# **See the source image**

TRAFFIC PREDICTION FOR INTELLIGENT TRANSPORT

SYSTEMS (ITS) USING MACHINE LEARNING

**Name: AGBOOLA TEMIREMI ADEGBAMIGBE**

**Student Number: 2063529**

**7CS041/ MSC PROJECT DATA SCIENCE**

School of Engineering, Computing & Mathematical Sciences

University of Wolverhampton

Wolverhampton, England; United Kingdom.

Submitted in Partial Fulfilment of the Requirement of the MSc in Data Science

**Supervisor: Dr. Nahar Vinita**

**Project Coordinator: Dr. Andrew Gascoyne**

**23December 2022**

# **Abstract**

In recent years, Intelligent Transportation Systems (ITS) have gotten more and more attention. Vehicular Cloud (VC), intelligent traffic controls, etc. are only a few of the outstanding applications that have been presented under the umbrella of ITS due to the quick development of vehicular computer hardware, vehicular sensors, and urban infrastructures. These apps can make our traveling environment safer, more effective, and more fun. However, in order to implement these applications, a precise and effective traffic flow forecast system is required, which provides a chance for ITS applications to anticipate potential road conditions. Many prediction techniques, including mathematical modeling techniques, parametric techniques, and non-parametric techniques, have been proposed in order to improve traffic flow prediction performance. How to develop a reliable, effective, and efficient vehicle traffic prediction system is a perennially popular issue. The aim of this study is to develop a machine leaning model that can predict traffic flow. This study reviewed and compared the performance of the traditional base models against the efficiency of the recurrent neural network (RNN) in predicting traffic flow. The findings revealed that the kernel-based model – support vector regression performed better than the example-based model- K-nearest neighbor regression. Overall, the efficiency of the recurrent neural network I predicting traffic flow was higher than the base model. This study adopted the RNN as the most efficient model for the intelligence traffic system (ITS).

**Keywords**: Intelligence transport system, machine learning model, kernel-based model, example-based model, recurrent neural network.

**Table of Contents**

[**Abstract** 1](#_Toc122066825)

[**CHAPTER ONE** 3](#_Toc122066826)

[**INTRODUCTION** 3](#_Toc122066827)

[**1.1** **Background** 3](#_Toc122066828)

[**1.2** **Rationale** 4](#_Toc122066829)

[**1.3** **Aims** 5](#_Toc122066830)

[**1.4** **Objectives** 6](#_Toc122066831)

[**1.5** **Research Approach** 7](#_Toc122066832)

[**1.6** **Structure of the Report** 7](#_Toc122066833)

[**CHAPTER TWO** 8](#_Toc122066834)

[**Literature Review** 8](#_Toc122066835)

[**2.1** **Introduction** 8](#_Toc122066836)

[**2.2 Intelligence Transportation System** 9](#_Toc122066837)

[**2.3** **Technologies in Intelligence Transport System** 10](#_Toc122066838)

[**2.3.1** **Sensing Technology** 10](#_Toc122066839)

[**2.3.2** **Video analysis** 11](#_Toc122066840)

[**2.3.3** **Computational Technologies** 11](#_Toc122066841)

[**2.4** **Machine Learning-Based Method** 12](#_Toc122066842)

[**2.4.1** **Regression Model** 15](#_Toc122066843)

[**2.4.2** **Example-based model (KNN)** 18](#_Toc122066844)

[**2.4.3** **Kernel-based model (SVM)** 21](#_Toc122066845)

[**2.5** **Neural Network-based Model** 25](#_Toc122066846)

[**2.5.1** **Feed Forward Neural Network (FFNN)** 27](#_Toc122066847)

[**CHAPTER THREE** 30](#_Toc122066848)

[**Methodology** 30](#_Toc122066849)

[**3.1** **Introduction** 30](#_Toc122066850)

[**3.2** **Data collection** 30](#_Toc122066851)

[**3.3** **Processing of Data** 31](#_Toc122066852)

[**3.3.1** **Data normalization** 31](#_Toc122066853)

[**3.3.2** **Moving Average** 31](#_Toc122066854)

[**3.4** **Training algorithm** 32](#_Toc122066855)

[**3.5** **Model design** 33](#_Toc122066856)

[**3.5.1** **Kernel based model** 33](#_Toc122066857)

[**3.5.2** **Support vector machine** 35](#_Toc122066858)

[**3.5.3 Artificial Neural Network** 42](#_Toc122066859)

[**3.6** **Model Performance Evaluation Metrics** 45](#_Toc122066860)

[**3.6.1** **Root Mean Square Error** 45](#_Toc122066861)

[**3.6.2** **Accuracy** 46](#_Toc122066862)

[**3.6.3** **Precision** 46](#_Toc122066863)

[**3.6.4** **Recall** 46](#_Toc122066864)

[**3.6.5** **Area under curve (AOC)** 46](#_Toc122066865)

[**3.6.6** **F- Measure** 46](#_Toc122066866)

[**CHAPTER FOUR** 47](#_Toc122066867)

[**DATA ANALYSIS AND RESULT PRESENTATION** 47](#_Toc122066868)

[**4.1** **Introduction** 47](#_Toc122066869)

[**4.2** **Dataset Overview** 47](#_Toc122066870)

[**4.3** **Exploratory Analysis** 48](#_Toc122066871)

[**4.3.1** **Data Processing** 51](#_Toc122066872)

[**4.4** **Modeling** 52](#_Toc122066873)

[**4.4.1** **Model Performance Evaluation Metrics** 55](#_Toc122066874)

[**4.5** **Traffic Flow Forecasting** 57](#_Toc122066875)

[**CHAPTER FIVE** 61](#_Toc122066876)

[**SUMMARY AND CONCLUSION** 61](#_Toc122066877)

[**5.1** **Summary** 61](#_Toc122066878)

[**5.2** **Conclusion** 62](#_Toc122066879)

[**5.3** **Limitation of study** 62](#_Toc122066880)

[**References** 64](#_Toc122066881)

# 

# **CHAPTER ONE**

# **INTRODUCTION**

## **Background**

The term intelligent transportation system (ITS) can be referred to as tools used for information analysis. To increase safety and efficiency, ITS is utilized to regulate communication technology for the road transportation sector. A variety of applications that are part of an intelligent transportation system are used to gather information, reduce traffic, enhance traffic management, protect the environment, and boost the advantages of transportation. According to (Meena & Mahrishi, 2020) Intelligent Transportation Systems (ITSs) were accepted in the world congress conducted in Paris, in 1994. ITS has different applications in communication technology, electronics, and computer to provide information to the traveler to enhance the efficiency and safety of the road transportation system. The benefit of ITS is that it provides safe and smooth road transport movement. It can also be used to decrease carbon emissions from the standpoint of environmental friendliness. It offers several opportunities to automobile or automotive industries to improve the security and safety of travelers.

(Boukerche & Wang, 2020) reported that with the development of urbanization and the acceptance of automobiles, the problems related to transportation are becoming more challenging, accidents are common, traffic flow is crowded, and the deteriorated traffic environment. The current world demands technology to solve such issues. Traffic prediction refers to forecasting the density and volume of the flow of traffic, mainly for the purpose of managing the movement of the vehicle, reducing congestion, and creating optimal (energy-consuming and least-time) routes. Most of the issue is solved through ITSs which aimed to improve the accuracy of the traffic flow estimation. It is considered the most critical element for advanced public transit, traffic control systems, as well as succeeding advanced public transportation systems.

(Chen & Chen, 2019) concluded that as per the historical and real-time traffic data obtained from numerous sensor outlets like inductive loops, sensors, radars, crowdsourcing, mobile Global Positioning System, as well as social media are used to identify the traffic flow. Since the prevalent use of emerging technology and traditional sensors, traffic data is set off and there emerges the need for large-scale data related to transportation.

The main role of ITSs is to analyze the information on traffic. This is mainly used to regulate communication technologies for road transit to enhance efficiency and safety. ITSs include a broad range of different applications which is used to access information, control congestion, enhance traffic management, decrease environmental effects and enhance the benefits of transit. ITS means the different kinds of needs in the field of transport along with similar policing. By using sensors in smartphones, it becomes easy to track/detect the density and speed of traffic.

In this context, machine learning is the most popular system used for predicting the flow of traffic. Machine learning (ML) permits to development of predictive models that study large commonalities of heterogeneous data from distinct sources. Machine learning algorithms can successfully extract information about the flow of traffic without the interference of humans.

## **Rationale**

(Guerreiro, et al., 2016) write that in the busy world, one of the major issues is traffic congestion and this is due to overpopulation. Due to traffic issues, there are several other issues associated with this like economic problems and health problems in society, negative impact on the environment. Another reason is urbanization and the popularity of automobiles; the issue of transportation becomes serious. Hence, this situation needs to be changed. Decreasing the traffic flow becomes an urgent need of society. Accurate prediction of traffic is the most challenging task because this data is complex and dynamic. The zone traffic depends on the road capacity, kind of road users, weather, time, traffic policies, time of the day, events, and others.

(Zantalis, et al., 2019.) find two ways to solve this problem; here first way to solve the problem of transportation is by constructing more highways and expanding more lanes on the road. But expanding the capacity of the road will result in more serious traffic situations and high infrastructure costs. Whilst, another way to improve the environment of traffic is to set an accurate and efficient transportation system that assists people in arranging resources of transportation, scatters the traffic flow formerly it is overloaded, and provides more copious on-road entertainment. Executing ITS using Machine Learning can help in solving this issue as this is a system that integrates a variety of technologies like transportation communication systems. Hence, ITS with machine learning will benefit in improving traffic efficiency, increasing the capacity of roads, comfort traffic congestion, and reducing environmental pollution and traffic accidents (Chen, et al., 2020) Also, such traffic control strategies decrease the expenses of the construction, and it proves to be cost-effective models for the traffic managers and the government.

## **Aims**

The purpose of this project proposal is to develop the ITS using machine learning that aims to be completed from previous literature and executing strategies to reduce the traffic flow and provide real-time information to the user. By examining different models of machine learning and ITS, the project can determine different aspects that can be executed and supported by the design of ITS using machine learning. There has been an increase in the use of ITS as this application brings an efficient, enjoyable, and safe transportation environment for the people. Future research requests to implement ITS for the general people so that it becomes easy for them to track the traffic flow before selecting a particular road. The aim of this project is to create the ITS with machine learning for the general public so that they can use it for making the right transportation decision before leaving the current place.

Research questions

* What are the key problems or issues faced by the people without having real-time information on traffic flow?
* Why it is important for people to predict the flow of traffic before leaving their current place?
* What are the different benefits that people can get from using ITS for traffic predicting?

