WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Traffic Prediction for Intelligent Transport System Using Machine Learning

Sasikala M, Kiruthika S, Deepa Sree A, Karthika K, Jyothi G.

Dhirajlal Gandhi College of Technology, Salem 636 309.

ABSTRACT

Machine learning is used in traffic prediction for intelligent transportation because it is a powerful tool that can process and analyse large amount of data to generate accurate predictions. While Random Forest Algorithm can be effective in some traffic prediction tasks but it is not that better efficient and not computationally expensive, especially as the size of the dataset grows larger. The main aim of our project is to develop a tool for predicting correct and timely traffic flow info using Support Vector Machine (SVM). SVM works relatively well when there is clear margin of separation between classes, it has shown high accuracy in predicting traffic patterns compared to other machine learning algorithms, such as neural networks, decision trees, and Random Forest Algorithm. To address this gap, proposing a deep learning approach and build a deep neural network model based on Support Vector Machine (SVM). SVM can handle high-dimensional data, which is common in traffic prediction datasets that contain many variables such as traffic volume, time of day, and it is relatively memory efficient. SVM algorithm has several advantages that make it a better choice in many situations, especially when dealing with high-dimensional, time efficiency, and accuracy. The goal of this project is to develop associate degree applicable machine learning tool which may predict the traffic for intelligent transit exploitation GPS, speed, direction, and start-end junction options. This could be more applicable for risk analysis and unsafe area identification rather than decreasing the emergency response time. The goal of this research is to establish a deep learning SVM neural network model that can address some of these problems. The deep learning approach provides automatic representation learning from raw data. Instead of common classification and regression approaches, we propose an anomaly detection approach. It is based on the difference between prediction and actual values. It eliminates the cumbersome process of labelling the data as 'Traffic' and 'non-Traffic' and dealing with imbalanced data. Furthermore, the detected anomalous data represents hazardous situations in addition to potential Traffic. SVC extends SVM to the case of multiple classes. A SVM is powerful because it uses contextual information when mapping between input and output sequences. Several detailed features such as speed, flow, and weather are used to train the predictive model and identify the abnormal dynamics that lead to a Traffic occurrence based on the weather changes

Keywords: Machine learning, Random Forest Algorithm, Support Vector Machine (SVM), High-dimensional data, Time efficiency, Accuracy, Data mining, Global positioning system (GPS).

1. INTRODUCTION

SVM works relatively well when there is clear margin of separation between classes, it has shown high accuracy in predicting traffic patterns compared to other machine learning algorithms, such as neural networks, decision trees, and Random Forest Algorithm. Support Vector Classification (SVC) is a type of machine learning algorithm that builds on the principles of Support Vector Machines (SVM) to classify data into different classes. SVM is a binary classification algorithm that aims to find the best hyperplane to separate data points into different classes. This hyperplane is chosen to maximize the margin between the closest points of each class, which helps to improve the generalization performance of the model. SVC extends SVM to the case of multiple classes. It achieves this by identifying the best separating boundary or hyperplane between the classes in the dataset. The boundary is chosen based on the maximum margin principle that SVM follows. In other words, SVC aims to find the hyperplane that maximizes the distance between the closest points of each class, ensuring that the model can generalize well to new data. The feedback connections provide a SVM the memory of past activations, which allows it to learn the temporal dynamics of sequential data. A SVM is powerful because it uses contextual information when mapping between input and output sequences. This can help the users to compute the safety measures which is useful to avoid Traffic. One of the key features of SVM and SVC is their ability to use kernel functions to map data to higher dimensions. This allows them to find more complex decision boundaries that can separate data points that are not linearly separable in the original feature space. The choice of kernel function is critical and can significantly impact the performance of the model. To address this gap, proposing a deep learning approach and build a deep neural network model based on Support Vector Machine (SVM). SVM can handle high-dimensional data, which is common in traffic prediction datasets that contain many variables such as traffic volume, time of day, and it is relatively memory efficient. SVM algorithm has several advantages that make it a better choice in many situations, especially when dealing with high-dimensional, time efficiency, and accuracy. Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systemsThe overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use Aside from the raw analysis step, it involves database and

