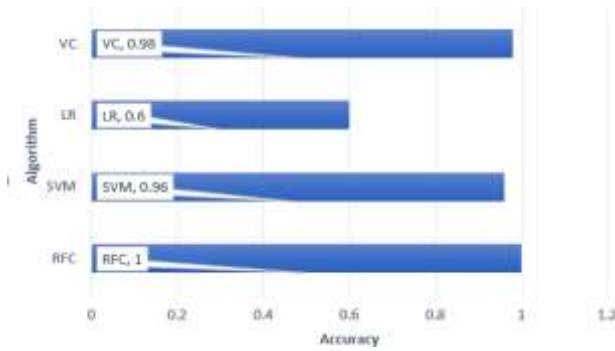
volume, speed, or congestion level, based on the input features, such as the time of day, weather conditions, and traffic flow parameters.

5. VALIDATION:

The trained model's performance is often assessed during validation using a different dataset called the validation set. The validation set is a portion of the larger dataset that is used to test the trained model's effectiveness rather than being used in the training phase.

5. RESULTS AND DISCUSSION

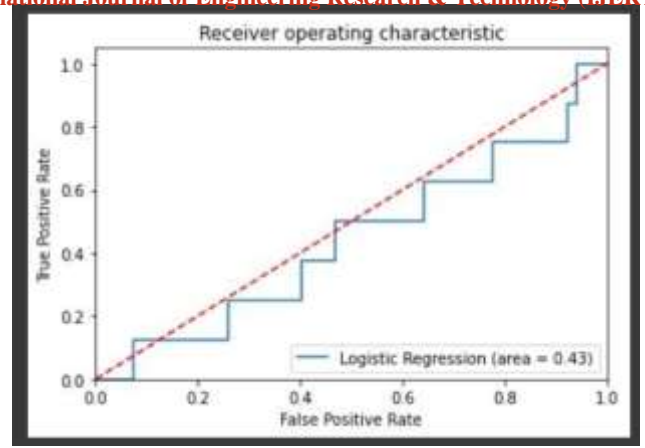
The ensemble voting classifier is a well-liked machine learning technique that combines various separate classifiers to enhance the performance of predictions in general. Using a combination of input features, such as traffic flow metrics, weather conditions, and time of day, an ensemble voting classifier could be trained on a dataset of traffic situations when predicting traffic using SUMO.



```
Accuracy Score of Random Forest Classifier : 1.0
```

```
Accuracy Score for SVC : 0.9666666666666667
```

```
Accuracy of logistic regression classifier on test set: 0.60
```



6. CONCLUSIONS

Traffic congestion prediction has drawn more attention. Every country is experiencing a problem with traffic congestion as a result of infrastructure development. As a result, projecting the congestion can assist governments to make plans and take the appropriate steps to avoid it. Big data and the development of machine learning have inspired researchers to use several models in this area. The methodologies were classified into three categories in this article. Probabilistic models are generally uncomplicated, but when many variables, such as the weather, social networks, and events, that influence traffic congestion are taken into account, they become more complicated. In this situation, machine learning, particularly deep learning, is advantageous. Since deep learning algorithms can evaluate a big dataset, their demand has grown over the years.

7. FUTURE SCOPE

This research shown that SUMO already offers a substantial framework and practical tools for the creation, validation, and assessment of massive traffic situations. Even though, there are still other features and extensions that are planned to improve and broaden SUMO. For instance, a fictitious Bologna scenario was created to demonstrate SUMO features that are not utilised in the real-world scenario, but it was not included in this presentation owing to page constraints. Intermodal railway simulation components including timetables, railroad crossings, railroad signals, and train dynamic models are already supported by SUMO. Many extensions are slated for the future in order to increase the possibilities of railway simulation.

- [1] Julian Nubert, Nicholas Giai Truong, Abel Lim, Herbert Ilhan Tanujaya, Leah Lim, Mai Anh Vu, "Traffic Density Estimation using a Convolutional Neural Network Machine Learning Project", National University of Singapore, 5 September 2018.
- [2] M Bando, K Hasebe, A Nakayama, A Shibata, "Dynamical model of traffic congestion and numerical simulation", *Phys. Rev. E* 51, 1035, 1 February 1995.
- [3] M Luo, SC Lin, IF Akyildiz, "Software defined network traffic congestion control", *Scientific Programming Volume 2017*, Article ID 3579540, December 2017.
- [4] Q Wang, J Wan, Y Yuan, "Locality constraint distance metric learning for traffic congestion Pitu, Larry Head, A real-time traffic signal control system: architecture, algorithms, and analysis.
- [5] Eigen, D., and Fergus, R. 2014. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. CoRR abs/1411.4734.
- [6] Gao, J.; Shen, Y.; Liu, J.; Ito, M.; and Shiratori, N. 2017. Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network. CoRR abs/1705.02755.
- [8] Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- [9] Singh, L.; Tripathi, S.; and Arora, H. 2009. Time optimization for traffic signal control using genetic algorithm.
- [10] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; and Wojna, Z. 2016. Rethinking the inception architecture for computer vision.
- [11] "RHODES to Intelligent Transportation Systems" Pitu Mirchandani, University of Arizona, 2005.
- [12] Yosinski, J.; Clune, J.; Bengio, Y.; and Lipson, H. 2014. How transferable are features in
- [13] B.-J. Chang, Y.-H. Liang, and J.-Y. Jin, "Adaptive cross-layer-based TCP congestion control for 4G wireless mobile cloud access," in Proceedings of the 3rd IEEE International Conference on Consumer Electronics-Taiwan, ICCE-TW 2016, pp. 1-2, May 2016.
- [14] M. Claeys, N. Bouten, D. De Vleeschauwer et al., "Deadline-aware TCP congestion control for video streaming services," in Proceedings of the 12th International Conference on Network and Service Management, pp. 100–108, November 2016.
- [15] J. Gruen, M. Karl, and T. Herfet, "Network supported congestion avoidance in software-defined networks," in Proceedings of the 2013 19th IEEE International Conference on Networks, ICON 2013, p. 16, December 2013.
- [16] J.-H. Wang, K. Ren, W.-G. Sun, L. Zhao, H.-S. Zhong, and K. Xu, "Effects of iodinated contrast agents on renal oxygenation level determined by blood oxygenation level dependent magnetic resonance imaging in rabbit models of type 1 and type 2 diabetic nephropathy," *BMC Nephrology*, vol. 15, no. 1, article 140, 2014.
- [17] P. Sun, M. Yu, M. J. Freedman, J. Rexford, and D. Walker, "HONE: Joint Host-Network Traffic Management in Software-Defined Networks," *Journal of Network and Systems Management*, vol. 23, no. 2, pp. 374–399, 2015.
- [18] H. Long, Y. Shen, M. Guo, and F. Tang, "LABERIO: dynamic load balanced routing in OpenFlow-enabled networks," in Proceedings of the IEEE International Conference on Advanced Information Networking Applications, pp. 290–297, 2013.
- [19] L. Lu, Y. Xiao, and H. Du, "OpenFlow control for cooperating AQM scheme," in Proceedings of the 2010 IEEE 10th International Conference on Signal Processing, ICSP2010, pp. 2560–2563, October 2010.
- [20] D. SreeArthi, S. Malini, M. J. Jude, and V. C. Diniesh, "Micro level analysis of TCP congestion control algorithm in multi-hop