## **Objectives**

The main objective of this paper is to prepare and design the ITS for individuals that will assist them in predicting the traffic flow before leaving the house so that they can choose an alternative path to reach the destination. Presently, the ITS is mainly used by traffic managers and the government to manage traffic or for preparing the traffic rules, but when it can be accessed by individuals, it will better help them to predict the traffic flow through real-time information so that they can decide their path accordingly. Following objectives have been set to attain the aim of the project:

* Evaluate the performance of several famous state-of-the-art non-parametric models and their Deep Learning structures on the dataset. The study focusses on the model’s accuracy and training cost and the efficient optimization algorithm, reducing the implementation cost of a prediction model with a complex inner structure.
* Propose a new traffic flow prediction system using Machine Learning-based model that has better prediction accuracy than the other state-of-art models.

## **Research Approach**

The adopted research for the development of ITS for the general public will be united with the action research method and pursues to uncover new ideas and thoughts on how ITS for individuals can be developed. Action research is a self-reflective and iterative research method that laid out the basics of the research framework. This method follows a cylindrical process divided into different stages i.e. evaluation, preparation, execution, and then it again restarts based on the outcomes of the previous cycle (Singh, 2015). The reason for selecting this research approach is that the project proposes to develop a new idea for solving the problems of the public, hence this approach helps in following the stages to attain the objective.

## **1.6 Structure of the Report**

This study will attempt to provide a digital solution to the research question of how information and communication technology especially in the aspect of machine learning modeling be used to predict traffic flow in transport intelligence system. A literature review to establish the current state of things in the use of machine leaning based model to develop traffic flow intelligence system (Chapter 2). Furthermore, a deep dive into the methodology implemented for this research work will be discussed (Chapter 3). Subsequently, the training, and implementation of the machine learning model for predictive analysis will be recorded in Chapter 4. Lastly, reflection and conclusion on the performance of the implemented machine learning models, and stating the further steps of the research work will be captured in Chapter 6

# 

# **CHAPTER TWO**

# **Literature Review**

## **2.1 Introduction**

This chapter provides a review of the literature on Traffic Prediction for Intelligent Transportation systems using Machine Learning. The chapter discusses the role of machine learning in developing an intelligent transportation system network. A review of past work on the historical conception of an intelligent transportation system, different models, and machine learning algorithms are presented.

## **2.2 Intelligence Transportation System**

The steady increase in the number of vehicles, especially for any urban city has called for an increase in the efficiency of the traffic management system. The necessity of developing a complementary effort to match the ever-increasing number of vehicle usage on a daily basis has been a major concern for most in developing countries. Based on the literature, one of the basic solutions is for the government to increase the capacity of infrastructure but this can lead to wastage of resources, time, and human effort (Falch, 2020). The evolution and development brought by information communication technology (ICT) in the last few decades, especially in the science and technology industry have made countries around the world look towards ICT in providing effective solutions by developing efficient traffic management systems. To break the barrier of cost and wastage of resources, ICT provides sophisticated systems such as Wireless Sensor Networks (WSN) and low-power-consuming sensors that are capable of creating a cost-effective modern intelligent transportation system. Cities can be mapped using Internet of Things (IoT) sensors in order to gather crucial data. Now that IPv6 is widely used, it is possible to provide a sensor node and an IP address for tracking and collecting real-time data on traffic flow (Guravaiah, et al., 2020). However, the nature of the data obtained is heterogenous and somehow complex which will always require high-level algorithms or tools to be able to extract useful information and provides efficient data analytics insight from the big chunk of data collected. According to (Khanna, et al., 2018), the degree of traffic congestion in a certain location of a city can be predicted using deep learning and machine learning techniques.

The concept of the Intelligent Traffic System has been around for decades, in 1994, the World Congress Command in Paris endorsed the concept of intelligent transportation (ITS). To increase the safety and effectiveness of road transportation systems, the ITS has applied computer, physics, and communication technologies to supply human data (Narendran, et al., 2022). An intelligent Traffic System can be described as a system that integrates a number of tools, including communication technologies, software, hardware, and traffic engineering principle to control a city’s traffic system that improves productivity and safety. Developed countries around the world are already implementing ITS through the use of different technologies (Dubey, et al., 2017).

## **2.3 Technologies in Intelligence Transport System**

In order to enhance transportation conditions, safety, and services, various technologies have been adopted for the Intelligent Transportation System. This section highlights and discusses briefly some of the technologies.

### **2.3.1 Sensing Technology**

The use of sensory technology in intelligent transportation systems is growing every day. Wireless sensor nodes are inexpensive and require little power to process data. For such systems, end-to-end latency and synchronization are essential. These sensors are designed to collect data from one or more sink nodes that are linked to one another via various long-distance connections, such as satellite, WiFi, WiMAX, etc. (Qureshi & Abdullah, 2013) established the different types of sensor networks such as thermal, seismic, infrared and magnetic which can adequately monitor specific vehicle conditions like temperature, direction, speed, etc. Also, the sensor nodes which was described as a tiny object similar to a matchbox; consists of four main components, Power unit, transceiver, sensing, and processing unit which are responsible for tracing vehicle locations (Qureshi & Abdullah, 2013).

### **2.3.2 Video analysis**

In using video analysis for ITS, there is a need for a smart camera together with a processing unit and a communication unit. Mounting, the video cameras will constantly monitor vehicle movements, and the captured video is compressed just to reduce the transmission bandwidth. The statistics gathered from the video cameras can be used to determine the frequency, average speed, and even the vehicle lane (Falch, 2020). Discussing some of the difficulties in the use of video analysis, (Leung, 2016) highlighted that the system is quite expensive to set up, and heavy downpours of rain or fog can affect the system.

### **2.3.3 Computational Technologies**

Advancement in technology has made computer science and its tools very resourceful in so many fields. Part of these rich tools includes mode-based or machine learning process control, deep learning, and artificial intelligence. For real-time applications that need real-time operating systems, powerful microprocessors, rich memory, and hardware, computational technologies in intelligent transportation systems offer various architectures and software. To address various issues in the transportation industry, numerous algorithms and computational programs have been created (Falch, 2020). Different machine learning models can be used for traffic prediction and management and various researchers have deployed these model-based algorithms for traffic prediction. (Navarro-Espinoza, et al., 2022) used five machine learning models; Gradient Boosting, Random Forest Regressor, Linear Regression, MLP-NN, and Stochastic Gradient Regressor to predict traffic flow without the need to completely change the existing traffic light system. Under the computational technologies, another useful technique is data mining algorithms like ANN and KNN which can be used to do traffic analysis, density estimation, and prediction of traffic management (Chavhan & Venkataram, 2019).

## **2.4 Machine Learning-Based Method**

Over the years, in an attempt to achieve better prediction of traffic flow, different prediction methods have been proposed. Researcher in science and technology fields have adopted both mathematical modeling methods and statistical modeling methods – both parametric and non-parametric. As of today, the most widely used of all the aforementioned methods is the non-parametric Machine learning-based method (ML). The non-parametric method requires no or less prior knowledge about the relationship among the different traffic patterns, less strict assumptions on prediction tasks, and is believed to fit the non-linear features that exist in traffic data (Wang, 2021). Under the non-parametric machine learning methods, there are different sub-classes models such as the regression model, support vector machine regressor, gradient regressor, kernel-based model, etc. According to (Wang, 2021), in designing an effective prediction system, it is important to always select the appropriate type of machine learning model using some established parameters and specifications. The use of accuracy, F1 score, recall, as well as the design goal of the model is used for appropriate model selection for any system prediction.

For this study, the machine learning models adopted for the traffic flow prediction will be based on the mathematical definition which is given as;

Where is the prediction function for the model used for the traffic flow, is the input data, represents the context features related to the task, for example, weather condition, social media information, policies, etc. represents the spatial feature related to the location where the prediction is made., and parameter represents the prediction results. Furthermore, is the number of spatial points corresponding to both the input and output, represent the number of channels of input and output, and stands for the length of the time-steps in the input and output. To have a clearer view of the parameters in the model, it is important to define the tasks input and output. Based on past research the input, is the traffic flow prediction while is the traffic speed prediction. (Okawa, et al., 2017) through their research collected traffic dataset from the freeway in the Greter Tokyo area and used traffic flow variable as the input to the prediction model. Also, (Pinlong Cai et al., 2016) used traffic speed dataset collected in Beijing, to predict the value traffic speed on Liuliqiao district road segment based on the current timestamp. In other research, some authors performed a multi-task prediction through the combination of both traffic flow and traffic speed. (W. Huang et al, 2014) predicted traffic flow and speed at the same time adding multi-task layer at the end of a Deep Belief Network. Also, the research of (Guo, et al., 2017) was based on predicting a daily traffic flow in the next six timestamps by using the iteration method and the auto-correlation coefficient method.

Machine learning is a broad topic with different types of learning algorithms based on scenarios and users’ perspectives. In this study, some of the selected models that will be reviewed are categorized by different ML algorithm theories as shown in figure 2.1

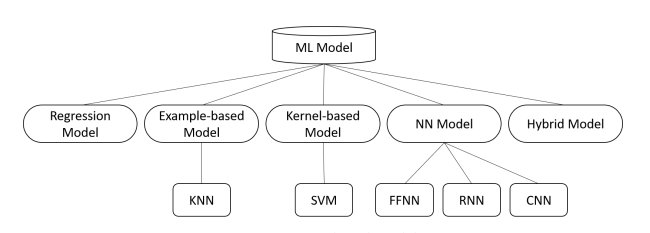


Figure 2.1 Machine Learning-based Models

*Source: Understanding ML* (Shai & Shai, 2014)

This study will select and compare the performance of the regression model, the example-based model (KNN), the kernel-based model (SVM), and the feed-forward neural network model (FFNN) in predicting traffic system flow. A brief explanation of the chosen model is presented below:

* Regression Model: This ML model studies the relationship between a numerical dependent variable and the independent variable, with the assumption of a linear relationship between the input and the output.
* Example-based mode: This ML model undertakes the prediction task by comparing the similarity between the input sequence and the historical data samples, thereby using the obtained samples to make the final prediction for the outcome variable.
* Kernel-based model: The kernel-based model uses a kernel function to map the input data into a high-order vector space, where the prediction tasks are easy to solve.
* Neural-Network Model: This ML model uses neurons just like the human brain, and information is simulated in such a way to pass through the neurons. Specifically, the input data is passed through different network structures in different types of NN models. The structure of the NN model requires that the input data be transformed into an activation signal through the use of activation functions.

The theoretical and application of the highlighted ML models on traffic prediction tasks will be reviewed next.

