Machine Learning and Deep Learning Approach in Traffic Flow Control and Prediction for Traffic Management System

Preeti Rekha Sahu

ABSTRACT

Congested roads is a fundamental urban areas' issue which is primarily contributed from unexpected vehicular population growth and inefficient traffic operation and control in some mechanisms. The field of intelligent transportation systems (ITS) has developed quickly inrecent years. A cost-effective method for managing and planning smart public transportation. ITS enhances traffic safety, and mobility reduces the externalities that arise through all the transportation-related activities. (ITS) applications require the minimum human intervention and are utilized for advance route planning and traffic control systems. ITS applications are efficient in large scale traffic data collection both in time and space and several studies use large scale traffic data for developing efficient traffic operation programs for a city. Large scale data collection programs have several potential applications in solving transportation related problems by developing robust traffic flow prediction models. In recent times, researchers apply novel tools such as Machine Learning (ML) and Deep Learning (DL) to predict real-time traffic. Real-time traffic prediction models are helpful for improved traffic control and efficient traffic management system. Statistical models, ML and DL models are used for traffic signal design, que length analysis, and delay minimization for traffic stream in an urban network. Inessence, these models help in minimizing travel time for users and thus reduces travel cost. This paper's goal is to present a thorough grasp of the use of ML and DL approaches to improve traffic flow prediction models with recommendations for ITS application in smart cities. The findings from this research may be applied by smart city managers for developing efficienttraffic management programs in the cities in India and elsewhere.

Keywords:- *ITS, Machine Learning, Deep Learning, Traffic flow Control and Prediction.*

ISSN No:-2456-2165

TABLE OF CONTENTS

ISSN No:-2456-2165

CHAPTER ONE
INTRODUCTION

A. Background

The idea of a "smart city" is emerging in response to the rapid development of information andcommunication technology (ICT) as a means of improving urban administration, environmental sustainability, and citizen quality of life. Cities must meet the demand from these residents by enhancing the services that improve the standard of living. While cities havechanged in terms of potential, these changes have also revealed numerous difficulties that mayaffect citizens' daily lives. The driving force behind this evolution has always been technology,which has profoundly altered our way of life over time. The connected worlds of physical objects, people, and technology are changing how we work, travel, socialise, and interact withour environment. The Intelligent Transport System (ITS) is a system that aims to enhance mobility and safety in transportation while also promoting citizen development by minimisingthe detrimental effects of traffic flow in cities. In the twentieth century, researchers in the US put forth the basic ITS concept. However, ITSs are currently gaining a lot of attention from academics and business since they can improve the safety, sustainability, and e-client of the transportation sector in addition to the conditions for vehicle traffic. Inconveniences brought on by clogged city streets and the impact of climate change on traffic can also be lessened. Modern society's major advancements in the automotive industry are being driven by communication technology. Mobile communications have revolutionised daily life for citizensover the past two decades by enabling ubiquitous information interchange at all times and locations. As businesses, academic institutions, and governments from all over the world investsignificant time and money in creating safer vehicles and infrastructure for the transportation of people and goods by road, it is anticipated that the use of such communication systems in vehicles will soon become a reality. Numerous national and international efforts devoted to vehicle networks can be used to verify these investments. Therefore, intelligent transport systems are a component of vehicular networks that can be used to forecast traffic flow in the context of smart cities. Thus, it is safe to argue that intelligent transportation The quality of lifein urban societies as well as the growth of the city are significantly impacted by traffic issues,therefore predicting traffic congestion is crucial for both citizens and governments and also However, it might be very difficult to comprehend and model the traffic flow conditions. Traffic congestion is a growing source of concern for people because it has negatively impacted both urban development and their quality of life. Cities all across the world have installed embedded sensors, such as in road networks, inductive-loop detectors and video imageprocessors are used to monitor the flow of traffic. The growing accessibility of data and serviceshas made it possible to forecast traffic conditions, such as projecting transport volume and speed over the entire city. Urbanization is changing here, and it is anticipated that The qualityof life in urban societies as well as the growth of the city are significantly impacted by traffic issues, therefore predicting traffic congestion is crucial for both citizens and governments. The main element of ITS that forecasts and prevents traffic congestion, efficiently controls and manages traffic, and plans the optimal route for travel is traffic flow prediction. Based on machine learning. Traffic flow predictions are made using techniques like Random forest algorithm, Support vector schemes, and convolution Neural Network (CNN) schemes. Deep learning (DL) models have the advantage simplifying when compared to other ML approachessince it performs more accurately and simplifies the data pre-processing procedure. As a result,the field of traffic flow prediction has recently paid a lot of attention to DL approaches. Moreover, one of the key components of a smart city, traffic flow prediction helps users to maximise their use of resources like time, fuel, and electricity.

B. Problem Statement

India is a developing country with second largest population. Transportation is an important factor in the economic development of any state. So, a large number of vehicles play importantrole in transportation causing congestion in traffic. Many researchers work in this field to avoidcongestion problems, specifically, in urban areas. Traffic control shall reduce, not only congestion but also improve safety. It also solves problems of fuel consumption, air pollution,and overall environmental issues due to the large use of vehicles. So, there is a need to solvethis problem of congestion management and the quality of traffic flow. The application of anintelligent transport system will help in mitigating this problem which can be used to predictthe congestion in traffic flow. Traffic flow prediction and development of better trafficmanagement system are important part of smart city for its sustainability. Therefore,developing. The safety of the traffic environment and a method to avoid traffic congestion areguaranteed by a scientific prediction model to reduce congestion. Congestion mitigation leadsto reduction of vehicle accidents and improve air quality. Many researchers have attempted todevelop traffic prediction models, but, there accuracy levels are questionable. Developedmodels are unable effectively deal with various features of time series of data leading to highererrors in predictions of the traffic flow. Therefore, extra investigation needs to be made to havea near accurate modelling system to solve the congestion problems. Better traffic flowpredictions can be made by analysing various variables. Finally, transportation quality can beraised with the successful deployment of smart intelligent transport system.

C. Motivation

City traffic can have a negative impact on society's health and environmental conditions. We all know that one of the most crucial ways urban civilization employs to get around on a dailybasis is through transportation. Smart city technology can enhance society's standard of living.Travel services that are efficient. One of the ways to offer this kind of service is through intelligent transport systems. For intelligent transportation systems to operate more effectivelyin the context of smart cities, prediction of traffic flow is crucial. To provide improved traffic flow forecast services, better prediction models have not yet been developed by many researches. The effectiveness of trajectory prediction is influenced by a number of things. As aresult, researching and

ISSN No:-2456-2165

analysing those elements aids in creating more accurate forecasting methods. The aim of this project is to investigate several modern traffic flow forecasting systems in order to identify methods for creating precise traffic flow predictions. As long as the method of accurate prediction is correctly examined and increase the quality of traffic flow prediction services in intelligent transportation systems. The development of reliable traffic flow prediction methods is crucial for both the government and users of urban roads. As a result, daily commuting experiences in urban societies will be better.

D. Objectives and Research Issues

The following research questions are examined in this study, all of which are based on the problem statement:

- *RQ1: What potential difficulties can prevent the effective application of forecasting system road traffic parameter?*
- *RQ2: What potential do recurrent neural network-based methods for predicting traffic parameters have?*
- *RQ3: What are the most recent machine learning architectures of traffic prediction for forecasting traffic flow, and what impacts does the suggested technique have on the performance of the selected model?*
- *RQ4: What advantages do deep machine learning techniques have over traditional or shallow machine learning methods when taking traffic flow data into account?*

E. Research Method

For the purpose of addressing the first three research questions, a background investigation anda current literature evaluation were conducted. Iteratively testing the proposed approach and the flow prediction goal in the null hypothesis function was done for several models. The mostadvanced deep ML models were then subjected to studies comparing them quantitatively to traditional ML and statistical forecasting methodologies. Then, a hypothesis based on results was created. The hypothesis is then put to the test again with predictions from many time- dependent models. Finally, in end conclusions are then made.

F. Contributions

This thesis makes a contribution by compiling the most sophisticated machine learning methodologies, from shallow to cutting-edge deep learning approaches, to make forecasts for traffic flows while optimising for the suggested objective function for the fundamental junctionlevel highway traffic flow. By using a 14 topological split of the highway network, it is possibleto report on the bi-directional flow function of individual roadways while taking into account net inflows and outflows. The suggested method is modular and is applicable to behavioural learning for network-wide traffic flow. Additionally, the method can assist in taking the bottlenecks into account while analysing congestion.

G. What is Machine Learning

In recent years, artificial intelligence (AI) has emerged as a popular buzzword. The more general phrase used to describe the systems and control methods that enable machines to do tasks that are thought to require intelligence is artificial intelligence. Since giving machines access to pertinent data will enable them to learn on their own, machine learning (ML) is the application of AI to machines. There are many distinct types of machine learning algorithms, but they all act on data, which is something they all have in common. Therefore, the two fundamental factors in any ML performance are data relevance and here therefore, the two fundamental factors in any ML performance are data relevance and accessibility. Another namefor ML is the subset of AI. It's acceptable to believe considering ML to be the most advanced technology available now is not incorrect.

➢ *The Workings of Machine Learning*

The group of methods known as "machine learning" are used to identify and carry out certain tasks from a given amount of data. The classifier's learning process, which later classifies andmakes predictions, is mostly driven by the data from a certain field that is available. The algorithm then develops its rules and functions either independently (unsupervised) or using features that users have manually chosen (supervised). The training phase is the process of teaching and perfecting an algorithm. The trained algorithm instance is then put to the test against the validation and test datasets, respectively, during the testing and training stages. Theperformances and accuracy are then categorised in a simulation environment using common benchmarking algorithms. This strategy is somewhat more practical.

➢ *Machine learning advances*

Data analysis is automated using machine learning, a type of analytical solution. Algorithms used in machine learning iteratively search for patterns in data and discover some useful hidden insights that may eventually enable a computer to function without explicit programming. The machine learning of today is very different from previous iterations of machine learning. Machine learning algorithms are being developed more quickly than ever these days since it isnow possible to use sophisticated mathematical operations on large amounts of data.

➢ *Self-Driving Cars*

The self-driving automobiles developed by Google and Tesla are two examples of extensively used applications that are currently getting a lot of attention in discussions. Due to the power of machine learning algorithms, they are all technically feasible.

➢ *Recommendation Systems*

Systems that utilise machine learning to suggest particular products to clients based on their purchasing power, preferences, and previous order history. These recommendation algorithmsare built on the basis of past purchases. The dynamic pricing model, which is the simplest pricemodel, is explained in. In accordance with the distribution of the dynamic pricing range groupto each individual client, it forecasts the product sale purchase based on the historical sale data,and 15 it also predicts the possibility that the products will be purchased by a certain customer.Customers' purchasing power or the dynamic price range that has previously been assigned viathe k means-clustering technique determine the selection of goods that are made available to them.

➢ *Study of the Sentiment on Social Media*

A data mining technique for emotive analysis of Twitter Tweets and language rules aggregationapproach, is a learning system that could forecast the kinds of attitudes people have about others.

➢ *Prevention Against Online Credit Card Fraud*

Machine learning techniques are currently being used by online credit transaction merchandisers to identify and foretell spam occurrences based on prior incidents. This enhances the degree of consumer happiness and the service provided by these credit merchandisers.

➢ *Filtering Email on Span*

For the classification understanding of spam filtering from incoming emails, the traditional machine learning example is still used today and includes email spam filtering. This facilitatesreading emails one at a time and, to some extent, helps protect PCs from hackers. Complex anti-viruses may employ a combination of classic and hybrid algorithms, whilst straightforward filtering techniques use a fundamental decision tree-based approach**.**

H. Commonly Used Machine Learning Algorithms

With a brief description of what they are used for, some of the most beneficial ML algorithmsavailable today are listed.

➢ *Artificial Neural Networks*

A neural network (NN) is a type of common classifier that mimics the decision-making abilitiesof the simplest human brain. The perceptron combinations that make up artificial neural networks (ANNs) are coupled in the form of several layers. Neural networks simulate the intricate input-output relationships that are otherwise challenging to understand and imitate in the actual world. Each individual perceptron's weights act as a stabilising chain for controllingthe data flow in the form of weight values and activation functions. Optical One of the severaluses for the ANN is the identification of electrical circuits. In order to comprehend the symbolsin an electrical circuit diagram, various ANN layers and activation functions are used.

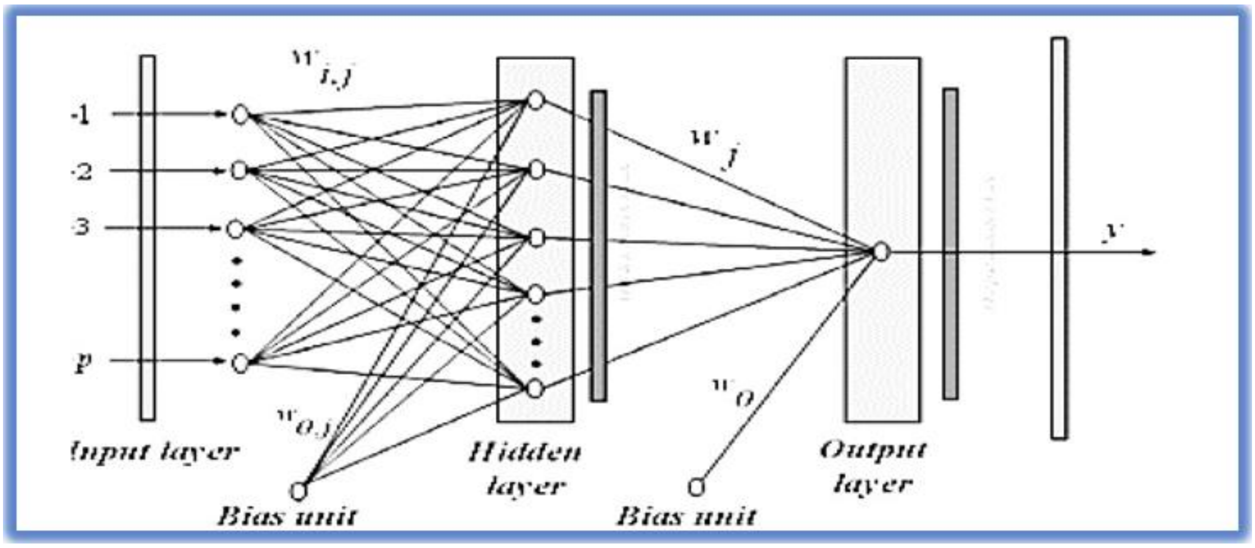


Fig 1 MLP Network (Multi-Layer Perceptron)

➢ *Decision Trees*

Data is categorised using decision trees in the classes that can be deduced from the data itself. Dynamic A more advanced solution that creates new classes or branches based on the data given in real-time is a decision tree. To understand decision trees in practise, a more appropriate way could be to use a conventional transport analysis solution. Providers of transportation services must devise solutions for fundamental issues like real-time decision assistance, handling insufficient data and utilising human judgement. Machine learning techniques are used to interpret the rules from the data and deliver offline planning solutions in order to improve the quality of transportation planning. The operators' time consumption in real-time planning systems is a major problem.