data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

2.SYSTEM DESIGN

2.1 EXISTING SYSTEM

In Existing System, Random Forest is a popular ensemble learning algorithm that combines multiple decision trees to improve the accuracy of predictions. The idea behind this algorithm is to create multiple decision trees using random subsets of the training data and randomly selected subsets of the features. The final prediction is made by aggregating the predictions of all the individual trees. The randomness in creating decision trees helps to reduce overfitting and increase the robustness of the model. Random Forest is effective in handling noisy and complex datasets, and can be used for both classification and regression problems. Since it creates multiple decision trees, it can capture more complex relationships between the input features and the output labels. Moreover, the algorithm can also handle missing values in the data. One of the main advantages of Random Forest is that it provides a way to measure the importance of each feature in the dataset. This can help in feature selection and data exploration. Additionally, the algorithm is relatively easy to implement and can be used with a wide range of datasets. Random Forest is effective in handling noisy and complex datasets, and can be used for both classification and regression problems. Since it creates multiple decision trees, it can capture more complex relationships between the input features and the output labels. Moreover, the algorithm can also handle missing values in the data. One of the main advantages of Random Forest is that it provides a way to measure the importance of each feature in the dataset. This can help in feature selection and data exploration. Additionally, the algorithm is relatively easy to implement and can be used with a wide range of datasets. However, there are also some drawbacks of Random Forest. The algorithm can be computationally expensive, especially when the number of trees and features is large. Moreover, the final model can be difficult to interpret, as it involves multiple decision trees. Despite these limitations, Random Forest remains a popular and powerful algorithm for machine learning tasks. In traffic prediction, imbalanced data could be problematic if certain traffic conditions or events are rare but crucial to capture accurately, such as traffic accidents or severe weather conditions. Random forests are known to be less interpretable compared to some other machine learning models. Interpreting the relationships between features and traffic patterns might be more difficult with random forests.

2.2 PROPOSED SYSTEM

In Proposed system using Support Vector Machine (SVMs) for Traffic detection occurrence of weather changes like rain, foggy. It is a type of artificial neural network that is well-suited for processing sequential data such as text, audio, or video. SVMs have a recurrent connection between the hidden neurons in adjacent layers, which allows them to retain information about the previous input while processing the current input. Support Vector Machine (SVMs) are neural networks with feedback connections specifically designed to model sequences. They are computationally more powerful and biologically more reasonable than feed-forward networks (no internal states). The feedback connections provide a SVM the memory of past activations, which allows it to learn the temporal dynamics of sequential data. A SVM is powerful because it uses contextual information when mapping between input and output sequences. However, the traditional SVMs have a problem called vanishing or exploding gradient. Models are created using Traffic data records which can help to understand the characteristics of many features like driver's behavior, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measures which is useful to avoid Traffic. It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. the model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduce it. Here the road Traffic study is done by analyzing some data by giving some queries which is relevant to the study. What is the trend in the number of Traffic that occur each year, do Traffic in high-speed limit areas have more casualties and so on. This analysis aims to highlight the data of the most importance in a road Traffic and allow predictions to be made. Several detailed features such as weather, speed, and flow of upstream and downstream points are extracted from big datasets. SVM in traffic prediction is its ability to handle high-dimensional data with a small number of training sample. SVM has good generalization performance, which means it can make accurate predictions on unseen data. This is important in traffic prediction, where accurate predictions are required for future time periods. Multiple machine learning models, such as SVM model and Decision Tree model, were used as the candidate models. It was found that weather, crash time, and flow shortly prior to the crash occurrence are critical impacting factors for real-time crash prediction. The candidate models have low to moderate sensitivity to predict the crash occurrences due to limited sample size.

2.3 ALGORITHM

Input: Input data matrix, class information

Output: Set of Basis vectors

begin

Repeat

For every candidate example - examples not in current set of BVs

include it in the model efficiently Observe the generalization performance on the remaining points

end for candidate examples

Add that point to the BVs' list that gave better test error

Till the stopping criterion

End

3. SYSTEM ARCHITECTURE

The main goal of this system is to predict heart disease using data mining technique such as SVM Algorithm. Raw data set is used and then preprocessed and transformed the data set. Then apply the data mining technique such as SVM algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and then accuracy is calculated.