### **2.4.1 Regression Model**

Regression models are common and widely used in machine learning. A traffic forecast approach can be classified as parametric if it makes assumptions about how traffic patterns will be distributed. Regression models are used to anticipate traffic since they are simple to apply and suitable for such jobs on a non-complex traffic network. The accuracy of model predictability can be significantly impacted by the choice of mathematical models and their parameters. Under parametric modeling, there is always a clearly stated relationship between the model functions and the corresponding parameters between the input and output variables (Trevor, et al., 2009). Researchers have established that parametric methods like a regression model can give relatively good predictions based on the satisfaction of some underlying theoretical assumptions.

Recent studies reviewed show that when it comes to traffic prediction for the intelligent transport system, the regression model has been widely used. The outcome variable is considered to be a linear combination of other existing traffic variables, which means the researcher can obtain a good prediction result just by selecting the appropriate weight parameters and traffic variables. According to (Rice & Zwet., 2004), combining linear regression with time-varying coefficients yielded a good performance of the linear regression model on the small traffic dataset. The author established that the regression model performed better than the kernel-based model KNN even though did not give a specific prediction based on the accuracy measure of the data. Also, in the research work of (Jaimyoung, et al., 2000), the authors examined the relationship between future traffic flow on a given link, and the origin links and adjacent link’s traffic flow records. By using three combined prediction models, each of these was combined with different traffic variables such as current traffic, historical average, and upstream traffic, and a strong effect between upstream traffic and current traffic was observed. Another researcher combined linear regression with a stepwise-variable-selection method and tree-based method to make a prediction for traffic flow, using occupancy and historical travel-time information for freeway districts in Australia (Jaimyoung, et al., 2000). (Kaouther, et al., 2011) employed cross-validation technique while predicting traffic headway from 0-60 minutes, and established that using a more recent traffic dataset was very useful for short-term traffic forecasting. It was gathered that the use of a tree-based method can greatly improve the efficiency of finding the parameter by controlling the depth of the tree. However, getting prior knowledge of a given road section is completely dependent on the search range that would be initialized.

Highlighting the limitation of the linear regression method, (Xiang, et al., 2011) noted that part of the drawback of using the simple linear regression model is when the model could not outrightly capture the non-linear relationship in traffic flow. To better make provision for the stochastic variations in the traffic information, researchers will most time hybrid the regression model with other methods. (Li Li et al., 2015) provided a dual Lasso phase model with Granger causality theory to solve the traffic flow prediction, demonstrating yet another method for enhancing the model's ability to capture the non-linear aspects of traffic data. A historical average method would be used to preprocess the data in order to identify the residual trend of the initial input sequence before making the forecast. The first Lasso regression followed. The records that weren't related were filtered using the first lasso regression model. To lessen the impact of outliers on the predictability strength of the model, the robust Lasso regression, which is the second Lasso regression, is employed. The Alternating Direction Method of Multipliers was used, as mentioned in the study, to solve the second Lass regression phase in a two-step fashion. Also, (Maya, et al., 2017) combined a latent factor model with a bi-linear Poisson regression model for a short-term traffic flow prediction task using the short-term historical data from the relevant road segments to further address the issue of the model's robustness and accuracy for short-term prediction. Bi-linear Poisson regression was used as the foundational model. To address the issue of the time lag between various road segments, the authors included the concept of a convolutive mixture in the model. A stochastic variations Bayes model was added by the authors to the regression model to make it appropriate for live updating and prediction applications.

Furthermore, in the past ten years, various researchers who have adopted the regression model appraised the simplicity in terms of implementation but criticized the simplicity of the model structure as weak and not robust enough at describing the continuous features of the traffic dataset. Nevertheless, when it comes to the prediction tasks for small and not too complex structures traffic networks, regression model can still be considered as a good choice model due to less computational stress. Over a considerable amount of time, the regression model has experienced remarkable popularity. The right set of parameters has a big impact on how accurate the model is. Researchers in this field have experimented with a variety of techniques, from the least squares method to the stepwise-variable selection method and tree-based methods (Jaimyoung, et al., 2000). The model's capacity to detect brief traffic vibration has significantly increased due to the advancement of the parameter-finding method. In their model, the authors also made an effort to use the spatial relationship data between each road section. The regression can successfully complete the multiple-road-section prediction task with high accuracy and resilience, according to the most recent research (Maya, et al., 2017).

Based on the reviewed papers, the regression model can be seen to be applicable in traffic prediction tasks due to the fact building the model requires the use of small historical datasets. However, to a certain extent, it restricts the accuracy potential of the model because the model does not reveal a microscopic picture of the dataset, such as its seasonal variations, and can easily experience an overfitting issue.

The regression model equation is given as;

Where;

### **2.4.2 Example-based model (KNN)**

The KNN model is widely used for traffic prediction because it has excellent properties that perfectly capture the spatial relationship between the road segments in the traffic network. In some cases, the model has also been used to exclude traffic data that is unrelated to the current prediction objective. A non-parametric regression technique based on data is the k-nearest-neighbor model. Instead of creating a formal prediction model, it looks for the K nearest neighbors that match the values of the current variables and uses those K data to forecast the value at the following period. In this strategy, historical data is used to gather the knowledge required for prediction (Wang, 2021). Numerous trends and typical traffic status patterns are contained in the historical database. Each sort of data in the dataset reflects a potential evolution trend in traffic. Because the algorithm assumes that the historical dataset contains the relationships between all of the traffic system's components, the historical database's quality—specifically, whether it contains all of the potential future traffic states—has a significant impact on the precision and outcomes of predictions.

To make traffic predictions, many researchers have used the KNN model. According to (Brian L, et al., 2002) four major obstacles must be overcome for the KNN model to be used for traffic prediction. The first one is how to define an appropriate state vector, the second is how to generate a prediction or forecast, the thirdly is how to define an appropriate distance metric, which is related to how similarity between data points is determined, and the fourthly, how to manage the potential neighbor database. When utilizing KNN for traffic prediction, it is typical to first choose K candidates from the dataset using Euclidean distance, and then output the results using the weighted average approach. This technique was employed by (H Chang et al, 2012) to forecast traffic flow at various prediction intervals (15, 30, 45, 60-min). To solve the issue of k selection, The distance measure used by (Zuduo & Dongcai, 2014) to choose the more related neighbors was the correlation coefficient distance. The time elapsed between the present time and the prior time, in addition to the data records themselves, had an impact on the neighbors’ distance (no more than one day).

Also, to prevent the influence of the dimensionality reduction of k due to the phenomena overlap while using the usual selection procedure, the candidates for the prediction were chosen using the local minimum algorithm. (Zuduo & Dongcai, 2014) used a Linearly Sewing Principal Component technique (LSPC) to create the final prediction, which forces the prediction into a minimization problem. It was established that the model did better in the tests on various road segments than other models. The issue of spatial correlation effects most times occurred as a problem in temporal features in the traffic flow dataset. For predicting traffic flow, (Dawen Xia et al. , 2016) the authors presented a MapReduce-based KNN model. The fact that this model takes the spatial influence into account sets it apart significantly from the KNN described before. The supplied road segment, along with its upstream and downstream, determines the distance measure. By employing equivalent distances that were influenced by the grid distance between the two specified road segments, ( Pinlong Cai et al., 2016) first reduced the pool of probable possibilities. In order to exclude the road segment with the weakest spatial correlation, this spatial correlation ratio was compared to a predetermined threshold. To identify candidates for the final forecast, the Gaussian-based Euclidean distance was determined. Finally, a prediction regarding the distance metric was created using the Gaussian weighted average algorithm. Based on the papers reviewed, one of the advantages of using KNN for traffic flow prediction is the ability to detect and capture useful information about a complete road network and not only the segment.

The classification performance of KNN is determined theoretically by the distance of the query's *kth* nearest neighbor. For multi-class k-NN classification, Cover and Hart (1967) prove an upper bound error rate of

where is the Bayes error rate (the minimal error rate possible), is the K-NN error rate, *M* is the number of classes in the problem. For *M =2*,and as the Bayesian error rate approaches zero, this limit to “not more than twice the Bayesian error rate”. The *'K-Neighbors Classifier'* algorithm to produce the model, the *'n-neighbours'* parameter was set to '5'. The *'n-neighbours'* variable's value is chosen at random, which can give the optimum by iterating a range of values, then fitting and storing the predicted values in the knn variable.

### **2.4.3 Kernel-based model (SVM)**

The Support Vector Machine (SVM) is a statistical learning theory model useful for classification and regression problem proposed by (Vapnik., 2013). SVM can avoid the drawback of being more easily prone to local optimum than other nonlinear prediction models since it is based strictly on statistical theory and adheres to the structural risk reduction concept. Many people refer to SVM as Support Vector Regression when it is used to solve a regression problem. SVM is frequently utilized in classification problems (SVR). In order to solve regression problems, SVM's main principle is to establish a hyperplane as the decision surface for a given training sample, ensuring that all sample points are in close proximity to the hyperplane and that the total deviation of the sample points from the hyperplane is kept to a minimum.

The hyperplane is often formed by a linear function when the classic SVM model tackles regression issues, however the traffic prediction is a nonlinear regression issue. In light of this, the researchers used the kernel function to convert the traffic prediction problem into a high-dimensional linear regression problem. As shown in table 2.3, is replaced with which is known as the kernel function. The kernel function enables to operate in high-dimensional space with low-dimensional input data, and we are not required to understand the transform function.

**Table 2.1: Kernel Function**

|  |  |
| --- | --- |
| **Kernel Name** | **Kernel Function** |
| Linear |  |
| Gaussian (RBF) |  |
| Polynomial |  |

Estimation of traffic prediction in (Vanajakshi, 2004) contained the first instance of SVR being employed. SVR was used to forecast the volume of traffic at a specific intersection. One hour was chosen as the prediction interval, and the previous five-time steps were utilized to forecast the upcoming hour, which may be thought of as a long-term application scenario. The testing data was gathered from 6 p.m. to 10 p.m. on the same day as the training data, which was taken from 8 a.m. to 4 p.m. Error rates on average were 6.03%. In (Vanajakshi, 2004), a Multi-Layer Feed-forward Neural Network (MLFNN) and an SVR using RBF as the kernel function were tested to forecast traffic speed over a 2-minute period. The model produces speed data for the upcoming two, four, and six minutes as its output. The input data are the speed data from the previous 10 minutes. The authors' results demonstrated that SVM can outperform MLFNN when the training dataset is compiled from the previous two days. While the accuracy of NN grows and surpasses SVR when the training set is expanded, the Mean Absolute Percentage Error (MAPE) of SVR only marginally increases. This indicates that SVR is more advantageous when the training set is limited. SVM is a kind of kernel-based model, as earlier stated. The accuracy of the model will be significantly impacted by the kernel function that is selected. RBF is the most well-known kernel function. Also, in (Manoel Castro-Neto et al., 2009) and (Wei-Chiang Hong et al., 2011) both authors employed RBF as their kernel function. Similarly, other researchers have attempted to take the kernel function into account for seasonal aspects. Two seasonal kernel functions—seasonal RBF function and seasonal linear function—were added to the SVM model by the authors (Marco, et al., 2013) in order to help it use the seasonal information present in the traffic records. The kernel function will additionally take into account the time interval data in relation to the selected seasonal period. The results obtained from the experimental procedure show that although the seasonal SVM model is more competitive than the SARIMA model, the training time for it is double. To demonstrate that their model is the most effective at forecasting future occurrences in the actual world, the authors looked at the training time and prediction time for a different model.