➢ *Other ML Techniques*

Random forests are among the most popular and well-known ML classification and predictionapproaches (RF). For in-depth data analysis, research, and understanding, variable selection may be utilised. Since it lessens the correlation impact by ranking the variables in a particular way according to their importance, RF has demonstrated greater performance when using various variable selection approaches. Data classification is aided by associations and sequencediscovery when used in conjunction with other methods. Data records must be converted into ontology-based event graphs in order to be categorised against previously collected data. These graphs serve as visual representations of time-related event sequences. Data conflicts between aggregated records represented as events could be resolved between aggregated recordsrepresented as events could be resolved with the use some methods in terms of events. In a Some of the developed machine learning (ML) algorithms that are frequently used alone or in combination with other algorithms on a variety of datasets to find the optimal solution. In figure 1.2, supervised procedures are broken down into standardised and fundamental ML techniques that have been used in past works of literature. Breakdown of supervised machine learning prediction methods is shown in Figure1.2

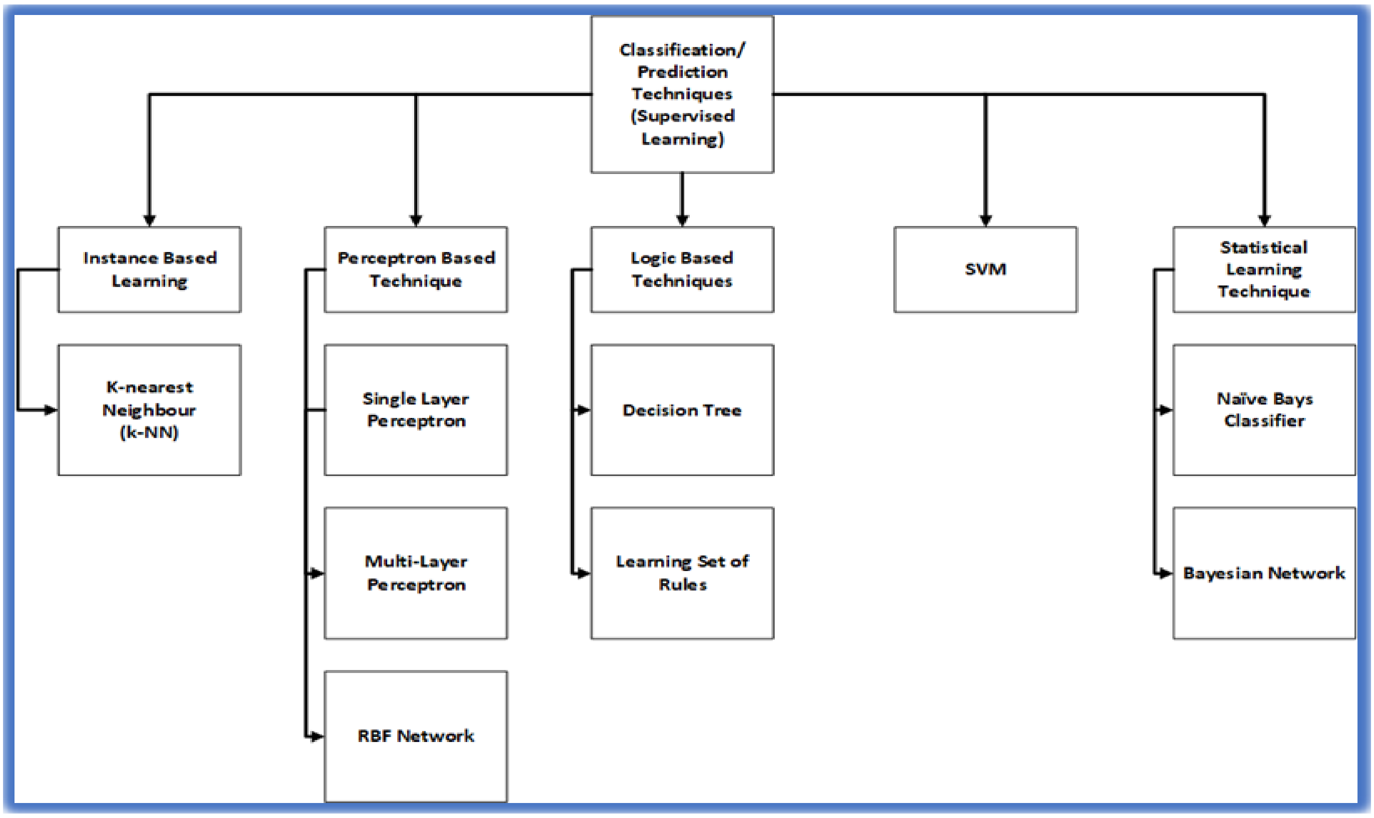


Fig 2 Breakdown of Supervised Machine Learning Prediction Methods.

Modern applied sciences now cover practically all aspects of artificial intelligence and machinelearning. A massive collection of data collected under varied circumstances—what we refer toas Big Data—allows AI and ML to generalise the behaviour of the process as a whole. Over time, numerous cutting-edge machine learning algorithms have been created and are now used as benchmarks. Performance and accuracy of all newly developed algorithms are compared tothose of these conventional methods. ML can be built or tweaked for various factors to generalise data behaviour, whether it is supervised or not, according to research presented in following chapters.

I. *What is Smart Transportation?*

Every day, data from transportation operations is produced in large quantities. Important data from the commercial transportation providers is generated by sophisticated electronic ticket machines (ETM) and the servers that run them. Due to the size of the data, a deep and manualinspection is insufficient, assuming it is even attempted. A thorough understanding of the raw data is desperately needed.

➢ *From the Perspective of Commercial Transport Operators*

The raw dataset contains a variety of indirect characteristics that are solely discovered by modifying the dataset's known visible features, such as bus departure times at each stop along the route, passenger dwell times (DT), the types and quantities of tickets used at each stop, theuse of smart tickets rather than Concessionary passes, cash transactions, and other characteristics. To achieve the indirect characteristics, we're after, which might not always be obvious physically but are more important for machine learning behaviour, we constantly needto modify the dataset's raw features in some way. An example of a wait time division that considers.

ISSN No:-2456-2165

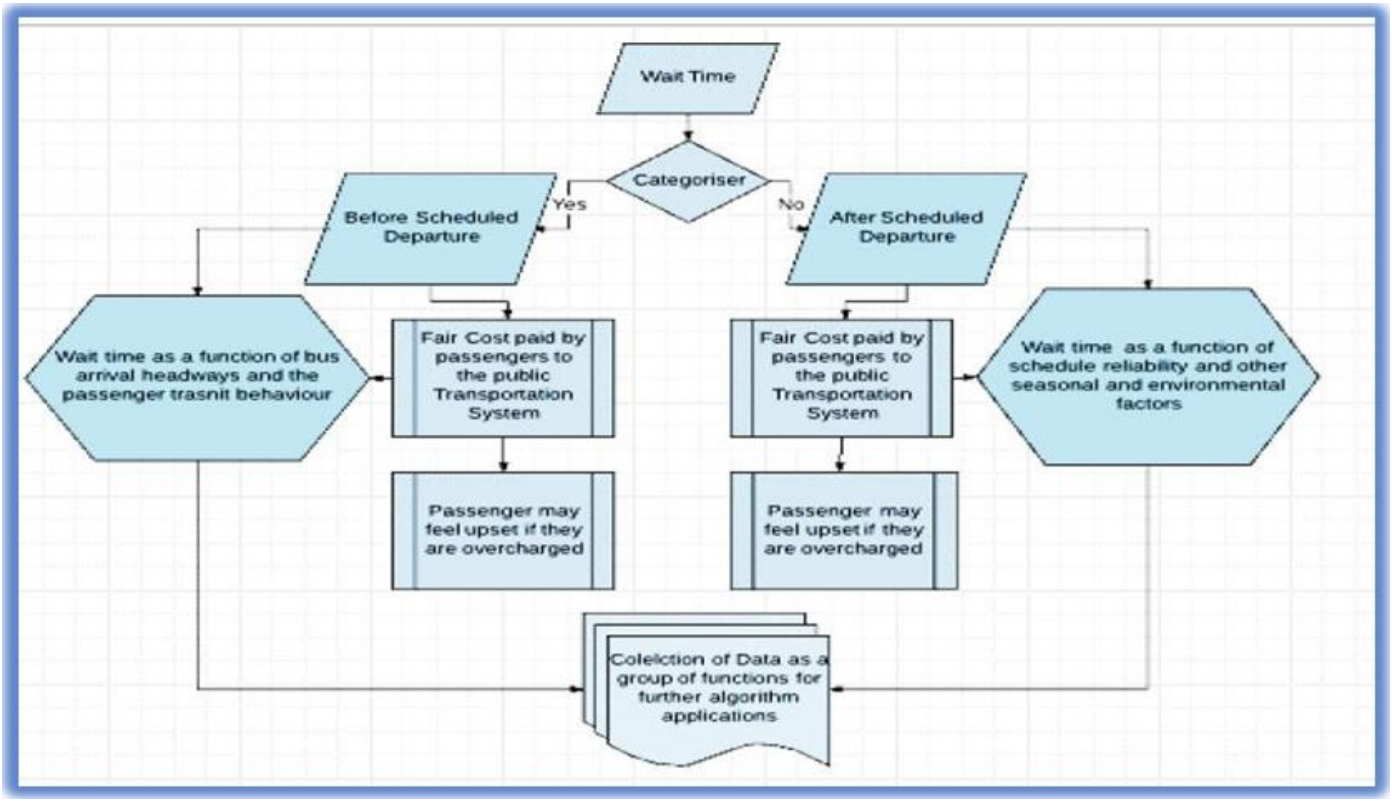


Fig 3 As Observed by the Transit Operators, Typical Wait Categories.

A generalised example of a sample smart transportation model that uses ML algorithms tosimulate, say, bus deviation behaviour is presented as follows:

Bus deviation prediction proposed general model base solution=Artificial Intelligence(ML)+Statistical models

Equation is the generic name for the model that mathematically depicts what the real-world ML architecture for bus deviation prediction scenarios looks like. Artificial intelligence (AI) methods, Internet of Things devices, and machine learning (ML) methods like Random forest,support vector machine, etc. are all included in equation 1.2. The definitions of these concepts would become much more evident as we developed the algorithm and reached the culmination the execution of the thesis. This academic project 's objective is to modify an AI algorithm tailored exclusively for the creation of transportation problem solutions that could make use of already-existing algorithms for improved result. Based on traffic and real-time data, a multi- modular, schedule-based, enhanced, and smoother mode of transportation.

➢ *Flow Limitation Caused by Congestion*

An intelligent transportation system is made up of numerous intelligent processes that each have a modular approach but coexist peacefully. For efficient mobility and top-notch traffic management and safety systems, understanding vehicular traffic congestion is essential. The consequent traffic congestion has the opposite effect on network traffic flows. According to a more empirical perspective, a sudden breakdown is what causes the traffic jams to occur. Whatwas formerly a motorway traffic road sees a steep drop in vehicle speed and an increase in car density. The extremely complicated spatiotemporal behaviour of the traffic networks is identified by later study. A traffic simulation is required to interpret the empirical aspects of the traffic breakdown.

J. Thesis Structure

Concluding the thesis with the conclusions and prospective future work, the scenario is built from the tiny problem definition to the potential questions and offered mythos, solutions through to the most recent machine learning model designs. are the main objectives of the various chapters with sections and subsections? The thesis remaining chapters are structured as follows: The detailed subject review for the traditional statistics, machine learning, and deeplearning methodologies for forecasting traffic road parameters is presented in Chapter 2. Modern frameworks for predicting traffic flow are presented in Chapter 3. The chosen and usedarchitectures' theoretical details are n Chapter 3. The chosen and used architectures' theoreticaldetails are presented in the next chapter. Chapter 5 provides research technique. While chapter 6 discusses the trials, findings, and quantitative evaluation of the suggested frameworks. The thesis is finally concluded in last section which also offers the future directions that will be based on this thesis and the final word on how well the prediction model performed.

ISSN No:-2456-2165

CHAPTER TWO
OVERVIEW OF TRAFFIC FLOW FORECAST APPROACHES, FROM CONVENTIONAL TO MODERN WAYS

A. Introduction

A general overview of the traffic forecast problem was provided. The pertinent literature is evaluated in more depth. This chapter is structured to include a brief idea, discussion of studieson the prediction of traffic parameters, a detailed. An analysis of each manuscript in comparison is provided, contrasting the methods used, the features chosen for decision-making,the experimental setup adopted using algorithmic models based on some statistical or data- driven machine learning (ML), sample algorithm test simulations, and the reported results in the manuscripts. This chapter also examines other closely related general engineering applications for data prediction that are relevant to traffic.

B. Aims

The background research on traffic flow and traffic journey time inference models and methodsis presented in Chapter 2. This chapter's final goal is to synthesise the many statistical and machine learning techniques that were discussed in the literature review. The optimal algorithms for the development of our model, in response to the observed deficiencies, and they are further explained in the following chapter.

C. Traffic Flow Analyses and Predictions Based on Literature Studies: ABrief History

According to a review of current literature, numerous writers have made significant contributions to the fields of traffic incident analysis, prediction, and their connections to traffic jams. It provides a straightforward visualisation-based method for displaying traffic collisionsfrom historical data as a dynamic radial circles on a map overlay. Traffic Origins are coloured circles that are drawn on the map and reflect various road circumstances, such as high traffic, breakdowns, and congestions. The visual indicators of the incident's location, of excessive traffic flow, and of breakdowns are the traffic origins, while the radius of those origins indicatesthe general area in which the traffic would be adversely affected. Once According to a survey of the literature, forecasting traffic flows can be roughly divided into two categories: parametric approaches and approaches based on statistical techniques for time series forecasting. With these procedures, understanding of data distributions is typically assumed. These traffic process-based prediction approaches primarily use simulations of traffic systems, road activities, and driver behaviour factors. The parallels between fluid dynamics and vehicle traffic flow serve as the foundation for the macroscopic traffic prediction models. The complexity of parameter estimations and the difficulty in creating a simulation test environment that closely resembles real life are the drawbacks of employing many macroscopicprediction techniques, on the other hand. Moreover, the accuracy of the anticipated traffic parameters is important impact on the forecasts. The creation of the ideal traffic flow predictionmodel can benefit from both the statistical ML and macroscopic techniques. However, the investigation of statically driven by data to sophisticated ML approaches for traffic-related predictions is the exclusive emphasis of this research. The primary distinction between ML andtraditional models based on analytical techniques are what ML is thought of as a "black box" that discovers connections between inputs and outputs to forecast traffic variables. ML modelscan be challenging to increase learning. ML models may adapt to the evolving societal norms seen in the data with continued training. Section 4 has a thorough discussion of the chosen models. The literature states similar traffic information must be utilised for both the models for statistical and computational learning in order for comparisons to be meaningful. It is challenging to locate a comparison of models using the similarities data in various comparativesettings across the literature.