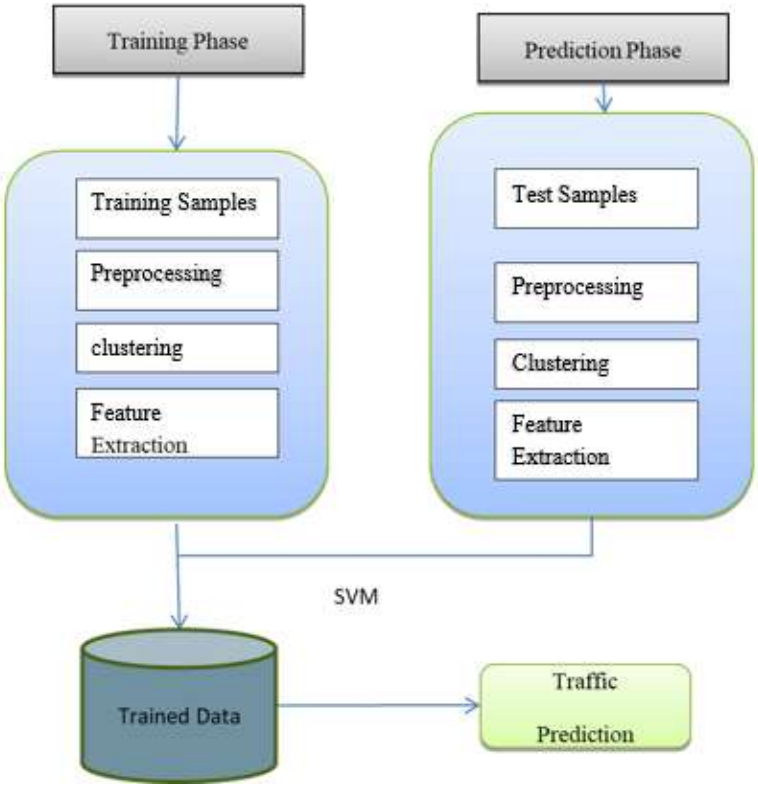

Fig 3. Architecture diagram of System

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Analysing data that has not been carefully screened for such problems can produce misleading results. The pre-processing module for traffic prediction is to clean, integrate, and transform the data so that it can be used effectively in a machine learning model to predict traffic patterns. Pre-processing plays a crucial role in traffic prediction as it involves transforming raw traffic data into a suitable format for analysis and modelling. The goal of pre-processing is to enhance the quality of the data, reduce noise, handle missing values, and extract relevant features that can be used to build accurate traffic prediction models. Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, make the software that can learn how to classify the data items into groups. For example, can apply classification in application that "given all past records of employees who left the company, predict which current employees are probably to leave in the future." In this case, divide the employee's records into two groups that are "leave" and "stay". Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. In a library, books have a wide range of topics available. Different from classification, clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. To make the concept clearer, can take library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without hassle. The task of accurate forecasting in extreme conditions is difficult mainly due to the complex nature of Traffic. Another problem with Traffic prediction/detection is the scarcity of Traffic in both space and time.

4.MODULES DESCRIPTION

4.1 Dataset Acquisition

In this module, upload the datasets. Gather the data from hospitals, data centres and cancer research centres. The collected data is pre-processed and stored in the knowledge base to build the model. The „Diagnosis‟ attribute is used to predict the heart disease with value “2” for patient having heart disease and “1” for patient having no heart disease. The „patient ID‟ attribute is used as a key and others are input attributes. A data acquisition module for traffic prediction typically refers to a system and gathers relevant data from various sources to feed into a traffic prediction model. This module plays a crucial role in providing the necessary input data for accurate and reliable traffic predictions. A well-designed data acquisition module enables the collection, processing, and integration of diverse data sources to provide accurate and timely information for traffic prediction models and systems.

4.2 Pre-processing

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Analysing data that has not been carefully screened for such problems can produce misleading results. The pre-processing module for traffic prediction is to clean, integrate, and transform the data so that it can be used effectively in a machine learning model to predict traffic patterns. Pre-processing plays a crucial role in traffic prediction as it involves transforming raw traffic data into a suitable format for analysis and modelling. The goal of pre-processing is to enhance the quality of the data, reduce noise, handle missing values, and extract relevant features that can be used to build accurate traffic prediction models.

4.3 Clustering

Clustering is a technique in data mining to find interesting patterns in a given dataset. The k-means algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters information’s into k groups, where k is considered as an input parameter. It then assigns each information’s to clusters based upon the observation’s proximity to the mean of the cluster. The cluster’s mean is then more computed and the process begins again. The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical data and related fields. K-Means algorithm is a divisive, unordered method of defining clusters. Clustering in traffic prediction by organizing traffic data into meaningful groups, extracting useful features, and enabling more accurate predictions and anomaly detection. Clustering helps identify underlying structures and similarities in traffic data, allowing for more accurate predictions and better understanding of traffic behaviour.

4.4 Feature Selection

In this module is used to select the features of the given dataset. Attribute selection was performed to determine the subset of features that were highly correlated with the class while having low inter correlation. By selecting the most relevant features, you can improve the accuracy and efficiency of your model. It can be used individually or in combination to select the best set of features for traffic prediction using machine learning.

4.5 Classification

A recurrent neural network (SVM) is a class of artificial neural networks where connections between nodes can create a cycle, allowing output from some nodes to affect subsequent input to the same nodes. This allows it to exhibit temporal dynamic behaviour. Derived from feed forward neural networks, SVMs can use their internal state (memory) to process variable length sequences of inputs This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition Support Vector Machine are theoretically Turing complete and can run arbitrary programs to process arbitrary sequences of inputs. classification module will depend on the quality and quantity of your data, the chosen features, and the selected machine learning algorithm. Classification in traffic prediction involves predicting the traffic state or condition of a particular location or road segment at a given time, typically categorized as one of several predefined classes, such as "free-flowing," "moderate," or "congested."