After selecting an appropriate kernel function, it's crucial to optimize the SVM's parameters. The challenge of parameter optimization is to avoid entering a comparatively lower local optimal while maintaining a low time consumption. The SVM model's parameters were optimized (Wei-Chiang Hong et al., 2011) using the Continuous Ant Colony Optimization (CACO) algorithm. The concept of ACO was used to construct CACO (M. Dorigo, 1996), although the search space in CACO is configured to be consecutive. Another optimization algorithm was used in (Mei Duo et al. , 2017), to raise the SVM model's accuracy, the authors developed a technique called Genetic Particle Swarm Optimization (GPSO). The input sequence was divided into a number of sub-frequencies at various fluctuation levels using empirical mode decomposition (EMD). Some researchers have concentrated on the data pre-processing stage in an effort to increase the SVM's accuracy. To deal with the chaotic and non-stationary elements in traffic speed records, (Jin & Qixin, 2013) proposed an SVM model. The paper's fundamental idea is to map the data onto a high-level space that can easily capture chaotic and nonlinear properties. To do this, decomposition and wavelet modification were initially used to de-noise the raw data input into the SVM model.

To handle the prediction task under unexpected traffic scenarios, an online-SVM model (OL-Model) adopting RBF kernel has been created. During peak hours, the model outperformed the Gaussian maximum likelihood approach in both typical and unexpected traffic scenarios (Manoel Castro-Neto et al., 2009). The results show that while the OL-SVM did not perform as well as the GML technique in a typical scenario, it did perform better in an uncommon scenario. Multi-step prediction is still another issue in practice. The authors attempted to resolve the multi-step prediction challenges in (Zhang Mingheng et al., 2013). The structure of the input data is what distinguishes the model in this research from the original SVM. The input data may include records from the past, such as the most recent time period, as well as records from further upstream. The experiment's findings show that when the prediction interval is set to 1, inputs with upstream records and previous records perform better because, in a short time interval, the upstream traffic flow has a significant impact on the provided road section. The accuracy of all models using various types of inputs decreased from 1 to 3, even if the prediction interval increased. However, the model that used historical and upstream data fared better than the other models. The history data becomes more relevant to the prediction task as the interval widens. In the end, the authors also noted that the input including only the prior and historical records of the specified road portion was the optimum option for implementation when taking into account computation usage. In this section, we compared the SVM models applied to traffic prediction tasks during the previous 10 years, the majority of SVM models used RBF as their kernel function because of its excellent capabilities for handling non-linear fluctuation. We also observe that the accuracy of the SVM model can be significantly improved by data preparation, particularly the decomposition procedure. The SVM model served as the foundation of the online traffic forecast system used by the authors in (Zhang Mingheng et al., 2013). The explanation for this is because SVM has a lower computational need, which is appropriate for the quickly updated requirement in online traffic prediction applications.

SVM does have a weakness, however, as it is constrained by the nature of its structure and finds it challenging to handle current trends such as large traffic networks. Additionally, using grid search increases the cost of discovering acceptable hyper-parameters compared to KNN or regression models (Zhang Mingheng et al., 2013).

## **2.5 Neural Network-based Model**

The neural network-based (NN) models also can be categorized under the heading of non-parametric models. Non-parametric models don't always mean they have no parameters at all; rather, they might learn their parameters by studying past data (Brian, et al., 2002). These models can be used without any prior knowledge as long as there is enough historical data, which is advantageous in the current age of information explosion. The model creates an inherent connection between prediction target and historical data by learning historical data. These models can be used without prior knowledge as long as there is enough historical data, which is advantageous in the current age of information explosion. The model creates an inherent connection between the prediction target and the past data by learning it. According to various earlier research, the non-parametric method has a much higher degree of precision than the typical linear regression model because it is better able to capture the non-linear aspect of the traffic pattern (Simon Oh et al., 2015). However, there are drawbacks to the non-parametric approach as well especially when compared to other types of approaches, the cost of training a non-parametric model might be costly due to the model's strict requirements on the quantity and quality of historical data (Jabot., 2015).

The complicated linear system can be more accurately identified using the NN model, an abstract mathematical model. In the meantime, the NN model makes use of the black-box learning model, which is excellent at predicting traffic. To determine the natural relationship of the variables in the dataset, no empirical formulas are necessary. The NN model automatically develops and modifies the input and output map model by learning a large number of input and output samples. The NN model is frequently employed in traffic prediction because it has the characteristics of associative memory, good fault tolerance, and resilience. Additionally, the NN model has the ability to update the network in accordance with real-time traffic data, enabling real-time prediction. Numerous elements that can be detected by ANN affect the traffic system. It makes use of the study road section's historical data, as well as the pertinent road section, and numerous things that affect the transportation system, like the weather, road work, accidents, road conditions, etc.

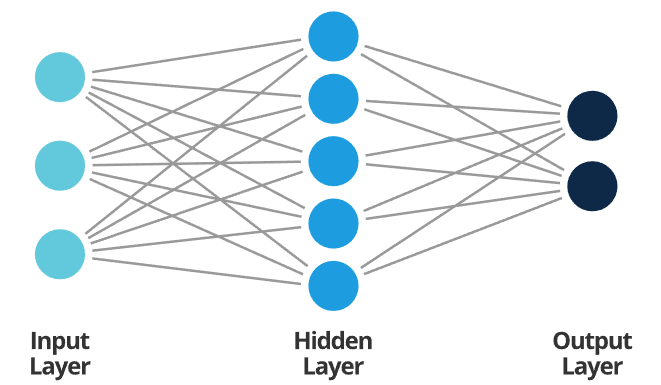


Figure 2.2: Artificial Neural Network Architecture

*Source: Smartsheet* (smartsheet, 2022)

However, for the neural network's training process a significant amount of data is required, and inadequate data could produce subpar prediction outcomes. The trained network can only be used for the current road section; if the traffic or road conditions change, the network will no longer be appropriate. As a result, the NN model's promotion capability is weak. Another issue is that each buried layer's amount of neurons must be calculated by experience. Network topology with too many neurons in each hidden layer will be large, taking a long time to calculate (W. Huang et al, 2014). The size of the number of neurons in each hidden layer determines the accuracy of the model prediction results. By implementation, there are different types of neural network models such as feed-forward, recurrent network, and convolutional network. This study will only adopt and review the feed-forward neural network (FFNN).

### **2.5.1 Feed Forward Neural Network (FFNN)**

Artificial neurons are the fundamental building blocks of neural networks; *Fig. 2.2* depicts a typical neuron structure. The input vector to the neuron can be described as if we suppose that a neuron accepts d inputs. The neuron's output will be:

Where is a d-dimensional weight vector, b is the bias variable to the neuron, and is the activation functions, which can enhance the expressive ability and learning ability of the network. Table 2.2 below shows the common activation function in neural network.

**Table 2.2: Activation Functions in ANN**

|  |  |
| --- | --- |
| **Name** | **Activation Function** |
| Logistic |  |
| Tanh |  |
| ReLU |  |

According on the order in which they receive information, each neuron in the FFNN is classified into various groups. Each group can be thought of as a layer of the brain. Each layer receives as input from the layer below it and outputs to the layer above it. Unlike a recurrent neural network, the entire network only transmits information in one direction; there is no reverse information transfer. The FFNN's internal calculation is shown (Shai & Shai, 2014) .

The one-hidden layer structure is the simplest and most user-friendly FFNN structure. (Jin & Shiliang, 2008) predicted traffic flow in a single layer using a one-hidden layer FFNN. The shallow FFNN was extended with a Multi-Task Learning (MTL) layer for a multi-interval prediction. The predictions focused on the traffic flow at time T in the future as well as the brief window before and after that. The other two tasks were used to use the correlations between various time intervals to increase the accuracy of the main job, which is the prediction for time T. The Levenberg-Marquardt algorithm was used as the optimization technique by the authors in order to help the NN converge more quickly and precisely. The experiment's comparison of the MTL FFNN with the single-task FFNN revealed that the MTL FFNN performed better than the single-task FFNN on various road segments.

One-hidden-layer Further research on FFNN's capacity for both long- and short-term traffic prediction was found in (Gültekin & Borat., 2010). NN receives the traffic volume records for the most recent time step, the weekday, and the month. The experiments had a few flaws, despite the authors' claims of hopeful results: The model was used to forecast traffic flow in the remaining years, despite the fact that its training data comprised only 20% of the entire year. The experiment's findings show that the model has a hard time adjusting to the volume shift. This can be the case because the dataset's non-linear features are too many for a simple structure FFNN to capture. The algorithm also has trouble adjusting to an unusual scenario in which the volume of traffic peaks during the day.

Extending the dimension of the input, or including more different sorts of traffic records, is one technique to increase the prediction model's accuracy. A shallow FFNN was used in (Kranti Kumar, 2013) to anticipate traffic flow on a road with a variety of vehicle kinds. Along with statistics about the volume of various vehicle kinds, the day of the week, and the time of day, the paper also considered the top speeds of various vehicles. The results of the experiment show that the Sigmoid activation function outperforms the tanh function. Momentum optimization is less effective than Levenberg optimization. Additionally, the model will be more precise if there are more hidden neurons. The impact of the speed records from the dataset was further examined by the author. The outcomes demonstrated that when more data was collected, the model's accuracy improved.

# **CHAPTER THREE**

# **Methodology**

## **3.1 Introduction**

The techniques employed to construct the prediction system will be introduced in this chapter. With not only suggests a prediction model but also offers implementation options for using the model in ITS. The information and control technologies used in this study can be divided into four basic parts:

A. Data gathering or collection B. Processing of data C. Modelling D. Decision-making system

## **3.2 Data collection**

The need for sufficient data is very important for implementing a machine learning model. The collected data will be used to train the mL model to make the model converge. A good dataset must provide sufficient types of pattern information according to different traffic prediction task. The fundamental data includes data on traffic flow, traffic speed, or both. The dataset should contain the geographic details of each data collection as well as the traffic patterns and connections between the data collectors when examining the spatial-temporal relationship. When examining the effects of holidays, notable occasions, or weather variations, the dataset may occasionally contain sufficient pertinent data. The dataset's aggregation interval is another important consideration. A longer aggregate dataset has a smoother trend of the time-series, which will have a significant impact on the structure of the prediction model. A shorter aggregation period will have more random and non-linear features. This study focuses on the dataset collected from the real-world on-road sensors for a particular time duration. These datasets include traffic information gathered from on-road loop detectors, toll booths, or other on-road or road-side equipment. The government or organizations then make these datasets available for research purposes.