D. In Light of the Literature Review, a Study of the Variables that Affect Traffic Prediction Models.

A number of variables affect how a model predicts traffic flow. Other factors besides the hyper parameters of the models include the environment that the supplied traffic parameters are usedare handled, the input data sample resolution, the prediction stages, the connections between the different traffic parameters employed, and the spatial-based temporal relationships concealed in the traffic variable data. The time series data's added seasonality and trend may also have an effect on the forecast's performance. Each of these essential components is coveredin the following subsections: algorithms or models.

- *Setting for the Use of Road Traffic Forecasts*

According to the literature review, two primary, different types of roadways that make up the environment of putting prediction models into practise are predicted for traffic parameters. There are two types of roadways: the urban linking roads and the highways, motorways, and motorways. The main distinction between the two is that because of uncontrolled connectionsand variable-sized crossings, Traffic dynamics on urban or linking roads are more challenging to comprehend. Highway prediction models are mostly employed for ITS applications since roads represent the foundation of the main long-traveling road infrastructure & this contains a few examples of road projections produced in relation to roads and highways. The arterial and linking road forecast models used are unreliable and Finding the best performing prediction model for the highway network roads is one of the objectives of this thesis.

- *Variables to Enter for Traffic Forecast*

The variables choice may be important and challenging for forecasting, but it is strongly related to the effectiveness and efficiency of traffic flow forecasting models. Sometimes, indirect techniques for extracting information, such as shared knowledge based on entropy theory, maybe used instead of only the raw feature values. Traffic flow volume, journey time, and speed data are some of the variable elements that are frequently taken into account by forecasting models. This form of changeable data is captured by on-site sensors that utilise some detectors or some sensors. Based on this piecewise switched linear traffic flow, together with traffic patterns, traffic density, and speed variables, are used to model complexity. The model is defined as a collection of partial derivatives of the important variables.

- *Traffic Flow Prediction: Seasonal Impacts and Spatial-TemporalPatterns*

The link between time and space has been extensively researched in respect to traffic flow andgeneral traffic predictions. In their research, numerous scholars have made enough mention of.Using road traffic temporal data in relation to spatial attributes has always been the aim. As thetrendiness of the series data is examined, traffic time series display seasonal and periodic behaviour. Regarding the spatiotemporal aspects, free flow lanes and motorway highways arestrongly related**.**

- *Different Road Conditions in Forecast of Traffic Flow*

Many factors can impact traffic flow. The simplest of these are minor accidents that happen onthe side of the road. Several studies have attempted to categorise various road situations in theirown ways. In general, there are two primary categories of traffic situations: regular traffic conditions and abnormal traffic conditions. Peak hours and post-accident conditions are two extremely high traffic conditions that are taken into account in for the proposed deep LSTM model. The suggested neural network (ANN)-based model is also focused on heterogeneous traffic circumstances. The use of conditioning-based traffic data for the suggested prediction models has also been the focus of other researchers.

E. Several Methods for Modelling in Light of the Literature Review, Traffic Flow and Congestion Behaviour, as well as the Limitations that Go Along withit

Prediction models in the literature review can be broadly divided into the following groups:

- *Methods Based on Simulations that are Parametric, Naive, and Macroscopic*

Parametric model approaches are the usual name for conventional methods that employ statistical techniques for time series forecasting. In parametric techniques, pre-existing data distribution information is presupposed. Most of the time, these model approaches deliver accurate short-term forecasts. Often known as naive techniques since they offer straightforwardtraffic estimates based on weighted averages of average means calculated in the most basic manner utilising data from the preceding interval, etc. Analytical methods typically perform poorly Long series data structures and longer time horizons are not suited for parametric approaches because they require a predefined set of parameters that are defined as a part of their mathematical and statistical equations.

Table 1 Methods Based on Simulations that are Parametric, Naive, and Macroscopic

Models	**Limitations**
Autoregressive Integrated Moving Average (ARIMA) Model	Might be used to make predictions for longerperiods of time, although the accuracy of thepredictions decreases as the prediction horizon gets longer.
Seasonal Autoregressive Integrated Moving Average (SARIMA)	Similar to ARIMA but takes into account thetime series' seasonality. The non-stationary seasonal series performs well since the stationary seasonal component is challengingto handle.
Auto Regressive Moving Average (ARMA) Model	Similar to the ARIMA model, it is better forshort-term projections. better suited for timeseries that are stationary. As a result, the integrated element doesn't have a big impacton the final estimates. Including the seasonalelement into performance can improve it.
Methods for Complicated Time Series Analysis and Filtering, Basic Smoothing, andExponential Smoothing	Biased significance cannot effectivelymanage real-time trends due to its bias towards the most recent observations. inadequate performance because of an uneven class or the seasonal elements
Geographical Weighted Regression, weighted moving average (WMA), and the weighted average approach (GWR)	Due to the complexity of the custom weightiness choice in traffic flow data and the trends filtering, significance is placed onthe observations with high weights.
Mean speed based on two-dimensional linear interpolation, Spatial-Temporal Hidden Markov models(STHMM),	With a macro method, estimated inference for the estimated variable is more accurate and benefits from higher data observation resolution.
Piecewise Method, Bivariate Move lets	For this strategy to be effective, a pattern library based on data intelligence must be created, which is challenging in a multi-trendtraffic environment.
Gaussian Model for Flow Data Imputation	It is necessary to infer additional parametersfrom the data, such as the mean and covariance among the data, in order to account for inaccurate estimations for incomplete data. Estimating prior data knowledge might be challenging.
Data fitting for linear, polynomial, power, and exponential curves using tensor-based multi-dimensional modelling	An objective function that takes into accountthe trade-offs between the relevant variablesis necessary for the modelling technique.

➢ *Techniques for Machine Learning that is Data-Driven and Non-Parametric*

This category includes models without a stable framework and without a set amount of constantvariables. neural network (NN), gated recurrent units (GRU), and recurrent neural network areexamples of deep learning long short-term memory (LSTM) (RNN). The majority of non- parametric methods also include data-driven models. To deliver the forecasts, they make use of the empirically based algorithms. Since they estimate the model parameters, they are free tomake any assumptions depending on the data generation and uncertainty; a well-known strategy is the use of neural networks. Prior to a few years ago, prediction of traffic parameter methods based on machine learning (ML) strategies were used. Non-parametric approaches areanother name for these data-driven methods. They have been applied without a predetermined structure or set amount of fixed parameters. K-nearest neighbours (KNN) and support vector regression (SVR) are two of the non-parametric methods for spatiotemporal traffic forecasting that have been tested the most frequently. However, the majority of these shallow ML systems work in a supervised fashion, which means that the dataset manual feature selection criterion determines how well they function. Using back propagation techniques in artificially connectedneural networks (ANN), a slightly more advanced dense supervised learning approach is used for traffic predictions as a result of the development of ML algorithms. While ANN outperforms traditional linear parametric models, it has difficulty learning simple time series data and locating global minima. Recent developments in deep recurrent neural networks (RNN) for dynamic sequential modelling, particularly in voice recognition, have shown tremendous promise. Nevertheless, gradient explosion for extra-long sequence training causes information loss and decreased performance in simple RNN. In order to estimate traffic, Fundamental have employed RNN variations known as long short term memory (LSTM) and gated recurrent units (GRU).

Table 2 Techniques for Machine Learning that is Data-Driven and Non-Parametric

Models	**Limitations**
K-means and hierarchical clustering, linear regression, and random forest are examples of K-nearest neighbour (KNN) models (RF).	performs better if the correlation between thedata is too low. Moreover, high dimensionaldata classification considerably degrades algorithm performance. Yet, traffic series data has a high degree of correlation
Support Vector Regression (SVR)	
Multi-Layer Feed Forward Neural Networks (ANN), Back-Propagation Neural Networks (BPNN), and other neural network variations	Although it outperforms conventional linearparametric models, time series data makes itdifficult to find the absolute global minimumduring the data learning phase because the common non-stationary data traffic series may contain many minimums.
Long short-term memory (LSTM) and gated recurrent units (GRU), two of the RNN's variants	Gives the possibilities to simultaneously learn over several time increments and is designed to deal with time series data prediction difficulties. Recurrent feedbackhas improved relative prediction performance over simple NNs.
Convolutional neural networks (CNNs) with many layers and deep learning models	In order to learn time stamp data, raw deep learning algorithms must be modified in some way, as the dependency may be a significant problem due to large data co-relations and dispersed trends, which are notalways visible while learning to classify images (same group of pixels might appear together that makes the classification job easier). If the input to CNN is formattedcorrectly, a similar principle is applied to leverage the spatial data from traffic series.
Bayesian Networks, Naïve Bayes, and Self-organizing maps	Naive Bayes, to mine spatiotemporalperformance trends for both individual links and the network as a whole. For calculating the probability distributions of connected features, Bayesian networks and naive Bayestreat them as mutually independent occurring events. This reduces computer complexity. They can be used for conditional probability- based flow estimates at the connection level farther down
Principal Component Analysis (PCA)	In the first place, PCA is a dimensionality reduction technique focused on identifying the most pertinent Eigen matrices of data variables. This could take into account variables that have little to no influence on the final forecast, making them unsuitable forpredicting traffic. Instead, they are frequentlyused to reduce the dimensionality of features.

➢ *Hybrid Models*

Parametric models typically display lower prediction accuracy and are responsive to parameteradjustment. On the other hand, hybrid models combine the strength of both parametric and non-parametric models and inherit properties from both to produce an accuracy model that performs better. Deeply complex models can also be categorised as deep learning models because each one makes use of a distinct feature or issue in the data.

Table 3 Hybrid Models

Models	Limitations
Space-Time Hidden Markov Model (STHMM), Exceptionally RandomizedTrees (ET), and Spatial-Temporal Random Field (STRF) for Parameter Learning	With high loads of correlated data, the segmented model's postulated additive (and uncorrelated) structure is less accurate. whenearlier data is more similar to the current data(evening and seasonal easily predictable trends).
Relevance Vector Machine Regression, Temporal Window based Support Vector Machine, and Least Square Support Vector Machine (RVM)	The failure of LS-future SVM's predictions of exceptionally high flow rates may be due to the sudden increase in traffic congestion and unanticipated patterns. It is occasionallynecessary to estimate the upper and lower boundary values in advance for RVM to predict the variable.
Fuzzy Logic Controlled Deep NeuralNetwork (FCDN),	Because the fuzzy rules must be established in advance and FCDN employs the fuzzy login in back propagation weight training, it is semi-unsupervised learning and challenging to deploy in traffic situations.
De noising Stacked Auto encoders, Stacked Auto encoder (SAE)	Although it over fits the input data, it lacks the time series-based data's intrinsic capacity to generalise. More recurrent layers mayimprove SAE learning.
Ada Boost	relates to an ensemble constructed iterativelyby reweighting the learning samples in accordance with how effectively the current ensemble predicts the goal variable. The weight increases in direct proportion to howbad the prediction is. Thus, the prediction algorithm determines the overall performance accuracy.
Gap-Sensitive Windowed KNN (GSW-KNN), KNN-PCA,KNN-RFE, Random Forests-PCA,Random Forests-RFE	combining the strengths and weaknesses of two ML algorithms. For time series traffic parametric data, GSW-KNN appears to be a good method for imputation of missing data.
Convolutional Neural Network (CNN) –Recurrent Neural Network (RNN) (DeepTransport) Convolutional LSTM NN (Conv LSTMNN)	the deep learning models that combine LSTM's temporal capabilities with CNN's ability to exploit spatial features. Generally Dealing with structurally lacking data is good. To learn spatiotemporal properties of surrounding roads. Important road segmentscan be differentiated based on their sequence by ranking them according to distance. goodfor mining road topology learning.

F. Summary

This portion examines multiple data prediction methods for the traffic-related factors found in the review. The latest techniques of diverse statistical flow forecasts are also investigated, although the emphasis were techniques for modelling using machine learning. Certain aspects of these closely connected fields are also investigated since the traffic prediction model and methodologies are flexible. Forecasting bus headway stops, forecasting congestion, and predicting passenger wait times at bus stops are all closely linked fields. The benefits and drawbacks of the literature-extracted techniques' adaptation to the suggested methodology arecontrasted in the next chapter. Modern methods for predicting traffic are reviewed in this chapter. Based on how well they are built, similar models are divided into three groups.

CHAPTER THREE
MODELS AND ARCHITECTURES

This chapter goes into great detail on the particular models that were selected. After that, the experiments employ the models. In section the justifications for selecting these particular models are discussed, and in section, the complete implementation information, including the frameworks utilised and the pipelines for each model used to find the appropriate model parameters, is listed.