5.CONCLUSION

In summary, machine learning-based traffic prediction offers promising solutions for addressing traffic congestion and improving transportation efficiency. Continued research and development in this field can lead to further advancements and the realization of smart, data-driven transportation systems in the future. Road Traffic are caused by various factors. By going through all the research papers, it can be concluded that Road Traffic cases are hugely affected by the factors such as types of vehicles, age of the driver, age of the vehicle, weather condition, and road structure and so on. Thus, we have built an application which gives efficient prediction of road Traffic based on the above-mentioned factors. By going through all the research paper, it can be concluded that road Traffic cases are hugely affected by the factors such as types of weather condition and road structure. Thus, we have built an application which gives efficient prediction of road Traffic based on machine learning.

ACKNOWLEDGEMENT

We would like to express our deep and sincere gratitude to professor **Ms. M. SASIKALA, M.E.,** for giving the opportunity and guidance throughout this research.

REFERENCES

[1] Pededn M. (2004) “World report on road traffic injury prevention”, Geneva: World Health Organization.

[2] M. Chang, L. Y., & Chen, W.C. (2005). “Data mining of Tree-based models to analyze freeway Traffic frequency” Journal of Safety Research, 36(4), 365-375.

[3] Tesema T. B., Abraham A., & Grosan, C. (2005). “Rule mining and classification of road traffic Traffic using adaptive regression trees”. International Journal of Simulation,6(10),80- 94.

[4] Ayushi Jain, Garima Ahuja, Anuranjana, Deepti Mehrotra, Data mining approach to investigate the road Traffic in India, proc of IEEE,2016.

[5] Global status report on road safety: supporting a decade of action, Geneva, World health organization, 2013.

[6] Handbook of Research on Computational Science and Engineering: Theory and practice.

[7] Peden, M. (2004) "World report on road traffic injury prevention". Geneva: World Health Organization. [8] M. Chang, L. Y., & Chen, W. C. (2005). "Data mining of tree- based models to analyse freeway Traffic frequency". Journal of Safety Research, 36(4), 365-375.

[8] Tesema, T. B., Abraham, A., & Grosan, C. (2005). "Rulemining and classification of road traffic Traffic using adaptive regression trees".

[9] World Health Organization [11], “Gender Specific Classification of Road Traffic Patterns through Data Mining Techniques”, IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012) March 30, 31, 2012.

[10] Ma, W., Zhang, Y., Zhou, J., & Yuan, Y. (2015). A survey on traffic prediction using machine learning techniques. IEEE Transactions on Intelligent Transportation Systems, 17(3), 697-709

[11] Liu, Q., Wang, D., Li, J., & Zhang, X. (2017). Short-term traffic flow prediction with spatial-temporal correlation in a large-scale network. Transportation Research Part C: Emerging Technologies, 79, 96-114.

[12] Lv, Y., Duan, Y., Kang, W., Li, G., & Gong, Y. (2015). Traffic flow prediction with big data: A deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), 865-873.

[13] Zhang, X., Gao, Z., Zhang, L., Zhang, J., & Xie, J. (2016). Short-term traffic flow prediction with a deep learning method. Transportation Research Part C: Emerging Technologies, 69, 182-197.

[14] Zheng, S., Chen, D., Chen, J., Liu, X., & Liu, Y. (2017). A long short-term memory neural network for traffic flow prediction using remote microwave sensor data. Sensors, 17(6), 1225.

[15] Lv, Y., Zhang, M., Gong, Y., Han, Z., & Li, G. (2018). Traffic flow prediction with spatial-temporal correlation in big data. IEEE Transactions on Intelligent Transportation Systems, 19(11), 3702-3711.

[16] Zhang, J., Zheng, Y., Qi, D., & Wang, R. (2018). Deep spatio-temporal residual networks for citywide crowd flows prediction. In Thirty-Second AAAI Conference on Artificial Intelligence.

[17] Li, Y., Peng, Z. R., & Ding, W. (2018). A survey on traffic prediction using machine learning. IEEE Transactions on Intelligent Transportation Systems, 19(2), 720-738.

[18] Chen, F., Jiang, Y., & Liu, L. (2019). A survey on urban traffic prediction based on big data. IEEE Transactions on Intelligent Transportation Systems, 20(3), 1022-1039.

[19] Ma, L., Sun, G., & Sun, X. (2015). A survey of big data analytics for traffic prediction with transportation cyber-physical systems. IEEE Transactions on Intelligent Transportation Systems, 16(2), 766-778.