## **3.3 Processing of Data**

There are different methods of processing data, especially data with some level of randomness or noise. This section will discuss the two popular selected approaches to data pre-processing.

### **3.3.1 Data normalization**

Mapping the original data records into a specified range can be referred to as data normalization. Because most activation functions employed in NN models, like the Sigmoid function or Tanh function, map the input into a fixed range, this is what is required. For the original records, the range, however, varies arbitrarily. In order to make the original records fit for the activation functions, we use the normalizing function. This can speed up the convergence of the model. To standardize the data on vehicular traffic flow in our study, we employ the MinMaxScaler function, which can be described as follows:

Where is the original data and is the data after scaling. Commonly, this standardized method is to scale the data between 0 and 1.

### **3.3.2 Moving Average**

According to (Xianglong, et al., 2019) and (Xiaolei Ma et al., 2015), too few stable vibrations will be present in vehicle traffic flow recordings with too short record intervals, which are not only unhelpful for making driving or control decisions but also have an impact on the prediction model's accuracy. However, for data on vehicular traffic, an interval that is too lengthy will lose many seasonal patterns. The Moving Average (MA) approach is applied in this study to convert a short interval record into the necessary interval duration. The formula below can be used to determine the MA for the first n records:

Where represent the record at time i. In order to obtain the following records, the formula below is used:

## **3.4 Training algorithm**

When developing a prediction system using machine learning-based models, especially when it is in a deep learning structure, we must take into account the time cost. A model that has been trained on a particular dataset is typically only applicable to that dataset; using the model to predict data from other datasets will significantly reduce its accuracy. It implies that all of the detector models on a particular transportation network might have a similar structure. If the system needs to build a new model for each detector and train each model individually on a relevant dataset to assure the correctness of the entire prediction system, the time cost will be significant.

In (Jiahao & Azzedine, 2020), the author presents an off-line optimization approach called desensitization to shorten the total training time while implementing an ML-based model. During implementation, the algorithm makes an effort to prevent duplicate training. As established in (Jiahao & Azzedine, 2020), despite the fact that data from various detectors collected at the same time of the day may differ, the overall trend in traffic flow is essentially the same over the course of a single day. Differences can be further reduced by employing normalization methods. With the help of some necessary parameter tuning, we may employ the parameters of a well-trained model across the entire network.

The desensitization algorithm still has drawbacks. The model is trained inside a while loop to satisfy the threshold criterion, as we can see in 1, to guarantee the model's correctness for a particular dataset. Focusing only on training accuracy, however, is likely to result in an overfitting issue, especially if the training dataset is limited. When the model's structure is relatively complex, this circumstance will become more prevalent.

## **3.5 Model design**

This section examines the framework and the design of each of the model adopted for this study

### **3.5.1 Kernel based model**

According to studies by (Schölkopf, et al., 2002) the new mapping method produced a scenario that is linearly separable, whereas non-linear kernels are needed to map the input information to a multi-dimensional space. Despite this, the kernel function is designed to create operations in the input space, not the possibly high-dimensional feature space. More so, the inner product does not require to be evaluated in the feature set; the mapping is accomplished simply replacing the inner product regarded as the kernel, , where is called a kernel function. If a kernel function K can be obtained, it can be utilized for training without having to know the specific form of. This is useful since our higher-dimensional feature space may be infinite-dimensional and thus impossible to compute at times. Symmetrically, ; By positive semi-definite . Then,

Substituting equation 3.37 into 3.36 becomes;

The Linear, Polynomial, Radial Basis, and Sigmoid kernel functions are the most often used kernel functions, with the basic mathematical function provided as:

In equation 3.52, *K* is the kernel function, *N* is the number of training data points, *xn* represents vectors used in the training process, *x* stands as an independent vector, and *b* are the parameters obtained by maximizing the objective function. However, each of the kernel functions given above has a specific parameter that must be improved in order to achieve the best results.

Furthermore, when the data is linearly distinguishable, such that, it can be segmented using a single line, the Linear Kernel is utilized. It is one of the most often utilized kernels. The function is:

The Sigmoid Kernel also known as the Hyperbolic Tangent Kernel is very popular for support vector machines due to its origin from neural networks

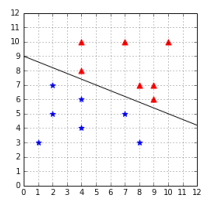
### **3.5.2 Support vector machine**

Examining a supervised binary classification problem. Given that the data is expressed as and where is the number of samples, for class and for class .

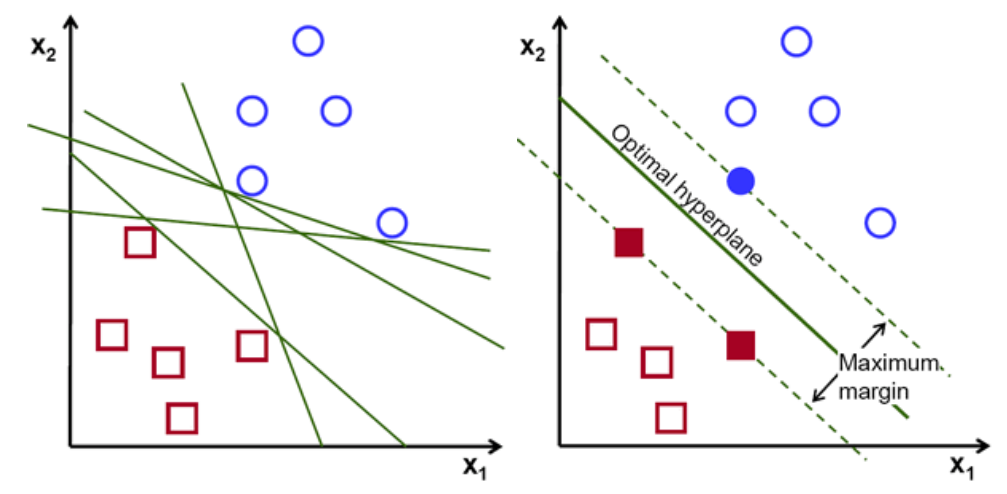
Assuming that both classes are linearly distinctive. It means that there is a possibility of finding at least an hyperplane defined by a vector with a bias , which can separate the classes without error as mathematically narrated in the first casedescribed later.

More so, more than one hyper-planes could be fitted to distinguish the two classes, but there exists a single optimal hyper-plane that is required to generalize better than other hyper-planes(figures 3.1 and 3.2)

The essence is to discover the hyper-plane that achieve the maximum margin between the classes. To determine the optimal hyper-plane, the support vectors have to be defined. The support vectors lie on two hyper-planes which are parallel to the optimal such as in the second case.



**Figure 3.1 A hyperplane separates a data**



**Figure 3.2: (a.) Possible hyperplanes and (b.) an optimal hyperplane separating a dataset**

**Estimation procedure of SVM:**

**First Case:** The **perfect separation optimization** equation model is described thus:

Equation 3.5 is presents the distance from a point to a hyperplane

(in essence, at (), , which is the maximum-margin decision boundary the we intend to maximize).

The optimization procedure that produces and is described thus:

Note:

Equation 3.11 expresses a perfect separation at coordinates . But then, the distance of nearest coordinate to H can then be derived as presented in equation 3.12.

Rescaling and ;

Let , cw, b

Note that is the typical vector to the hyperplane and is the input vector.

The separating hyperplane is the plane. The nearest points lie on the planes . The margin is

The quadratic, constrained optimization procedure is a means of maximizing the margin.

Note that is the training instances, and is the right output of the Support Vector Machine for the training instances. The value is is +1 for the positive instances in a class and –1 for the negative instances.

Employing a Lagrangian expression, this optimization challenge can be transformed into a dual form, being is a Quadratic Programming problem in which the objective function is only reliant on a set of Lagrange multipliers,

A one-to-one interaction exists between individual Lagrange multiplier and individual training instances. Immediately the Lagrange multipliers are concluded, the typical vector and the threshold can be gotten from the Lagrange multipliers by differentiating equation 3.19 with respect to and ,

Since can be pre-estimated before application, from the training data through equation 3.20, the amount of computation needed to examine a linear Support Vector Machine is the same in the number of non-zero support vectors.

,

Subject to the inequality constraints, and a linear equality constraint,, by putting both equation 3.21 and 3.22 into equation 3.19, it becomes:

Hence,

The dual nature of the optimization problem for support vector machines in equation 3.24 indicates that a search for the maximum-margin decision boundary is equal to a search for the support vectors; meaning that they are the training instances with non-zero Lagrange multipliers, and via they entirely conclude the decision boundary. More so, this shows that the optimization problem is entirely described by a pairwise dot products between training instances.

**Second Case: The Non-perfect separation optimization equation model:**

For non-separable categories, the optimization procedure requires to be modified efficiently. Mathematically, the maximal margin hyper plane for non-separable data is chosen by minimizing the cost function: Considering

where stands for a weight vector, and represents a bias value. To maximize separating margin, and minimize error, we can use objective function:

Note that is a regularized constant greater than 0 to carry out balancing between the training error and model flatness.

Variables are slack variables. The essence of the optimization task is to increase the margin as much as possible and decrease the number of misclassified instances with . The parameter is a positive constant that manages the relative influence of the two essential terms.

By introducing Lagrangian multipliers, for the constraints, results in the Primal Lagrangian:

A search for a stationary point of from equation 3.28 then begins by taking its partial derivatives with respect to , and , followed by setting its gradients to zero:

Utilizing and , putting equations 3.29, 3.30, and 3.31 in equation 3.28, produces the Dual Lagrangian below:

Hence,

### **3.5.3 Artificial Neural Network**

At the hidden layers, each neuron computes, a weighted sum of its input signals, , for and then applies a nonlinear activation to produce an output signal,. A neuron is mathematically represented by these pair of equations:

and

In this study, the rectified linear unit activation function is used to transform the output limited into an acceptable range which as well, been adopted. It can be given by the following expression in equation 3.45:

Thus, the output layer will be obtained by the function in equation 3.46

Where; denotes the output of the overall networks, represents the ReLU transfer function of the layer of the network, is the bias of the neuron in the output layer.

However, the process is varied by adjusting the weights, which could be easily done given the small size of the network. In a case where a large network of many neurons is needed, adjusting the weights could become tasking.

Methods of adjusting the weights have however been discovered. Thence, the type of ANN used in this study is a feed-forward multilayer perceptron (MLP) with back propagation (BP) learning algorithm.