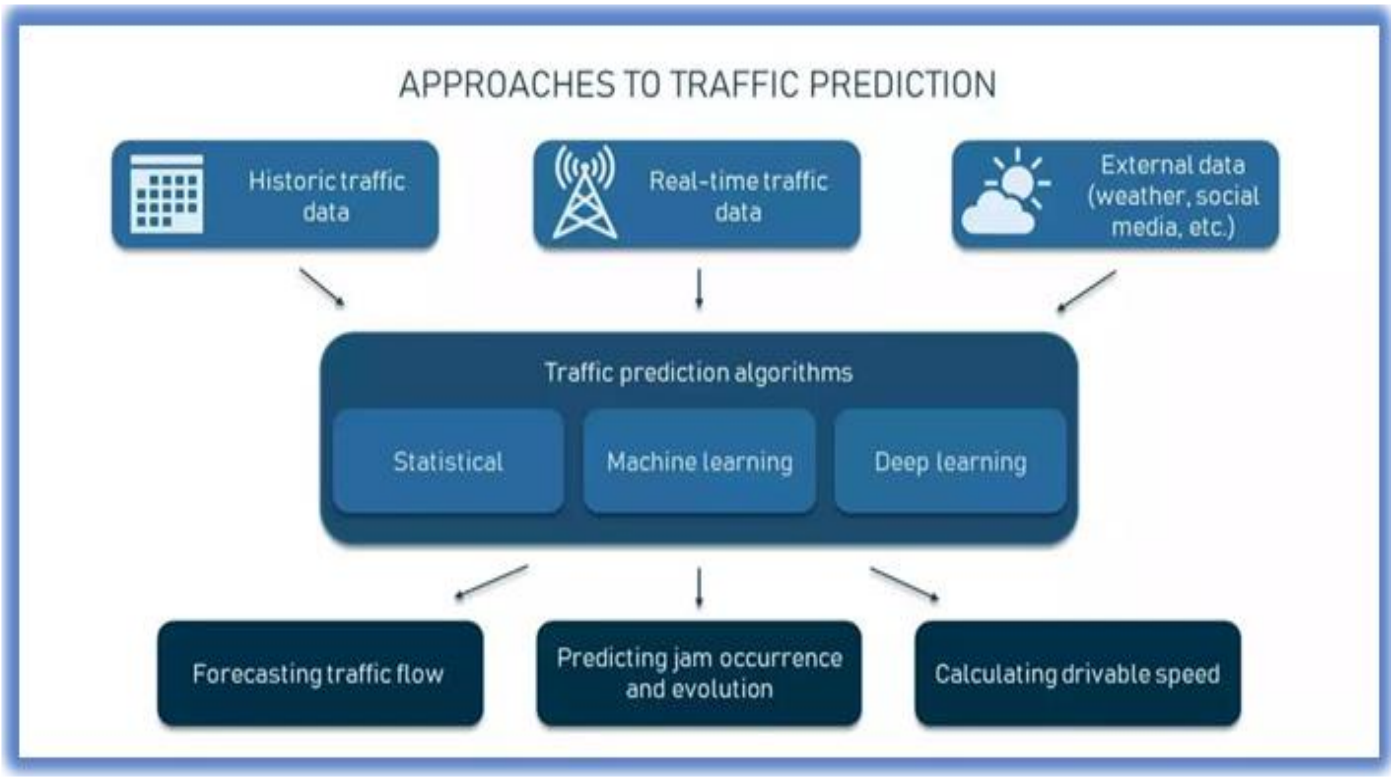


Fig 4 Approaches to Traffic Prediction

A. *Selected Models Theory*

The models that have been used are detailed in this section. According to studies, non- parametric models are preferable than parametric models for the problem-learning componentbecause they are more adept at generalising complicated data and are better able to adapt to its patterns, such as forecasting traffic data. The underlying statistical distributions in the data are assumed by parametric tests and procedures. Since the link between the input and output traffic variables is nonlinear and poorly understood, parametric techniques are typically the first to be used. The non-parametric pattern recognition-based algorithms, which are a subset of these, appear to be more suitable because they are good at locating similar traffic situations needed to provide predictions. Ten data-driven models are included in this list: The Autoregressive Integrated Moving Average (ARIMA) Model Seasonal Autoregressive Integrated Moving Average (SARIMA), Random Forest Regression (RFR), Linear Regression, Long Short-Term Memory (LSTM), and sophisticated hybrids of Backpropagation Long Short Term Memory with Neural Network (DCNN-LSTM), GRU models. The necessary sections provide a full mathematical reasoning as well as the justifications for the models used.

➢ *Models or Algorithms*

Time series data broadly defined in two categories, univariate series and multivariate series. To analyse both the data, broadly three categories of algorithms or methods used for analysis,these are

- Classical/Statistical methods: Moving Averages (MA), Auto Regression (AR), Exponential Smoothing, Autoregressive integrated moving average (ARIMA), Sessional SARIMA, Trigonometric seasonality Box-Cox transformation ARIMA errors Trend (TBATS), Vector Regression VAR, VARMAX etc.
- Machine Learning: Linear regression, Random Forest, XG Boost etc.
- Deep learning methods, Recurring neural networks (RNN), (Long short term memory)LSTM, GRU, Transformers.

➢ *ARIMA (Auto Regressive Integrated Moving Average) Model*

One of the most well-liked traditional approaches to time series forecasting is ARIMA. Basedon its own prior values, or its own forecasting errors and lags, it predicts a given time series. ARIMA is made up of three parts.

- *Auto Regression (AR):*

A model in which an alterable variable regresses on itself lagged, or prior, values are one that exhibits this phenomenon. The number of prior data points to be usedis determined by this order value: Calculate the p value typically by testing many values and determining which model met the minimum (AIC).

To enable the time series to stop fluctuating, the Integrated (I) symbol symbolises the differencing of raw observations (i.e., The difference between the data values and the prior values is used to replace the data values).

- *Moving average (MA):*

Takes into account the connection between a measurement and a residual error that results from using lagged observations with a moving average model. Presentation of the ARIMA model mathematically.

Several other variations of this model, including ARMA (Autoregressive Moving average), SARIMA (Seasonality with ARIMA), and SARIMAX, are also used for forecasting (Inclusionof exogenous variable).

$Xt=\sum\varphi jxt-j+\sum\theta jet-j+et$ –Forecasting expression

$\sum\varphi jxt-j$ – Auto Regression of order p

$\sum\theta jet-j$ −Moving average of order q

$\{et\}$ −Residuals

- *Seasonal Autoregressive Integrated Moving Average Model (SARIMA)*

The most widely used statistical model, the ARIMA model, is a useful starting point for comparison with other machine learning methods. The order differencing, of the moving average (MA) model, and auto-regressive model (AR) make up the three simple models that make up the ARIMA model. The total model is frequently expressed as ARIMA because p, d,and q are utilised to the model's features as representations (p, d, q). The data from a non- stationary univariate series are represented by the integrated term d. The parameter "d" establishes the delay between the response and the calculation of the difference. For seasonal non-stationary data, The Augmented Dicky Fuller (ADF) test is used throughout the experiments to test the time series' stationarity because ARIMA anticipates that the input timeseries will be seasonal and stationary. As its name implies, SARIMA (p, d, q, P, D, Q, m). Theorder for the AR model is P, the seasonal component order for the integration model is D, the seasonal component order for the MA model is Q, and the seasonal component order for the seasonal lag consideration is m, which stands for the cyclic seasonal period of time steps, the SARIMA model was used to analyse the time series.

$$\text{SARIMA }(p, d, q)\, x\, (P, D, Q)\, m= (\Phi\, (L)\, (L^{m})\, \Delta^{d}\, \Delta m^{D}\, yt= \theta 0+ \Theta(L^{m})\, \theta(L)\epsilon t)$$

Where t is the Gaussian white noise process with zero mean and variance, L is the lag operator,and m is the seasonal length. The difference and seasonal difference operators, with d and D standing in for their respective orders, are d and m^{D}. The non-stationary time series is convertedinto a stationary time series with the use of the difference procedures. The autoregressive lag polynomials (L and L^{m}) are multiplied to create the AR component of this model. Moreover, the moving average lag polynomials (L) and (L^{m}) are used to represent the MA component.

- *Random Forest Regression (RFR)*

The data-driven Random Forrest Regression (RFR) is chosen as one of the models for contrast from a useful standpoint because it fits quite well, fits quite quickly due to its small number of parameters, and has an intuitive fit model that enables visualising the order in which variables are ranked according to how important they are in contrast to other machine learning models. The outcome is the average of the regression-based trees that make up RFR. A selection of m data variables is represented by n nodes in each tree. The independent variable values used to generate the node leaves are chosen so that the sum of squares involving that independent variable is minimised when the dependent variable is averaged on each side of each value.

- *Pseudocode 1: Training for RFR*

Information: X training samples, m characteristicsResult: skilled model

- ✓ Choose k features at random from m given that k>m;
- ✓ Determine the node d with the best split point based on k characteristics;
- ✓ Node split using additional best splits to form daughter nodes;
- ✓ Repeat steps 1 through 3 until the maximum number of leaf splits equals the number ofnodes;
- ✓ Repeat steps 1 through 4 until n=max number of trees are divided.

ISSN No:-2456-2165

- *RFR Prediction Data for Pseudocode 2:*
  Test samples Y, characteristics pResult: Estimated Value

- ✓ Iteratively choose p characteristics from each sample of Y, and use the trained model's ruleof built trees to forecast and store each tree's predicted results;
- ✓ Determine the number of votes or nearly identical votes for each scenario projected; 3 As theresult of the trained model, take into account the outcome that received the most votes or the average of closely comparable outcomes.

➢ *Deep Learning Methods*

Numerous data science domains make considerable use of deep learning techniques. This is used to benefit analysis that deals with time series. Context vectorising is a technique used withrecurrent neural networks to describe sequential data with the time step index t. Context vectoring serves as "memory," capturing data regarding calculations made thus far andallowing RNNs to recall past data, allowing them to retain data of long and variable sequences.RNNs can therefore receive one or more input vectors and output one or more output vectors.RNNs and deep neural networks are conceptually similar. They have output vectors, hidden states, weight vectors, and input vectors. The patterns or context are captured by the concealedstate.

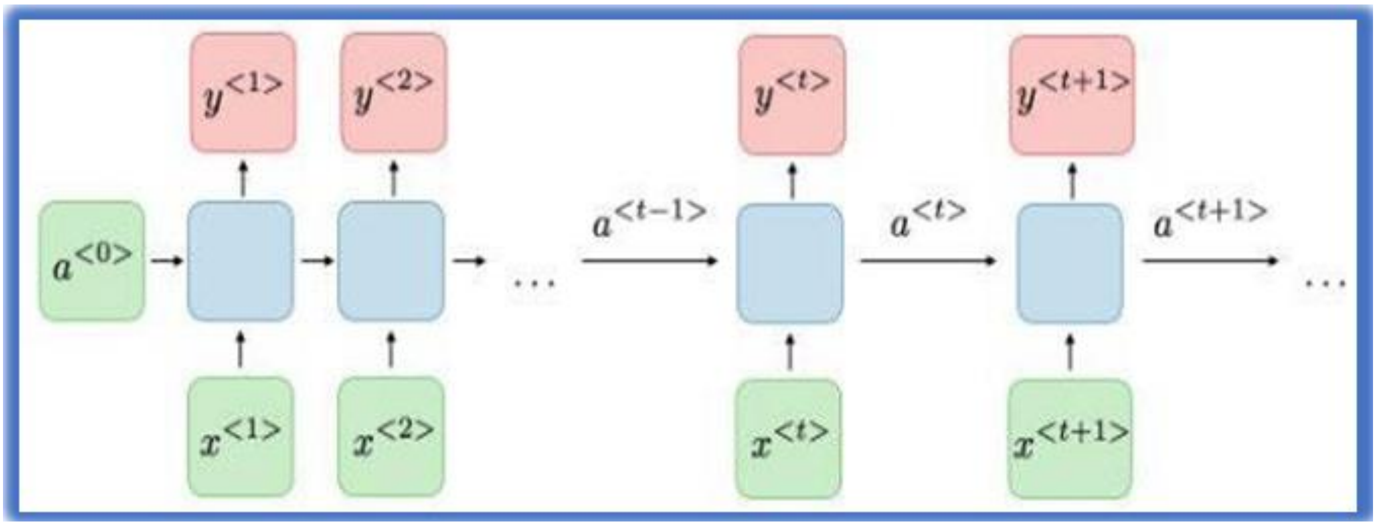


Fig 5 RNN Network

➢ *Convolutional Neural Network (CNN)*

The deep learning family of convolutional neural networks (CNNs) makes advantage of the spatial structure of input (such as photographs) learn the characteristics of the data so that the final design can learn anything meaningful. This model aids in identifying the data's local properties. The FFBNN deep model extension is the CNN model. with extra convolutional layers at the input as well as further hidden layers. Convolutional layers, pooling layers, and afully linked layer make up a conventional CNN. Except for the fact that convolutional layers are applied at the beginning of the typical ANN model and pooling layers are applied in between the ANN layers, the CNN inherits all the features of an FFBNN model. Convolution is a function-to-function procedure. depicting one state, and its pixel values represent values ofthe scaled inputs, serves as CNN's input. Convolutional layers function as a filter that, when included in the model, highlights particular traits of the input feature vectors. In order to produce a feature map, it acts similar to a custom automatic feature detector. When strides ontothe input feature matrices, functions have between two and five elements per dimension, and on the remaining elements they are deemed zero. Kernels are the resulting tiny matrices that represent the filter function groups. Each kernel cell's element-wise multiplication with the matching feature at a specific point of the convolutional kernel.

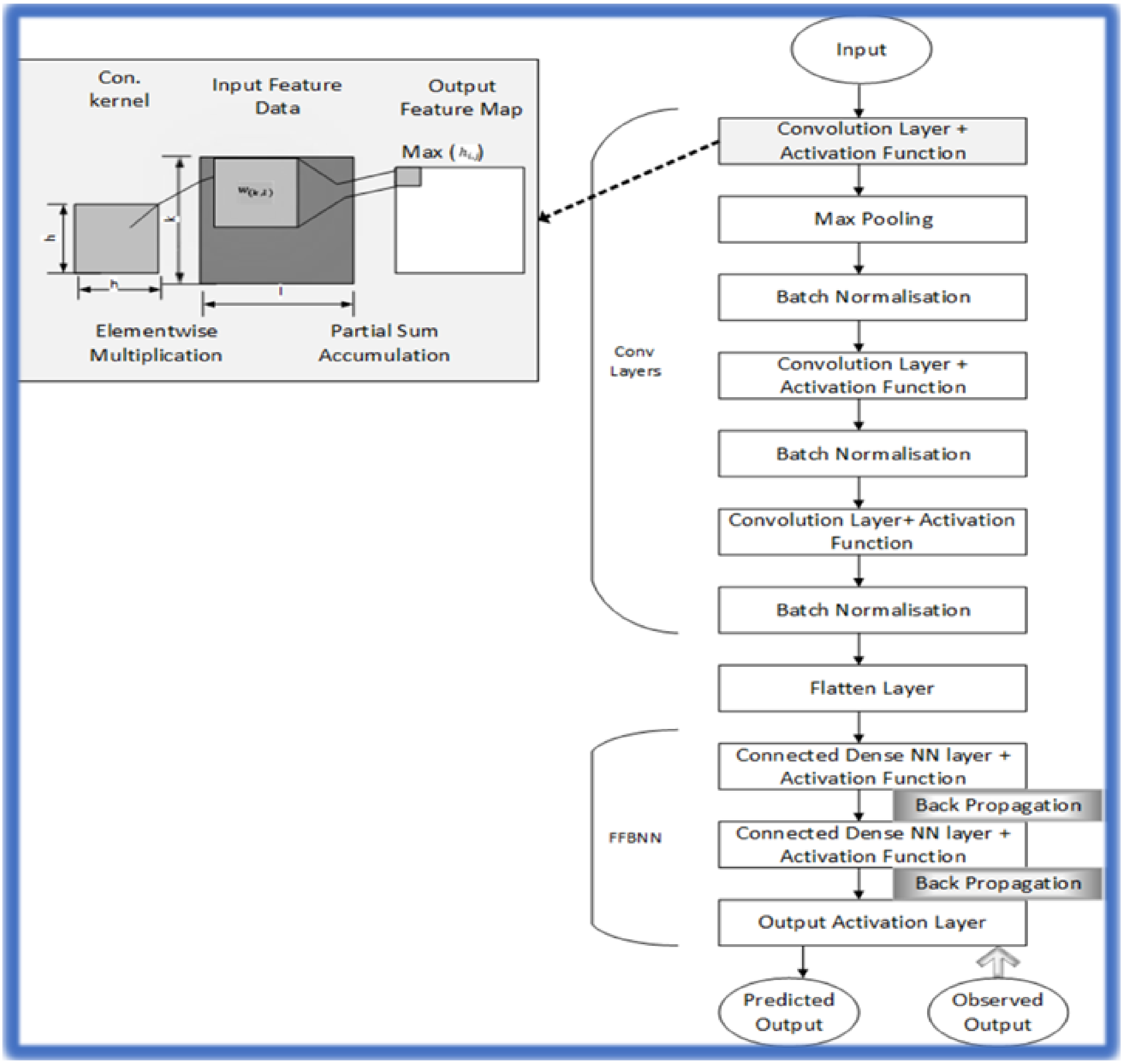


Fig 6 Steps in Convolution Operation (Left) and CNN (C-FBNN) Model (Right).

➢ *Briefly about LSTM*

Long data dependencies are present in extended time sequence. The traditional neural based network is unable to learn these long-term dependencies. Long Short-Term Memory (LSTM) based on recurrent neural networks (RNN) are utilised for this. The difference between an RNN and a feed-forward neural network is that an RNN is a recurrent value learner since each neuron unit's output is returned to its source. A more sophisticated version of RNN is LSTM. LSTM were first developed for language processing approximately two decades ago, where they were effective for improving speed while memorising dependencies across time in the data. An LSTM is a type of storage block that memory cells can control using their individual input, xt is the feature input to the memory unit whereas it, ft, ot, ht , ct represents the input gate's output., the final cell state output, the final output gate output, the final memory unit output, and the forget gate output, respectively.. Wxi, Wxf, Wxo represents, in turn, the weights between the input layer and the input gate, the input layer and the forget gate, and the input layer and the output gate. Similarly, Whi, Whf, Who are the weights assigned to the input layer, forget gate, and output gate, respectively, in the recurrent hidden layer. Similar to what the subscript says, Wci , Wcf , and Wco are the weights connected to the input gate, forget gate, and output gate, respectively, as well as the cell state. Each of the gates has a variable named b that represents all of those variables, according to the formulae. The activated sigmoid function is represented as. The last LSTM memory unit of one layer is passed as an input to the subsequent layer memory unit, while the output of the hidden recurrent unit, ht, is sent from the earlier LSTM memory device the subsequent LSTM unit. Backpropagation are used for training the LSTM model layers for several optimisers, and the final model forecasts are selected as best performing layers parameter.

ISSN No:-2456-2165

$$i_t = \sigma\,(x^t\, W_{xi} + h_{t-1}\, W_{hi} + C_{t-1} W_{ci} + b_i\ )$$

$$f_t = \sigma\,(x^{\,t} W_{xf}\ + h_{t-1}\, W_{hf} + C_{t-1} W_{ci} + b_f)$$

$$O_t = \sigma\,(\,^t\, W_{xo} + h_{t-1}\, W_{ho} + c_t W_{co} + b_o)$$

$$c_t = f_t\ C_{t-1} + i_t\ \cdot \tanh\,(x_t\ W_{xc} + h_{t-1}\ W_{hc}\ + b_c\ ) h_t = o_t \cdot \tanh\,(c_t)$$

$$y_t = W_{yh}\ h_t + b_y$$

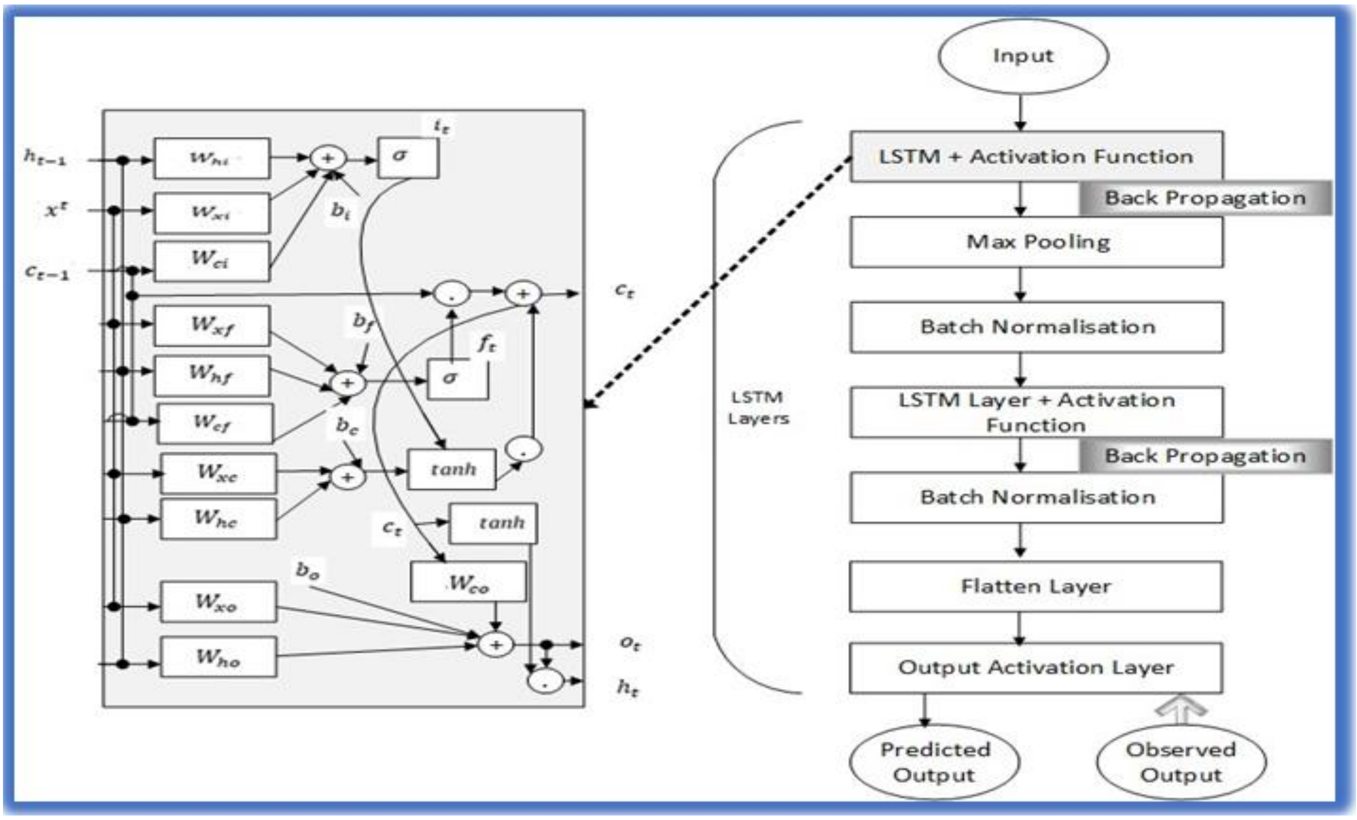


Fig 7 LSTM Memory Unit Structure (Left) and Stacked LSTM Model (Right).

Long Short Term Memory (LSTM) is one of the popular model based on RNN being used forprocessing sequential nature of data.

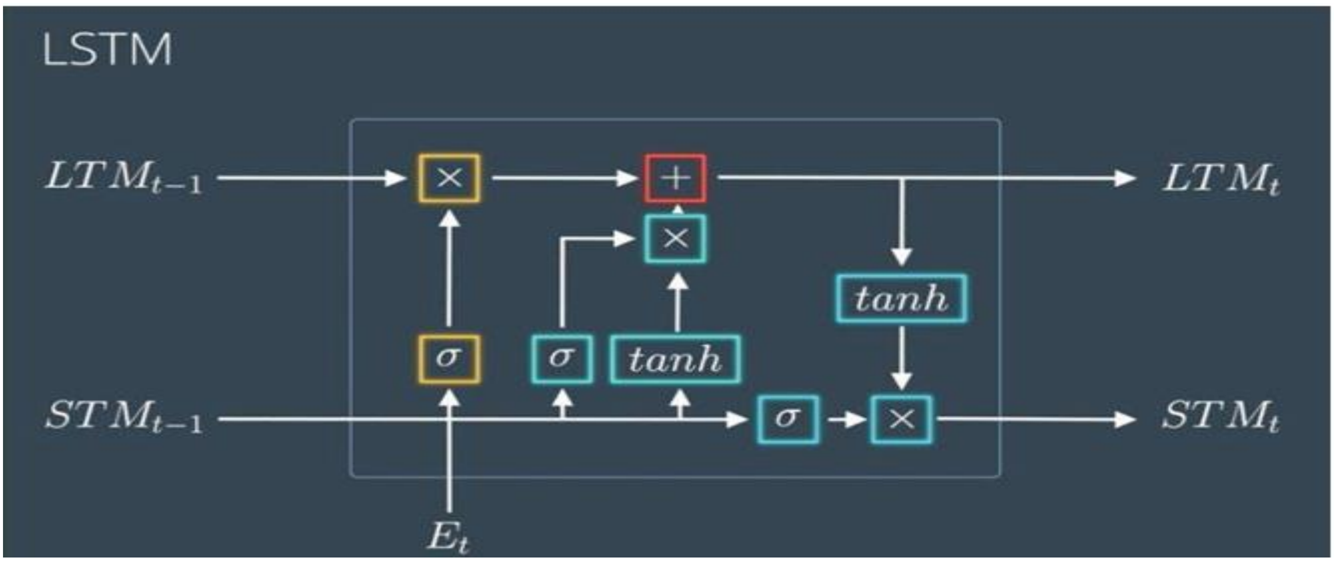


Fig 8 LSTM Structure Courtesy https://www.analyticsvidhya.com

ISSN No:-2456-2165

➢ *The Architecture of LSTM:*

The notion of gates is used in LSTM to simplify and optimise calculations for both Long Term Memory (LTM) and Short Term Memory (STM).

- Forget Gate: As the LTM goes through the forget gate, it forgets information, but that isuseless.
- Learn Gate: Event (current input) and STM are coupled so that the necessary knowledge wepreviously acquired through STM can be applied to the current input.
- Remember Gate: In the Remember Gate, which functions as an updated LTM, LTMinformation that we have not forgotten, STM information, and Event information are joined.
- Use Gate: This gate additionally forecasts the outcome of the current event using LTM,STM, and Event.

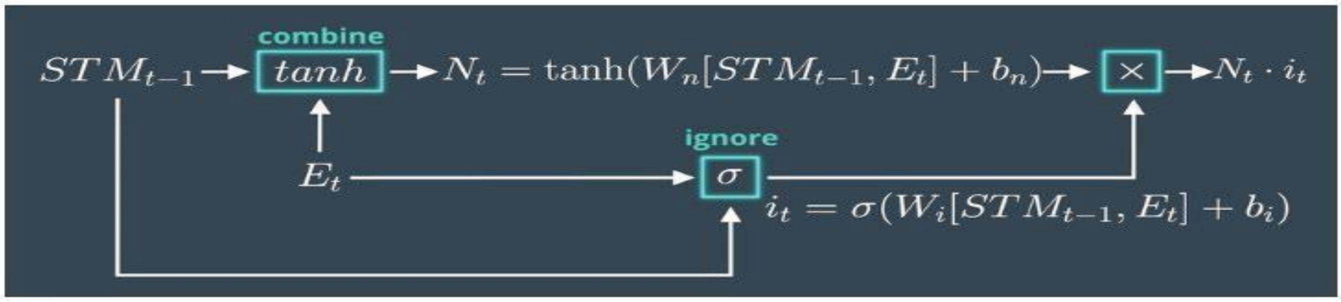


Fig 8 Mathematical Intuition of LSTM Structure

B. Hardware and Software Implementation Details

Python 3.7.11, a well-known simulation and programming language, is used for all development and testing. The libraries, dependencies, and binaries for data science and machine learning were compiled using the open source distribution Anaconda. Python is a well-known using dynamic language interpretation that is quick and appropriate for real-time dealing with applications. Since scientific computing libraries already in existence and intensive C-written processing is easily integral with python, the popularity of python has increased. The existing substantial and widespread community support of Python developers is one of the other causes.

➢ *Data Exploration Library*

The panda's library for Python was used for data exploration and preliminary analysis. Pandas handles data exploration by transforming information into frames and tabular data structures with columns. Pandas2 is a useful scientific tool that makes it simple time series resampling data, index the data frame again, or to group using any column headings to better comprehend the visualisation of data plots created from data frames directly.

➢ *ML Implementation Library*

Model designs that have already been addressed in earlier parts are implemented utilising a variety of open-source library packages. The two most popular machine learning libraries are TensorFlow4 and Keras3. A high-level machine learning API called Keres is built on top of the Tensor Flow framework and is written in Python. Because to its quick experimentation capability, Keras is recommended over Tensor Flow because it can build the sophisticated network from scratch in minutes as opposed to longer time with the Tensor Flow library. Convolutional and recurrent neural networks are also supported by Tensor Flow deals with model computing as graphs, which enables the development of novel architectures based on fundamental unit constructions. A few Tensor Flow capabilities include simple model creation and graph computations that can be handled by both CPUs and GPUs, making it simpler to use Keras API on top of it. Scikit-learn5 is a Python library that was used to prepare the training and validation datasets as well as to pre-process the data before training. Model flow pipelines were created to make it simple to locate the most effective criteria for each unique model using the scikit-learn library's grid search function allows for end-to-end model training with the ideal parameters. Appendix A lists the model parameters with the best performance. It's interesting to note that The functions in scikit-learn are designed to work with Keras high-level prediction functions.

- https://www.anaconda.com/distribution/
- https://pandas.pydata.org/
- https://keras.io/
- https://www.tensorflow.org/
- https://scikit-learn.org

CHAPTER FOUR
METHODOLOGY AND CONTRIBUTIONS TO THE RESEARCH

A. Introduction

The review of the literature is covered in chapter 2 along overlook of various techniques for methods for analysing and predicting traffic congestion, travel times, and traffic flow built around cutting-edge machine learning algorithms. The selected models are covered in more detail in chapter 3. This chapter includes information on the potential datasets, a breakdown ofthe dataset's applicability for the research approach, and a recommended procedure for the machine learning models that have been selected.