**Back propagation learning algorithm of Feed forward neural net**

For the weight adjustment and threshold coefficient, the back-propagation algorithm employs the steepest-descent learning approach. In different terms, the created artificial neural network's training approach is the back propagation algorithm. Recall equation 3.40:

where s s are the inputs, 's are the respective weights applied.

In this situation, activation is solely determined by the inputs and weights associated with them. The neuron is said to be linear if the output function equals the activation. However, this demonstrates that we are simply attempting to fit a straight line. Then we'd use a sigmoid activation function in conjunction with the linear activation function. The hyperbolic tangent, logistic sigmoid, bipolar sigmoid function, and the (Elliott, 1993) transfer function are all examples of sigmoidal functions. The logistic sigmoid function will be considered for the purposes of this study. The backpropagation algorithm can be employed to train the developed Feed Forward Neural Network (FFNN). Consider a three-layer multiplayer feed forward. Its functioning can be explained using the equation below:

Where, and are the respective outputs of the and layers of the networks. is the bias vector of layers of the networks, where M is the number of layers of the neural network.

The neuron of the first layer obtain input:

Equation 3.42, gives the initial condition for Equation 3.41. The outputs of the neurons in the last layer can be seen as the overall networks outputs:

The task is to train the network with associations between a specified set of input-output pairs {(, ( where is an input to the network, and is the corresponding target output. As each input is applied to the network, the network output is compared to the target.

The sigmoidal activation function combines nearly linear behavior, curvilinear and nearly constant behavior depending on the input value. The sigmoid function has non-linear characteristics, monotonically increasing and continuous differentiable.

In the training process, the weights are being adjusted in a way such that the error is minimized. This is achieved by observing the difference between the actual output and the desired output

In order words, to adjust the weights appropriately, the back propagation will calculate how the errors depend on the outputs, inputs and weights. Thereafter we employ an appropriate method, steepest descendent rule to adjust the weight to meet the required target.

## **3.6 Model Performance Evaluation Metrics**

### **3.6.1 Root Mean Square Error**

The Root Mean Square Error (RMSE) is frequently used in assessing performance of each model for both training and predicting data.

RMSE =

stands for the actual value and as the predicted value.

### **3.6.2 Accuracy**

Accuracy is the referred to as the percentage of right predictions the trained algorithm could make on the test dataset. The formula is given as:

### **3.6.3 Precision**

Precision is the fraction of important instances (in essence, True Positives) amidst the overall instances that were predicted to be in a particular category. It is calculated thus:

### **3.6.4 Recall**

Recall is the fraction of instances that were predicted to be part of a category relative to the overall instances that are truly part of that category. It is calculated thus:

### **3.6.5 Area under curve (AOC)**

AUC gives an aggregate metric for the performance of an algorithm for all possible classification benchmarks. It can be referred to as the probability that an algorithm scores a stochastic positive instance higher than a stochastic negative instance. It is scale and threshold invariant. It is expressed as:

### **3.6.6 F- Measure**

F-Measure is an approach that binds both precision and recall together as one measure, while reflecting the essence of both measures. This is because both recall and precision cannot individually tell the overall performance of an algorithm.

# **CHAPTER FOUR**

# **DATA ANALYSIS AND RESULT PRESENTATION**

## **4.1 Introduction**

In this chapter, the details of dataset that will be used for traffic flow prediction task will be presented. Also, the use of base machine learning model, specifically the K-nearest neighbor regressor and support vector machine regressor, and artificial neural network will be evaluated. Comparison based on some selected performance metrics will be used to select the best model which will be used for traffic flow prediction. Moreover, data pre-processing technique used to ensure stationarity and the use of customized plots for data exploration will be discussed. In the end, selection of a new data instances will be used to test the predictive strength of the selected model.

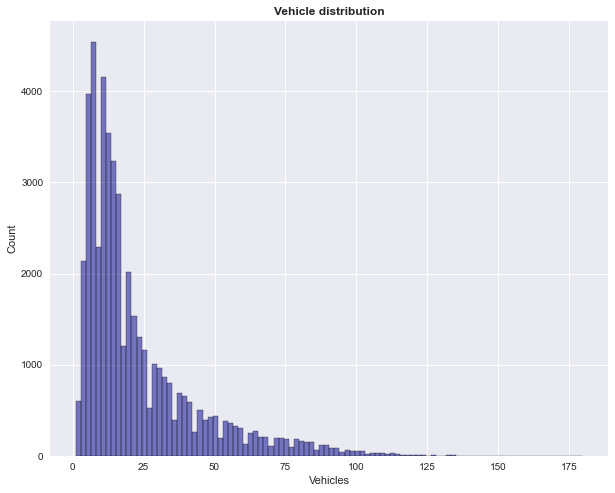
## **4.2 Dataset Overview**

Traffic is one of those annoying problems that affect many of us living in urban settings. One of the causes of traffic is the increase in urban populations. While the infrastructure is old and can only accommodate a limited population there is an influx of residents in search of livelihood and opportunities. Traffic congestions lead to an increased in the combustion of fuel. It further increases the carbon emissions causing air pollution. It also costs time and money. According to (INRIX, 2019), a transportation analytics and connected car services, 2020 report found that on average, Americans lost 99 hours a year due to congestion, costing them nearly 88 billion dollars in 2019, an average of 1,377 dollars per year. From 2017 to 2019 the average time lost by American drivers has increased by two hours as economic and urban growth continues.

This study will explore that dataset corresponding to four junctions which will be used to train the three selected machine learning models for traffic flow prediction. The dataset was sourced from [link](https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset). It contains 48120 observations of the number of vehicles each hour in four different junctions. There are 5 columns namely; DateTime, Junctions, Vehicles and ID. Four different sensors were placed on each of the junctions and each sensor were collecting data at different times or period or three years. The information covered by the sensors on each junctions differs significantly, some of the junctions have provided sparse data compared to others.

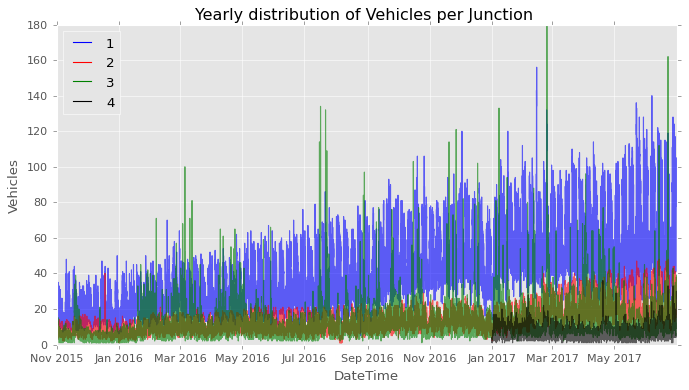
## **4.3 Exploratory Analysis**

Before a data can be set up for machine learning modeling, exploring the data to uncover patterns and examine the trends in the features present in the dataset is important. This section will explore the dataset through the use univariate, bivariate and multi-variate plot.



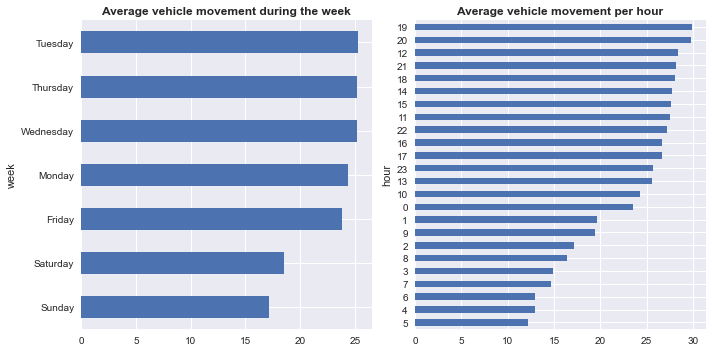
*Figure 4.1: Distribution of Vehicles*

The figure 4.1 shows the distribution of the vehicle numbers recorded by the sensor in the junction based on the observational time period. On the average, the distribution of vehicles was slightly above 20 but below 25. Also, the histogram curve shows that the distribution vehicles is heavily skewed to the right, this suggests randomness in the vehicle flow. This is somehow expected due to irregular movement of vehicles on any motorable road. More than 125 number of vehicles are seen on the road at a time which drags the tail of the histogram plot towards the right.

**

*Figure 4.2: Distribution of Vehicles per Junction based on timestamp*

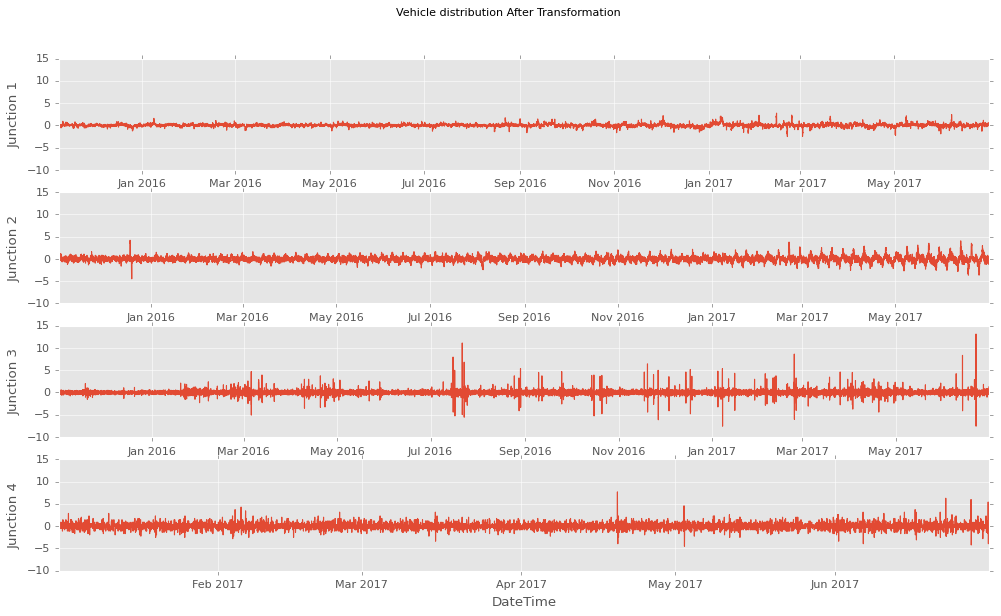
The figure 4.2 shows the number of vehicles on the different junctions based on timestamp. It can be observed that yearly, there is an upward trend for vehicle flow for all the junctions except for the fourth junction. It has been established earlier that the fourth junction has a sparse or limited data coverage which didn’t span over a year. There is a progressive influx of vehicle flow across the year especially for junction 1 and 2 relatively to other junctions. This may be due to summer break, nearness to business hubs and periodic activities at the junctions. Also, there seems to be a consistent flow of vehicle flow per month while the month of May, year 2017 recorded the highest number of traffic jam.