B. Data Description

One of those bothersome issues that many of us who live in urban areas deal with is traffic. The growing urban population is one of the factors contributing to traffic. There is an influx ofresidents looking for work and opportunity despite the outdated infrastructure that can only support a small population. Fuel combustion rises as a result of traffic congestion. It intensifiesthe air pollution-causing carbon emissions. Time and money are additional costs. According toa survey by INRIX, a provider of mobility analytics and connected car services, Americans lost 99 hours of work per year on average due to traffic in 2019, costing them $1,377 on average. As economic and urban expansion continue, American drivers have lost more time onthe road on average from 2015 to 2019. link https://inrix.com/press-releases/2019-traffic- scorecard-us/

I will be investigating the dataset of four intersections and developing a model to forecast traffic on it for this project. By providing a better understanding of traffic patterns, which will further help in constructing an infrastructure to eradicate the problem, this may help in alleviating theproblem of traffic congestion. There are 48,104 number of data set which is collected.

C. Investigation Area

The method and standards used to collect the data on traffic flow characteristics are describedin this section. This dataset is a compilation of hourly counts of vehicles at four intersections.There are four features in the CSV file:

- Date Time
- Junctions
- Vehicles
- ID

The traffic data comes from several time periods because the sensors on each of these junctions were gathering data at different times. Data from several of the intersections wasscarce or limited.

D. Dataset

Based on the data. For this paper's prediction and application of machine learning models for traffic flow, world site data are gathered. This prediction method had results. The data collection includes a record of the number of vehicles that are detected using road sensors as they pass through various intersections. From 2015 to 2018, traffic data was logged every hour.The. There are 48,104 different data sets that have been gathered. This dataset is a compilationof hourly counts of vehicles at four intersections. There are four features in the CSV file: Date/Time, Junctions, Vehicles, and ID. The traffic data comes from several time periods because the sensors on each of these junctions were gathering data at different times. Some ofthe junctions have provided limited or sparse data.

Creating a sub features out of Date Time. Namely:

- Year
- Month
- Date in the given month
- Days of week
- Hour

Table 4 Investigation Area

	DateTime	Junction	Vehicles	ID
0	2015-11-01 00:00:00	1	15	20151101001
1	2015-11-01 01:00:00	1	13	20151101011
2	2015-11-01 02:00:00	1	10	20151101021
3	2015-11-01 03:00:00	1	7	20151101031
4	2015-11-01 04:00:00	1	9	20151101041
...	...	...	...	...
48115	2017-06-30 19:00:00	4	11	20170630194
48116	2017-06-30 20:00:00	4	30	20170630204
48117	2017-06-30 21:00:00	4	16	20170630214
48118	2017-06-30 22:00:00	4	22	20170630224
48119	2017-06-30 23:00:00	4	12	20170630234

48120 rows × 4 columns

Table 5 Dataset

	DateTime	Junction	Vehicles	ID	Year	Month	Day	hour
0	2015-11-01 00:00:00	1	15	20151101001	2015	11	1	0
1	2015-11-01 01:00:00	1	13	20151101011	2015	11	1	1
2	2015-11-01 02:00:00	1	10	20151101021	2015	11	1	2
3	2015-11-01 03:00:00	1	7	20151101031	2015	11	1	3
4	2015-11-01 04:00:00	1	9	20151101041	2015	11	1	4
...	...	...	...	...	...	...	...	...
48115	2017-06-30 19:00:00	4	11	20170630194	2017	6	30	19
48116	2017-06-30 20:00:00	4	30	20170630204	2017	6	30	20
48117	2017-06-30 21:00:00	4	16	20170630214	2017	6	30	21
48118	2017-06-30 22:00:00	4	22	20170630224	2017	6	30	22
48119	2017-06-30 23:00:00	4	12	20170630234	2017	6	30	23

48120 rows × 8 columns

E. Data Preparation

Flow-based forecasts are operationalized through a multi-stage procedure. The procedure begins with several of real-time data streams that contain series of data for every relevant connections or nodes of the initially selected a road system serving the research area. The trained classifier model is tested against the incoming real-time data, just like with any other machine learning technique, to predict the prediction variables. The verification results are produced on the set of validation set of validation in a traditional machine learning model implementation to assess the performance effectiveness the tested algorithms' prediction accuracy.

- *Data Cleaning*

Around 15% of the variables in the raw link dataset were missing. It is crucial to preserve the inherited trends in the traffic data owing to persistent patterns like seasonality, as well as otherenvironmental influences. In order to infer the missing data, the backward fill method is used. Using the, the flow value is imputed previous interval's first recorded worth in the backward fill approach. Until all of the missing values were imputed, this imputation process was continued. Even though the data inconsistency was fixed, if the rate of missing values is too large, this strategy may endanger the inherit data attributes.

- *Data Integration*

For each considered link, 48180 data samples are used in total. In order to create an array of 48180, number of considered flows at the junction, they are reshaped in accordance with equation. The current experiment takes into account the 48180 x 4 dimensions using the link predictions connected to patch 1 node 2 as a test example. Only when multiple features are taken into account for a node's predictions does this reshaping take place.

- *Data Normalisation*

The next step is "intra flow linkages normalisation," which further generalises and normalisesthe data by sizing it for the lowest and highest values contained within each data column. The reshaped dataset is lagged by one-time interval, two-time interval, and three-time intervals in order to further prepare it the experiments section for further information on supervised training for short-, medium-, and long-term forecasts.

- *Data Reduction*

40% of The validation set is created using the original dataset. in order to construct the trainingand validation sets for the ML. For each selected model, K fold cross validation is carried out three times using the validation and testing data. The sequencing of the training and validationensembles is crucial since the data is time series data with subsequent intervals. Consequently,following each training iteration, the tail end 40% series data are taken into account for the training models for validations.

- *Data Discretisation*

With the dataset that was first made available, there are twelve periods in a twenty-four-hour time window for the multi feature prediction model scenarios. Only the twelve intervals, whichare separated by two hours each, are taken into consideration in order to accelerate the ML models' training and provide them a more generalised a visual representation of the daily sequential data.

- *Dependent and Independent Data Variables*

According to a survey of the literature, numerous different kinds of variables have been utilisedto analyse traffic movement prediction issues. Spatial variables (such as a connecting road's traffic flow), temporal elements (such as the time of the collected data), and seasonal factors are some of the frequent variables that have been employed in past studies (weather conditions). Field variables from the AADF and DFT datasets can be divided into independent variables and dependent variables, according to a general knowledge of the dataset's field variables. Theapplicability of our suggested objectives and model development, as well as maintaining the performance metrics in line with past studies, are all taken into consideration when choosing the dataset's dependent and independent variables. During the data preparation process, whichinvolves obtaining data from various sources and he network region, produced for several timeperiods, such as 15, 30, 45, and 60 minutes. Together with developing a preliminary understanding of the data, it is necessary to filter the trends for several independent variables, such as the day's kind. The findings of the data preparation described in subsections 4.4.1-4.4.4for the sample region taken into consideration for the experiment.

F. Methodolgy

Various machine learning (ML) algorithms are utilised for traffic prediction, but in this study, we employ the linear regression, random forest regression, Arima, and Sarimax models. We will discover the traffic forecast model with the highest accuracy among these methods. The initial collection of traffic data is crucial for forecasting traffic flow. The first collection of traffic data using ML algorithm aids in estimating the traffic. The variables and numbers havenow been converted to numerical values as part of the pre-processing of the data. Data is splitinto two sets, a training data set and a test data set, after pre-processing stages. The ML methodand training data are then selected. The algorithm will run on the practise data and output.

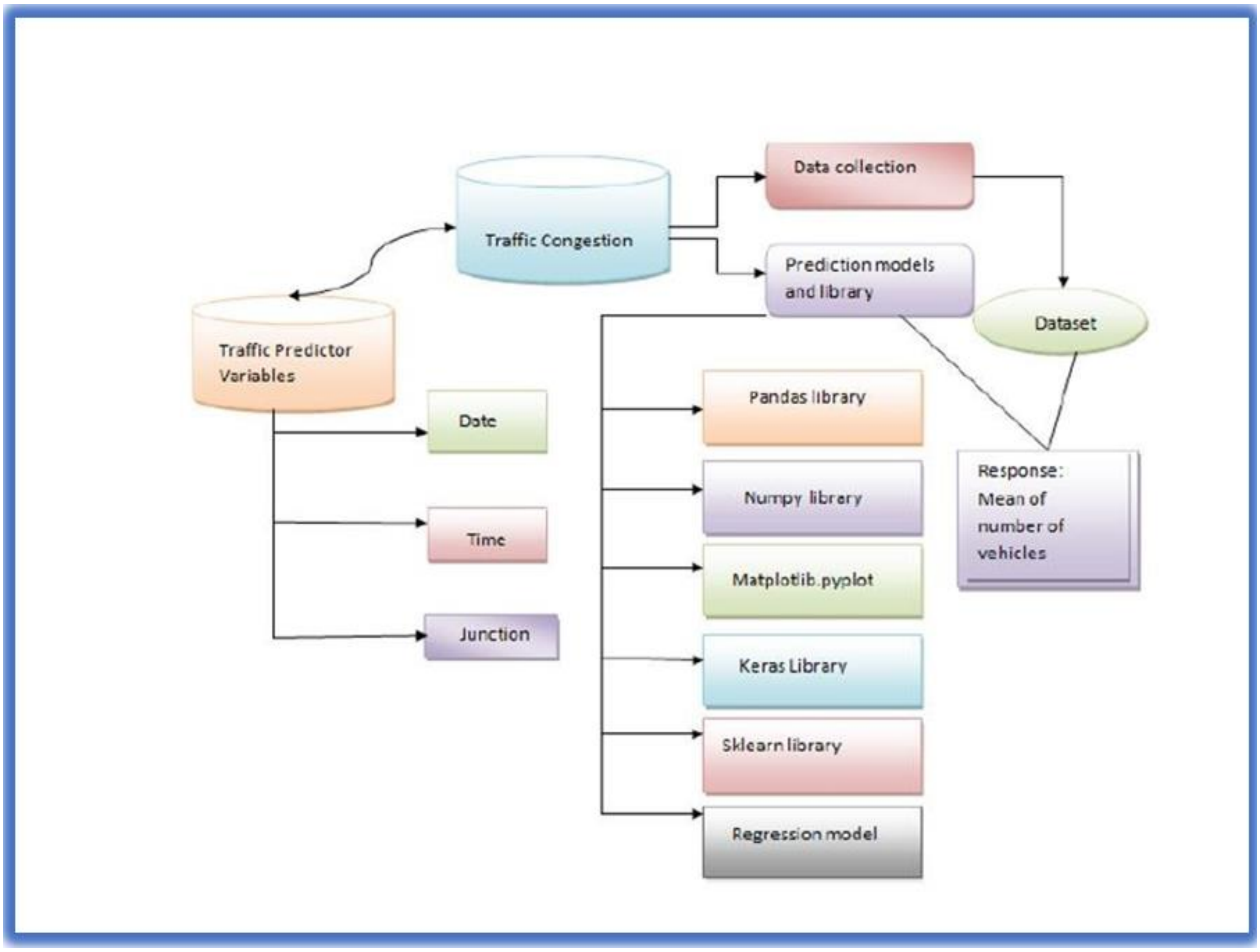


Fig 9 Methodology Overview

G. Summary

In this chapter, potential datasets for the data collection are thoroughly reviewed along with how well they fit the objectives of the study and the field of study. The MIDAS dataset is putthrough a series of data preparation stages, including cleaning, integration, normalisation, datareduction, and discretization techniques, following a full data description. The dataset's potential for dependent and independent variables is also investigated. The partitioning of thenetwork into patches and then nodes with connected traffic roads is shown. At the end of this ML and DL models' proposed topology-based network methodology is explored in detail inthis chapter.

ISSN No:-2456-2165

CHAPTER FIVE
TESTS AND FINDINGS: ANALYSIS OF THE SUGGESTED FRAMEWORKS

This chapter does a preliminary examination to select the most effective analytical techniques. The proposed study methodology for forecasting traffic flow at highway junctions is then discussed. This chapter's major goal is to give the results of a thorough investigation into the traffic flow forecast utilising statistical approaches, machine learning, and deep learning methods. This portion describes the framework of the investigation execution based on the selected ML and DL model methodologies, the recommended methodology, and the reported findings. The experimental conditions for the situations, which are provided in section 5.2, aredoes a data correlation investigation.

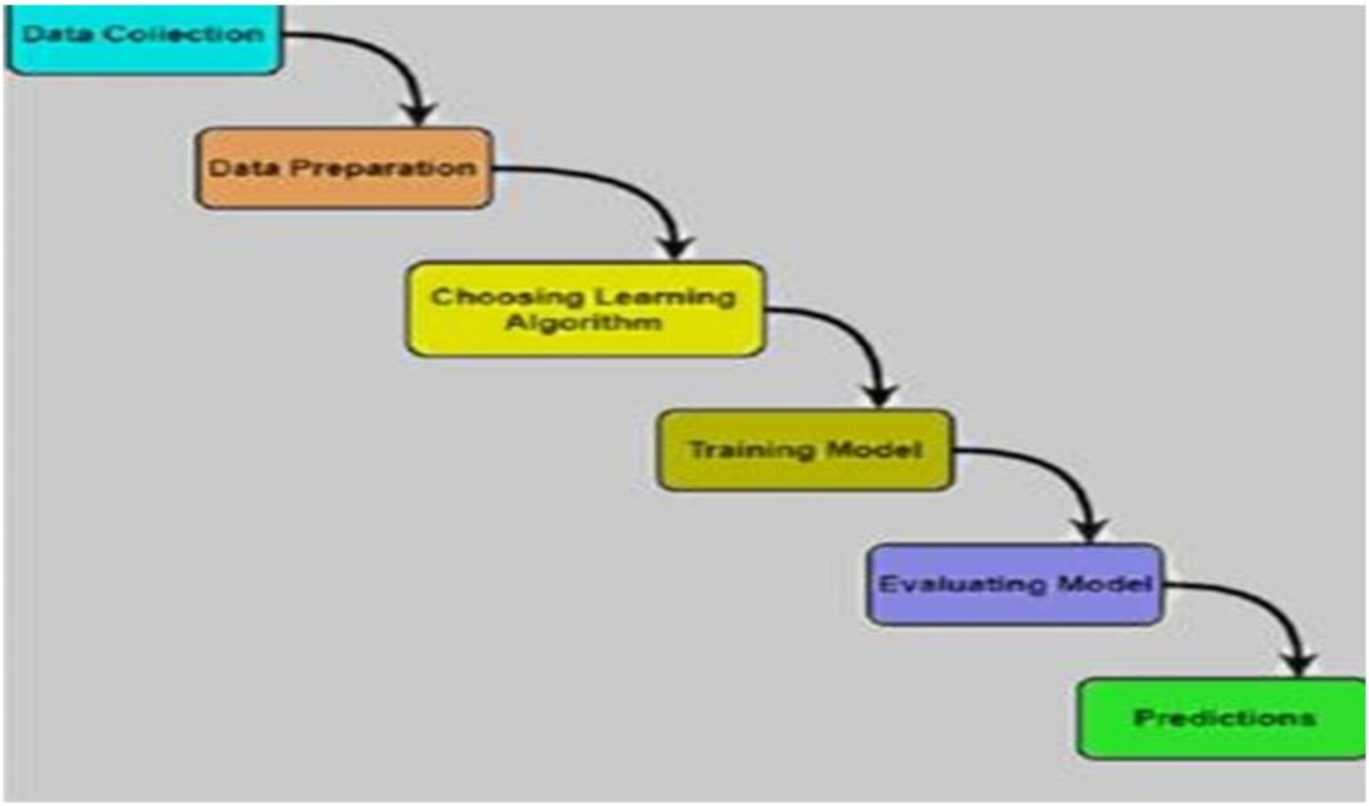


Fig 10 Methodology

A. *Experimental Environments*

This section introduces the performance indicators for reporting the top-performing individual models and the evaluation techniques for contrasting several models. In addition to testing andtraining of the suggested models, the chosen dataset is further examined for correlation patterns. For the purpose of analysing the benefits and drawbacks of the suggested approaches, various modular architectures are combined.