*Figure 4.3: Average movement of vehicles per week and hour*

The plots in figure 4.3 shows the average movement of vehicles during the week and hourly. Based on weekly traffic flow, Sundays enjoy a smoother traffic with lesser vehicles on the road. Steady traffic flow was recorded for Monday through Saturday, and Tuesday seems to be the busiest day with highest average number of recorded vehicles. On the average, the hourly vehicle distribution plot shows that the 19th hour of the day seems to be the peak for number of vehicles flow across all junctions. The 19th hour corresponds to the dusk period of the day where almost all workers and road users are probably returning from work or going for an evening outing. Average movement of vehicles was observed declining towards the early hours of the day.

### **4.3.1 Data Processing**

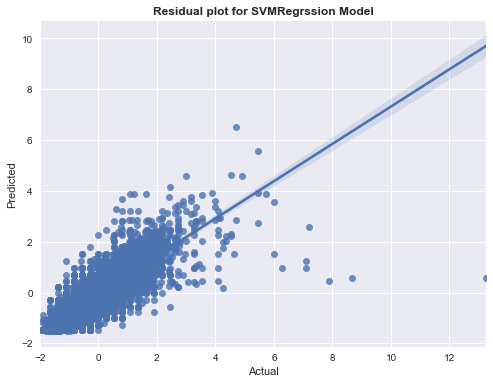
****

*Figure 4.4: Vehicles distribution per junction after transformation*

Data preprocessing is one of the important steps in Machine Learning, the quality of data and the useful information that be derived from can significantly affect the ability of any model to learn and predict well. Therefore, data preprocess is highly needed to be ensured before feeding data into any model. Data preprocessing cut across different layers ranging from handling null values, data standardization, variable encoding to detection of serial correlation. The traffic data used in this study does not have any issue of Null values but presence of variability or randomness in data distribution as earlier established. Due to presence of outliers in the distribution of vehicles per junctions for the yearly coverage, stationarity or transformation of the data becomes imperative before modeling. To standardize the data, a standard scaler function was used to transform the values such that the mean of the values becomes 0 and the standard deviations is 1. The time series plot in figure 4.4 shows that the vehicle distribution across the four junctions was stationary.

## **4.4 Modeling**

In this section, comparison among the three selected famous non-parametric model was presented. For the support vector machine (SVM) the liner kernel function was used to estimate the relationship between the dependent y and the independent variables . For the K-nearest neighbor, the Euclidean function was used to estimate the relationship between the dependent and independents variables. Furthermore, for the artificial neural network, the multi-layer recurrent neural network (RNN) was used to estimate the traffic data. Each model was trained, validated and tested on the traffic dataset. Different lag features were generated for the four junctions based on the past three continuous records of the traffic dataset to predict the traffic flow in the next time interval. The generated lag features were randomly split into 70% train, 15% validation and 15% testing set. By random shuffling of the dataset, training inertia caused by the regular vibration in the original dataset was avoided, and the average of the cross-validation method’s results was adopted to also avoid evaluation deviation.



*Figure 4.5: Residual plot for Support vector machine regression (SVMR) model*

Measuring the significance of the model, a residual plot as shown in figure 4.5 was used as a model diagnostic test. A linearly trend association was observed between the predicted and the actual values, this implies that the residual of the SVMR model is normally distributed; therefore, the model is fit for predictions. However, slight variation and very few outliers could be observed in the fitting of the actual versus the predicted plot.

**

*Figure 4.6: Residual plot for the KNN regression model*

Measuring the significance of the KNNR model, a residual plot as shown in figure 4.6 was used as a model diagnostic test. A linearly trend association was observed between the predicted and the actual values, this indicate that the residual of the SVMR model is normally distributed; therefore, the model is fit for predictions. However, slight variation and very few outliers could be observed in the fitting of the actual versus the predicted plot.



*Figure 4.7: Residual plot for the RNN model*

The artificial neural network model was also used to train the traffic dataset based on the generated lag features. The diagnostic plot as shown in figure 4.7 produce the linear relationship between the predicted traffic flow and the actual recorded traffic flow. Compared with the other models, the residual plot of the RNN model produced a more perfect linear relationship between the predicted and the actual value.

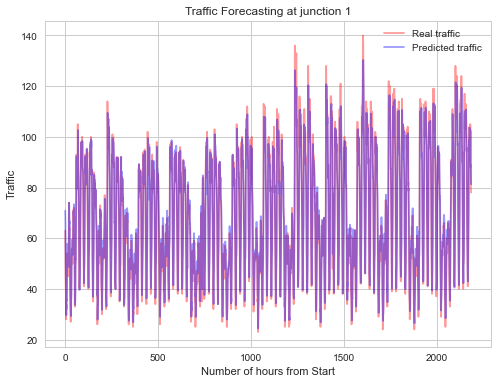
### **4.4.1 Model Performance Evaluation Metrics**

To evaluate the accuracy of the machine learning models, the performance metrics of each of the model was evaluated. As shown in table 4.1, all models performed very well with root mean square value (RMSE) below 1. The RMSE and MAE vale in RNN model is smaller than the other two model. This shows that the RNN model yielded a better predictive accuracy on the testing data. Also, the RMSE and MAE of the SVMR model was lower than the KNNR model. The SVMR model yielded a higher R2 than the other two model. In summary as observed from the table 4.1, analyzing the performance of each model on the traffic dataset. The RNN model have the best accuracy over all of the models. Since it can be established that the artificial neural network model outperformed the traditional base model (SVMR and KNNR); therefore, the RNN model will be adopted in this study for traffic flow prediction or forecast.

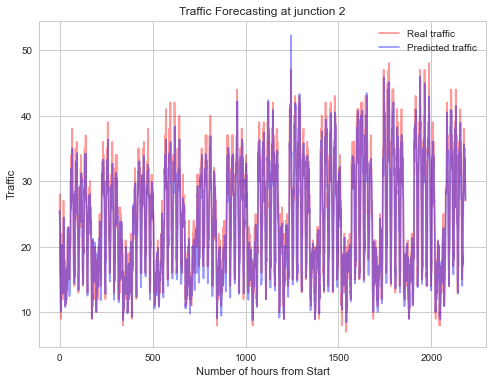
**Table 4.1: Evaluation metrics of the Machine learning models**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Models** | | |
| **Metrics** | **SVMR** | **KNNR** | **RNN** |
| Mean Square error | 0.32768 | 0.38426 | 0.14231 |
| Mean absolute error | 0.38661 | 0.43001 | 0.36949 |
| Root mean square | 0.57244 | 0.61988 | 0.37724 |
| R2 | 0.67233 | 0.61574 | 0.48611 |

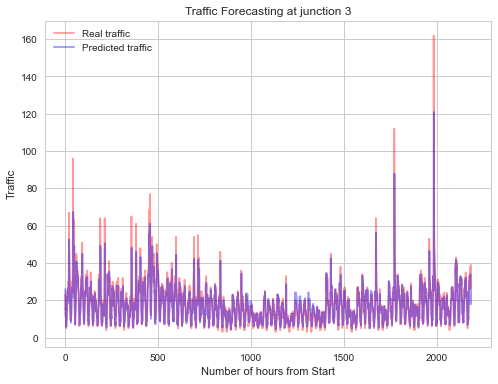
## **4.5 Traffic Flow Forecasting**



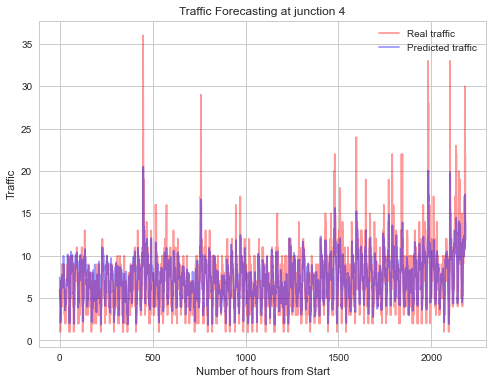
*Figure 4.8: Traffic prediction at junction 1*



*Figure 4.9: Traffic prediction at junction 2*



*Figure 4.9.1: Traffic prediction at junction 3*



*Figure 4.9.2: Traffic prediction at junction 4*

Using the obtained artificial neural network model for prediction of traffic flow for junction 1. It can be observed from figure 4.8 that the model predicts the traffic flow based on number of hours from start at junction 1 almost perfectly. The non-linear distribution of vehicles per hours on junction 1 was predicted with good level of accuracy by NN model.

For junction 2, as shown in figure 4.9, the pattern of traffic flow prediction at this junction per hour is similar to the event of junction 1. It can be observed that the flow of vehicle at this junction per hour is fluctuation and nonlinear in distribution. Adequate vehicle diversion and monitoring devices are needed in order to avert traffic congestion.

The prediction of traffic flow for junction 3 as shown in figure 4.9.1 by the RNN model was good with relatively high precision rate. Traffic congestion was observed to increase steadily across the period.

For junction 4, the performance of the RNN model in predicting traffic flow also was observed good with higher precision as shown in figure 4.9.2. The distribution of vehicles at this junction seems steadily and progressive compared to the other assessed junctions. The predicted instances match almost perfectly the real or actual traffic record for the particular time period.

Overall, the performance of the RNN model in traffic flow prediction was excellent with deeper ability to model a non-linear flow in vehicle distributions based on each observed junction.

# **CHAPTER FIVE**

# **SUMMARY AND CONCLUSION**

## **5.1 Summary**

In this study, firstly, a review and testing of some popular non-parametric traffic prediction models was done, as well as the use of a deep-learning model on traffic prediction task on the collected traffic dataset. The analysis structure adopted the convectional performance evaluation metric to compare the accuracy and the overall fitness for the base models and the deep learning model. From the analysis and modeling results, it was established that machine learning based models and the deep learning model structure yielded a good prediction accuracy, especially while the model has sufficient training data. However, relatively to the other non-parametric model (KNNR), the support vector machine regression (SVMR), has advantage on implementation cost and gave a lower root mean square error. Before modeling, a thorough data exploration was performed in order to determine the behavior and trend pattern among the variables contained in the dataset. Randomness due to presence of noise or outliers was detected, and stationarity was achieved through data scaling and pre-processing. This greatly improved the learning rate and accuracy of all models especially the traditional machine learning based models.

Secondly, a review of the predictive accuracy of the models through the use of diagnostic plot was performed. Based on the regression plot, the residual estimate of the ML-based model, though, produced a linear relationship between the actual and the predicted traffic flow but a slight deviation from normality was observed as the volume of prediction increased. However, this was not the case of the deep learning model which produced an almost perfectly linear relationship between the observed and the predicted traffic dataset.

Based on methodology, selection of suitable model for the traffic flow prediction is important, therefore the use of RMSE, MSE, MAE and R2 became necessary in order to judge the best performing model. Despite achieving varying values (almost all lower that 1) for each metrics across the models, the deep learning ML model was observed to be consistently lower in terms of MAE, MSE and RMSE score relatively to the based ML models. Comparing the R2 score, the two ML-based models produced a higher R2 score than the deep learning model. This may be due to over estimation of the R2 score by the ML-based model which was penalized by the deep learning model.

Lastly, based on the lower metrics achieved the deep learning model was selected as the best model, and was used for the traffic flow prediction task. The RNN model obtained was used to predict the traffic flow across the four junctions per hour. The RNN model predicted the flow of vehicle on the four junctions accurately.

## **5.2 Conclusion**

This study has established the role of machine learning models in Intelligence transport system. The use of a deep learning model gives a much higher efficiency and accuracy in prediction than the traditional machine learning based models. Also, it was established that machine learning models especially deep learning models have high robustness when facing a traffic prediction tasks. The reviewed efficiency of the three models suggested that the deep learning model will be the best model in predicting traffic flow in intelligence transport system.

## **5.3 Limitation of study**

Further study on traffic prediction for intelligence transport system using large transportation network prediction should consider combining some of the model used in this study with an intelligence transport system, such as vehicle control or congestion control system. Proper evaluation mechanism should also be adopted to see the model can improve the performance of the interface. Furthermore, this study did not consider the spatial relationship change of traffic flow at different times of the day and no inclusion of such feature.

# **References**

Pinlong Cai et al., 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C Emerging Technologies 62.*

Boukerche, A. & Wang, J., 2020. Machine Learning-based traffic prediction models for Intelligent Transportation Systems. *Computer Networks, 181,* p. 107530.

Brian L, S., Billy M, W. & R Keith, O., 2002. “Comparison of parametric and nonparametric models for traffic flow forecasting”.. *Transportation Research Part C: Emerging Technologies 10.4,* pp. 303-321.

Brian, L. S., Williams, B. M. & Keith, O. R., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies 10.4.*

Chavhan, S. & Venkataram, P., 2019. Prediction based traffic management in a metropolian area.. *Journal of Traffic and Transportation Engineering .*

Chen, C. et al., 2020. An edge traffic flow detection scheme based on deep learning in an intelligent transportation system.. *IEEE Transactions on Intelligent Transportation Systems, 22(3),* pp. 1840-1852.

Chen, X. & Chen, R., 2019. A review on traffic prediction methods for intelligent transportation system in smart cities. *12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI),* October.pp. 1-5.

Dawen Xia et al. , 2016. A map reduce-based nearest neighbor approach for big-data driven traffic flow prediction. *IEEE access 4,* pp. 2920-2934.

Dubey, A., Lakhani, M., Dave, S. & Patoliya, J., 2017. Internet of Things based adaptive traffic management system as a part of Intelligent Transportation System (ITS).. *In 2017 International Conference on Soft Computing and its Engineering Applications (icSoftComp),* December.

Falch, M., 2020. Intelligent Transportation System in Developing Countries.. *ICTE 4.*

Guerreiro, G. et al., 2016. September. An architecture for big data processing on intelligent transportation systems. An application scenario on highway traffic flows.. *IEEE 8th International Conference on Intelligent Systems ,* pp. 65-72.

Gültekin, Ç. B. M. S. & Borat., O., 2010. “A neural network based traffic flow prediction model. *Mathematical and Computational Applications 15.2,* pp. 269-278.

Guo, J., Wang, Z. & Chen., a. H., 2017. *“On-line Multi-step Prediction of Short Term Traffic Flow Based on GRU Neural Network”.* Bangkok, Thailand, s.n., pp. 1-11.

Guravaiah, K., Kavitha, A. & Velusamy, R. L., 2020. *Data Collection Protocols in Wireless Sensor Networks.* s.l.:IntechOpen.

H Chang et al, 2012. Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences. *IET intelligent transport systems 6.3,* pp. 292-305.

INRIX, 2019. *INRIX: Congestion Costs Each American Nearly 100 hours, $1,400 A Year.* [Online]   
Available at: https://inrix.com/press-releases/2019-traffic-scorecard-us/

Jabot., F., 2015. Why preferring parametric forecasting to nonparametric methods?. *Journal of theoretical biology 372,* pp. 205-210.

Jaimyoung, K., Benjamin, C. & Peter, B., 2000. “Day-to-day travel-time trends and travel-time prediction from loop-detector data”. *Transportation Research Record: Journal of the Transportation Research Board 1717,* pp. 120-129.

Jiahao, W. & Azzedine, B., 2020. The Scalability Analysis of Machine Learning Based Models in Road Traffic Flow Prediction. *2020 IEEE International Conference on Communications. IEEE.*

Jin, F. & Shiliang, S., 2008. Neural network multitask learning for traffic flow forecasting.. *IEEE International Joint Conference on Neural Networks.*

Jin, W. & Qixin, S., 2013. Short-term traffic speed forecasting hybrid model based on chaos–wavelet analysis-support vector machine theory. *Transportation Research Part C: Emerging Technologies 27,* pp. 219-232.

Kaouther, A., Azzedine, B. & Pazzi., a. R. W. N., 2011. sign and evaluation of context-aware and location-based service discovery protocols for vehicular networks”. *IEEE Transactions on Intelligent Transportation Systems 12.3 ,* p. pp. 717–735.

Khanna, A., Goyal, R., Verma, M. & Joshi, D., 2018. Intelligent Traffic Management System for Smart Cities. *International Conference on Futuristic Trends in Network and Communication Technologies,* February.pp. 152-164.

Kranti Kumar, M. P. a. V. K., 2013. Short term traffic flow prediction for a non urban highway using artificial neural network. *Procedia-Social and Behavioral Sciences 104,* pp. 755-764.

Leung, K. H., 2016. *Indonesia's summary transport assessment. (White paper)..* [Online]   
Available at: https://www.adb.org/sites/default/files/publication/217196/ino-paper-15-2016.pdf

Li Li et al., 2015. “Robust causal dependence mining in big data network and its application to traffic flow predictions”.. *In: Transportation Research Part C: Emerging Technologies.*

M. Dorigo, V. M. a. A. C., 1996. “Ant System: Optimization by a Colony of Cooperating Agents. *Trans. Sys. Man Cyber. Part B 26.1,* pp. 29-41.

Manoel Castro-Neto et al., 2009. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions”.. *In: Expert systems with applications.*

Marco, L., Matteo, B. & Frasconi., P., 2013. *IEEE Transactions on Intelligent Transportation Systems 14.2,* pp. 871-882.

Maya, O., Hideaki, K. & Hiroyuki, T., 2017. Online Traffic Flow Prediction Using Convolved Bilinear Poisson Regression. *Mobile Data Management (MDM), 2017 18th IEEE International Conference on. IEEE,* pp. 134-143.

Meena, G. S. D. & Mahrishi, M., 2020. Traffic prediction for intelligent transportation system using machine learning. *3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE),* February. .pp. 145-148.

Mei Duo et al. , 2017. A short-term traffic flow prediction model based on EMD and GPSO-SVM. *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE,* pp. 2554-2558.

Narendran, K., Monishraj, R. & Srivinas., D. S., 2022. Traffic Prediction for Intelligent Transportation System Using Machine Learning. *International Journal of Research Publication and Reviews,* March, Volume Vol 3, pp. pp 1790-1794.

Navarro-Espinoza, A. et al., 2022. Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithm. *technologies,* 10 January.

Okawa, M., Kim, H. & Toda., a. H., 2017. “Online Traffic Flow Prediction Using Convolved Bilinear Poisson Regression”. *Mobile Data Management (MDM), 2017 18th IEEE International Conference on. IEEE,* pp. 134-143.

Pinlong Cai et al., 2016. “A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting”. *Transportation Research Part C: Emerging Technologies,* pp. 21-34.

Qureshi, K. N. & Abdullah, A. H., 2013. survey on intelligent transportation systems.. *Middle-East Journal of Scientific Research,* pp. 629-642.

Rice, J. & Zwet., a. E. V., 2004. A simple and effective method for predicting travel times on freeways”. *IEEE Transactions on Intelligent Transportation Systems 5.3,* pp. 20-207.

Schölkopf, B. et al., 2002. Support Vector Machines and Kernel Algorithms..

Shai, S.-S. & Shai, B.-D., 2014. *Understanding Machine Learning..* Cambridge: Cambridge University Press,.

Simon Oh et al., 2015. Short-term travel-time prediction on highway: a review of the data-driven approach. *: Transport Reviews 35.1.*

Singh, K., 2015. Creating your own qualitative research approach: Selecting, integrating and operationalizing philosophy methodology and methods. *Vision, 19(2),* pp. 132-146.

smartsheet, 2022. *Neural-network-application.* [Online]   
Available at: https://www.smartsheet.com/neural-network-applications

Trevor, H., Robert, T. & Friedman, a. J., 2009. The elements of statistical learning: data mining, inference, and prediction.. *Springer Science & Business Media.*

Vanajakshi, L. R. L. R., 2004. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. *Proc. IEEE IVS..*

Vapnik., V., 2013. *The nature of statistical learning theory.,* s.l.: Springer science & Springer science &.

W. Huang et al, 2014. “Deep Architecture for Traffic Flow Prediction: Deep Belief Net works With Multitask Learning”. *works With Multitask Learning”,* pp. 2191-2201.

Wang, J., 2021. Vehicular Traffic Flow Prediction Model Using Machine Learning-Based Model.

Wei-Chiang Hong et al., 2011. Forecasting urban traffic flow by SVR with continuous ACO”.. *Applied Mathematical Modelling 35.3,* pp. 1282-1291.

Xiang, F., Chung-Cheng, L. & Ke, L., 2011. “A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction”.. *Transportation Research Part C: Emerging Technologies 19.6,* pp. 1306-1318.

Xianglong, L., Danyang, L. & Shengrui, Z., 2019. Traffic flow prediction during the holidays based on DFT and SVR. *Journal of Sensors.*

Xiaolei Ma et al., 2015. Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS one 10.3.*

Zantalis, F., Koulouras, G., Karabetsos, S. & Kandris, D., 2019.. A review of machine learning and IoT in smart transportation.. *Future Internet, 11(4),,* p. p.94..

Zhang Mingheng et al., 2013. Accurate multisteps traffic flow prediction based on SVM. *Mathematical Problems in Engineering.*

Zuduo, Z. & Dongcai, S., 2014. Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transportation Research Part C: Emerging Technologies 43,* pp. 143-157.