➢ *Performance Measurements*

Model performances are evaluated using two prediction errors, and the results are then compared. based upon the examination of estimates the two prediction error, model performances are compared: The different model accuracies are measured using the mean relative error (MAE) and root mean square error (RMSE). A set of predictions' average magnitude of mistakes is measured by MAE. The MAE, which gives equal weight to each individual difference, is the mean of the absolute sample variances disparities between forecastand actual worth. The RMSE is used to compare the accuracy of various models.

$$\text{MAE}\ (y, y') = \Sigma|y_t - y_{t'}|$$

$$\text{RMSE}\ (y, y') = \{\tfrac{1}{T}\Sigma(|y_t - y_{t'}|)^2\ \}^{/12}$$

Where the traffic flows at time t, both current and predicted, respectively, are denoted by y_t andy_t' By giving bigger weights to larger errors, these performance indices allow for the relative residual error assessment by determining the linear score with the same weight as the RMSE and MRE that averages the prediction error. It is also crucial to examine how various models function for various node linkages between various junctions. The empirical distribution achieves this.

➢ *Examination Conditions*

The selected error measures, provide each directional link's mistake of the taken-into-account nodes. With the testing data, the Empirical distribution function (EDF) and k-fold validation are used to find the right accuracies between models. In this part, the process for doing this is first introduced. Last but not least, we also explain how the inaccuracy estimates for the result are compared to the filtered data.

➢ *Distributions of Empirical Errors*

The multivariate level projections for traffic flows are produced by several models. How effectively a model predicts for each link on the investigated nodes is what distinguishes various models based on performance. The cumulative distribution function (CDF), which takes the form of an EDF, is used to illustrate the differences in error measurements across various models. Assume, for instance, that x€ X, where X represents the model's performanceindicator in the form of the estimated RMSE or MRE for the predicted outcomes for each nodeconnection.

➢ *Comparing Error Distributions*

The dataset under discussion has events and other weather-related elements as intrinsic and extrinsic characteristics. The flow and speeds based on the various vehicle classifications, as well as the time, day, week, holiday season, and typical working day, are considered extrinsicfeatures. The same goes for intrinsic qualities. It is essential to analyse the dataset for these factors. Prior to assessing the performance measures, the data filter criteria for these features are restricted in order to contrast different model performances for these parameters. When different models perform better or worse based on the volume of traffic Moreover, the normalisation of the existing traffic volume is performed in specific situations at various periods.

B. Experiments

This portion outlines the experiments that were performed and explains how they made sense. The model performances experiment for several forecast time frames are first shown based onthe model architecture of choice and which lag in input data delivers the most accurate predictions. The second instance reports the effects of including other variables in addition to the flow data on the statistical model, machine learning models and deep model findings. Linear regression, Random forest, the Auto Regressive Moving Average (ARIMA) Recurrentneural network (RNN) variants, LSTM and GRU, are the building blocks of machine learningand deep learning models. To be considered as a superior option, the newly presented model must outperform the current models. The two most basic models that are taken into consideration here are HA and RFR. Given that both models represent two unique skills that they inherited. First and foremost, any methodology that wants to be regarded as a superior model needs to perform better than RFR, as this indicates that the model being evaluated.

➢ *Case 1: Interval of Prediction*

The interval of time between the most recent observed values and the future time step that themodel is attempting to forecast. For a better understanding of the behaviour and utility of the model, various time steps or prediction intervals are taken into consideration in this situation. As a result, three separate input data scenarios are used to design and test three different interval lengths: brief interval or single step interval or one-time step (15 minutes), medium interval (30 minutes), and long interval (60 minutes). The samples dataset utilised in this study are by default recorded at intervals of fifteen minutes. One interval index equals in this definition.

71 intervals of fifteen minutes each. However, as was noted in the pre-processing of the data, the two-hour intervals are taken into account for this experiment.

➢ *Case2: Adding Associated Variables*

In the second experiment, linked MIDAS dataset variables can be used to create feature vectors.The concept of incorporating weather-related variables to enhance flow predictions has alreadybeen taken into consideration in. However, in this study, we limit the trials to just flow data, but the method can be used with the addition of variables from the MIDAS dataset that correspond to additional traffic information. All Table 4 discusses the variables that are provided by the MIDAS dataset. 2. The selection of the factors taken into consideration in thisexperiment situation is as follows:

- Vehicle speed value (the site's measurement of the mean speed in km/h for all cars travellingin all lanes during a 30minute period)
- Overall carriageway throughput (the number of detected vehicles)
- Hence, the data is recorded using a 15-minute time index. The following list of potential recorded variables can be utilised in conjunction with flow and speed characteristics to create more useful feature vectors forms the basis for future issue solving:
- (Day of the week, regular weekdays, first, middle, and last weekdays of school breaks and bank holidays, and Day of the year)
- The goal function is maintained while a deeper multi-feature deep end model is trained usingthe days of the week.
- The movement of different vehicle categories (Total range of vehicle flow)

C. Analysis of Correlation

To investigate the autocorrelation of the feature selection and the relevance of the primary selected characteristics to other taken into consideration features (cross-correlation). The totalcarriageway flow is the key feature, and the time-lag versions of link flows are the supplementary features that were taken into consideration.

➢ *Auto-Correlation*

The auto-correlation test is used to analyse the flow data with a scheduled lag to examine howdifferent time intervals affect the carriageway flow values.. This study establishes. the amountof earlier time intervals (n-steps) that should be considered when making a prediction since they are important and have an impact on values that correspond to forward time intervals. model (n-steps). The data on the first traffic flow characteristics for the incoming connection are shown in Figure as an auto correlation graph. 15-minute time step elapses between each lag. The blue shaded area denotes the correlation coefficient's 95% confidence interval. If the correlation coefficient is greater than, then there is significant autocorrelation.

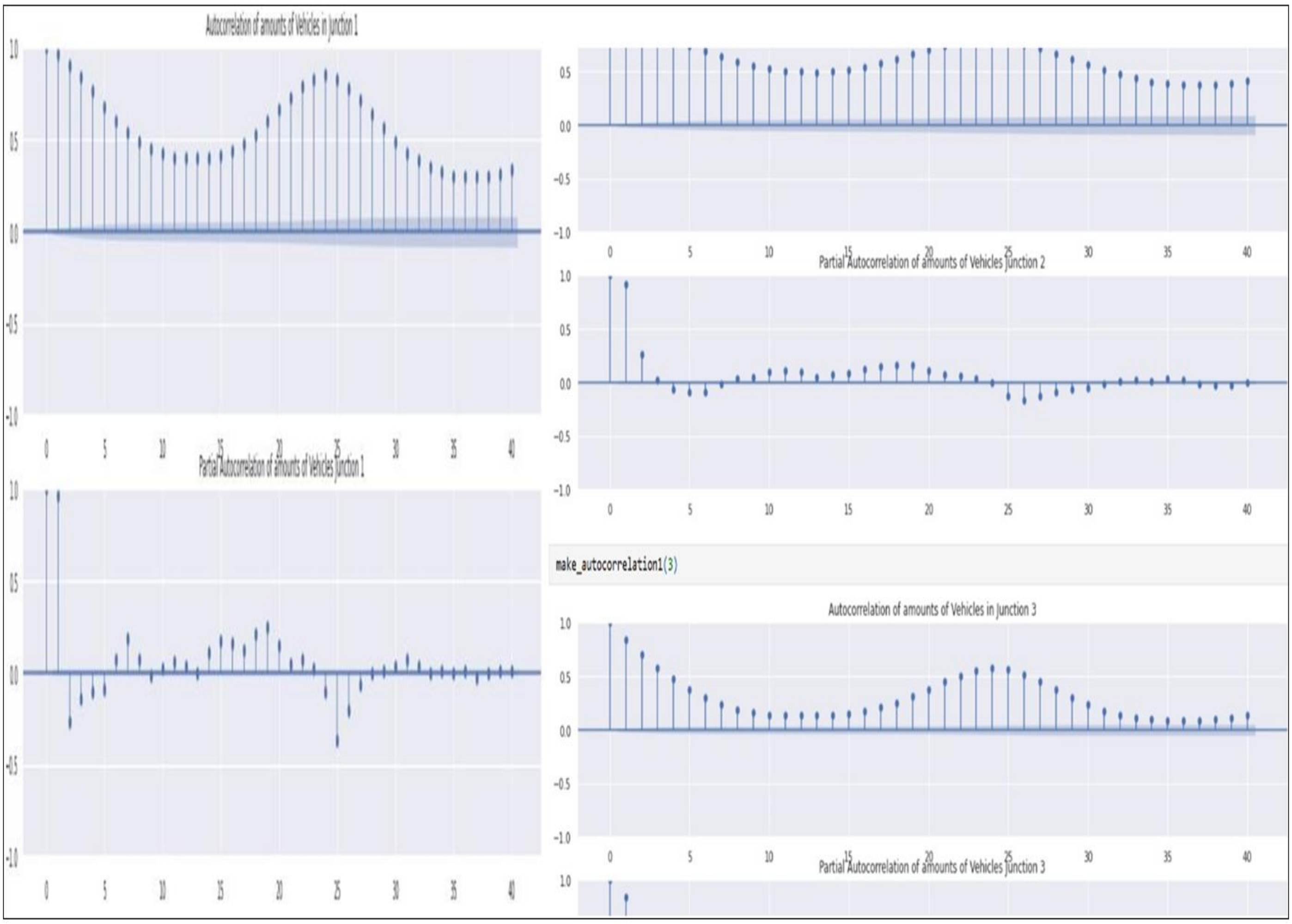


Fig 11 Autocorrelation of Amount of Vehicles in Junctions
Month and Year are not correlated positively.
Vehicles and Hour have a similar association to Years and Hour.

➢ *Cross-Correlation*

This is used to verify the reliance of traffic travels through multiple junction-connected links at various time increments. In figure 11, the cross-correlation results are displayed. Investigating the cross-correlation of linked traffic is required to fully comprehend the traffic flow characteristics that define traffic flow patterns.

- As shown in figure, While the cross correlation of the links utilising their own sluggish versions is nearly zero at any time interval, that of the connections with their own lagged versions depends on time lags.
- The lagging pairs display a linkage with a stronger linear correlation that are closest to the lagged versions, which suggests that the trend is continuing into the next time lag and graduallygoing away in the lags after that.
- Because of the related time lag's substantial trendiness across linkages, there is more cross correlation.
- Given that the traffic flow predominantly adheres to the flow conservation principle ratherthan any one-time lag concern, the joined road connection might have a nonlinear relationship Nevertheless, to the master inflow link.

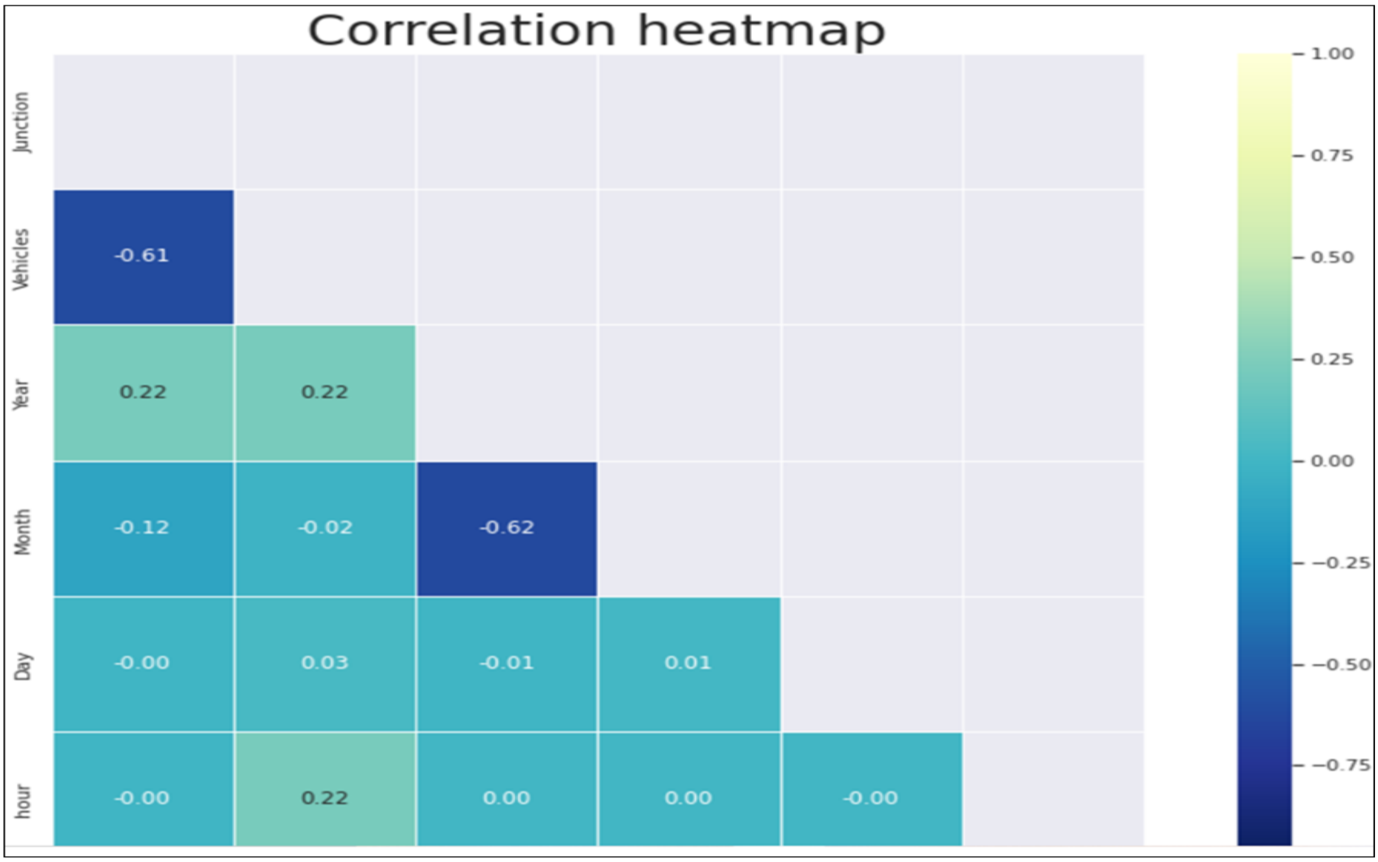


Fig 12 Cross Correlation

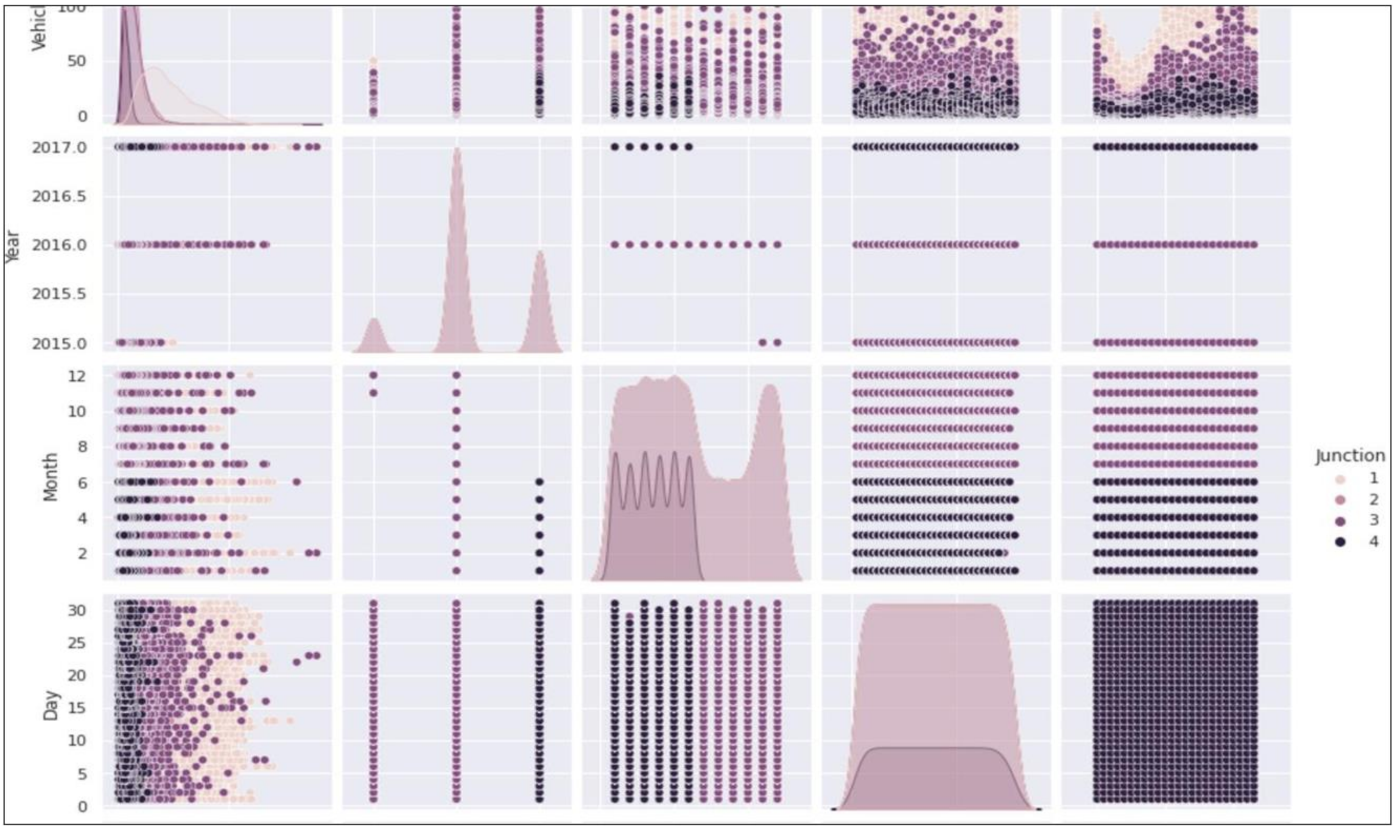


Fig 13 Seasonality Traffic Flows in Year, Month, Day, Week

Each of the four intersections has a different range of data. Just 2017's data are available for the fourth junction. The annual trend for Junctions 1, 2, and 3 has varying slopes. The first junction has a stronger weekly seasonality than the other junctions. For the aforementioned reasons, I believe junctions should be modified to suit each one's specific requirements.

D. Seasonality and Traffic Flow Trends

Most of the time during the day, there is a positive correlation between the link flow pairs. It goes without saying that volume of the traffic rises during peak hours and falls during off-peak hours when the daylight is waning. The fact that traffic flow is inversely related to vehicle speeds is also relevant. The average speed is lower than usual under bad weather and low visibility conditions, which reduces traffic flows. These evolving elements have a significant impact on traffic patterns, particularly during times of congestion. Seasonality and patterns may also play a role in traffic congestion. Figure 5.5 shows that in both plot combinations, Seasonality and patterns may also cause traffic congestion. Figure 5.5 shows that there is a level of flow in both plot pairs at a particular period of time that is highly associated, but the flow is less correlated in the lagged versions of the pair plots. This explains why traffic flow profiles are very depending on the seasons. Also, the gap between peak and regular hours is less of a difference during winter months with reduced visibility. The curve of the density distribution shifts of link pairs during the morning peak hours and regular hours of the day, asshown in figure in the trend plots, makes this variation in flow behaviour more obvious. The pattern of traffic may continue. It is clear from Figure 5.6's findings that there is a sizable Thetraffic patterns include seasonal elements. The residual traffic flows that are unaffected by a trendiness and seasonality characteristics of the flow profiles are essentially unchanged for every link. The seasonal plots show a clear seasonal reverse shift, which is also seen in the plots, as expected. demonstrating that the variation in traffic volume from summer to winter isreal. On the other hand, the trend charts show that the volume or flow of traffic does decline for while the winter season changes and the days get shorter.

E. Trend of Seasonality Traffic Flow

The inherent time-dependent characteristics of data are explained by the periodic dependency in traffic flow time series. This temporal arrangement shows that time-dependent consistencies require specific handling. These observations are idealised as consistent in terms of statistical modelling. The time series being stationary is alluded to in time series analysis. Due of the seasonality and trend features traffic flow time series in our situation has. In order to determine whether our applied forecasting algorithms are effective, the stationarity of time series must be examined. The time series' clear stationary test could be a visual plot. The monthly averaged original flow observations plot is displayed in Figure 6.8. The outcome of the ADF statistics test on the flow series is displayed in Figure 5.9. The unit root test is another name for the ADFtest. It reveals how trendy is it? there is in the time series. An auto regressive model is used byADF. The time series is non-stationary if it can be represented by a unit root, which is the null hypothesis. The series remains unchanged, according to a different theory the hypothesis can be verified or disproven by comparison. The ADF test statistic to crucial values and measuringhow far the two result values diverge from one another is significantly lower than each of the test's three crucial threshold values, indicating that is stationary.

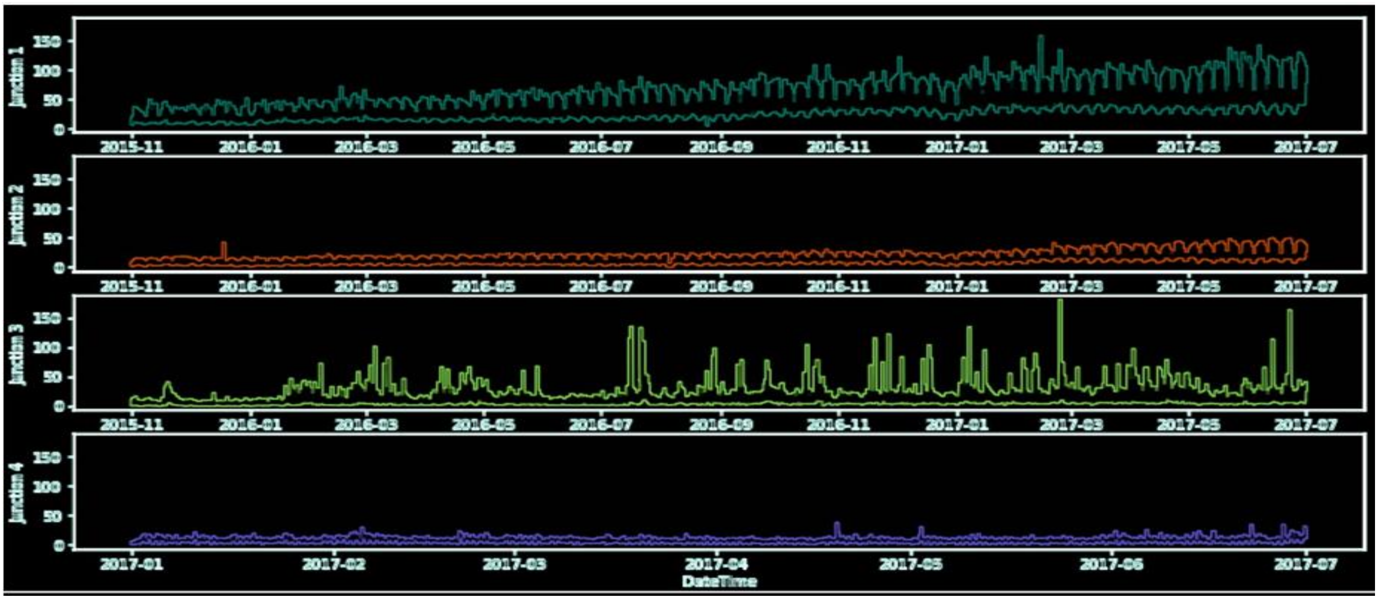


Fig 14 Data Frame before Transformation of 4 Junctions

If a time series lacks a pattern or seasonality, it is said to be stagnant. Nonetheless, we observeda weekly frequency and a rising tendency over time in the EDA.It is once again clear from the graphic above that Junctions one and two are trending upward. We will be able to see the weekly seasonality more clearly if we restrict the span. At this point, I'll skip that step and continue with the appropriate dataset transforms.

In light of the aforementioned observations, the following differencing procedure should be used to remove seasonality:

- I'll be using the difference in weekly numbers for Junction 1.
- At Junction 2, it is better to utilise the difference of consecutive days,
- While Junctions 3 and 4 will use the difference between hourly figures.

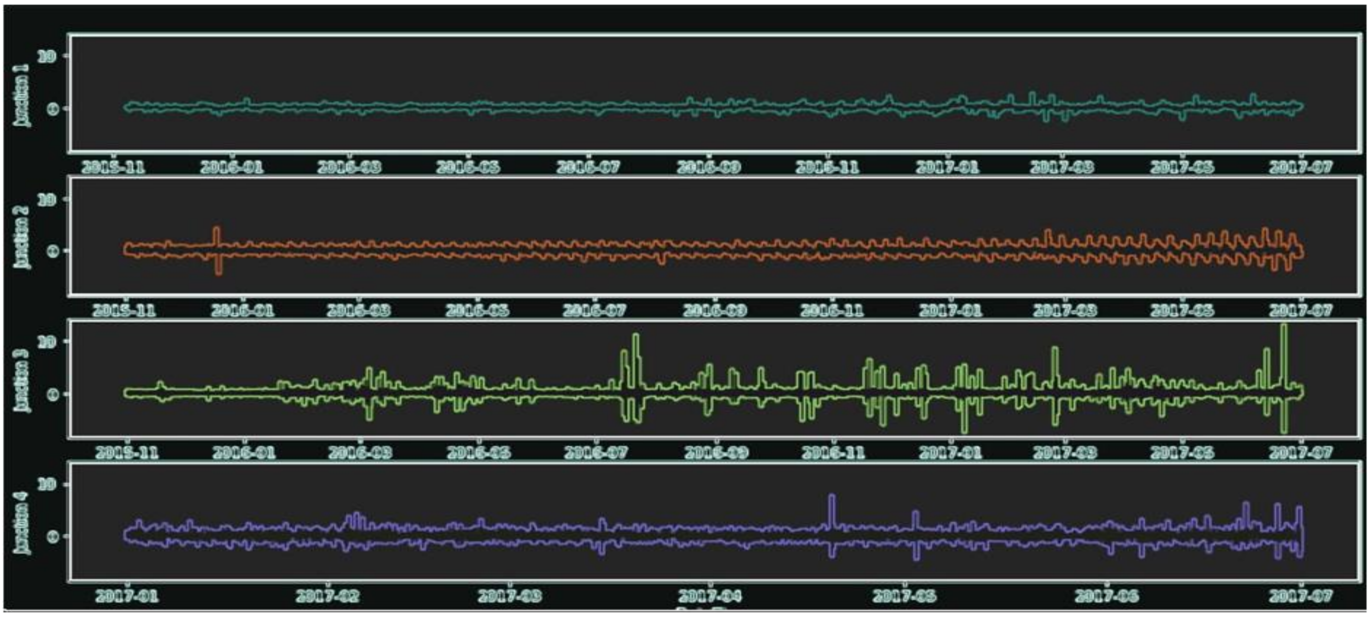


Fig 15 Plots of Transformed Data Frame of 4 Junctions

The aforementioned charts appear to be linear. An Augmented Dickey-Fuller test will be run to make sure they are stationary.

Fig 16 Show Amount of Vehicles by Junction, Each Junction by Day Month

ISSN No:-2456-2165

With the exception of the fourth junction, all junctions have shown a yearly rising tendency. As was already stated above, the fourth junction contains scant data that doesn't go back morethan a year. We can observe that around June, there is an increase in traffic at the first and second crossroads. This, I assume, may be related to summer vacation and related activities. The data is fairly consistent from month to month and across all dates. For a particular day, wemight notice activity peaks in the morning and evening, and a decline during the night. This was expected to happen. Due to fewer automobiles on the roadways on Sundays than other days of the week, traffic flows more smoothly. The traffic is steady from Monday through Friday.

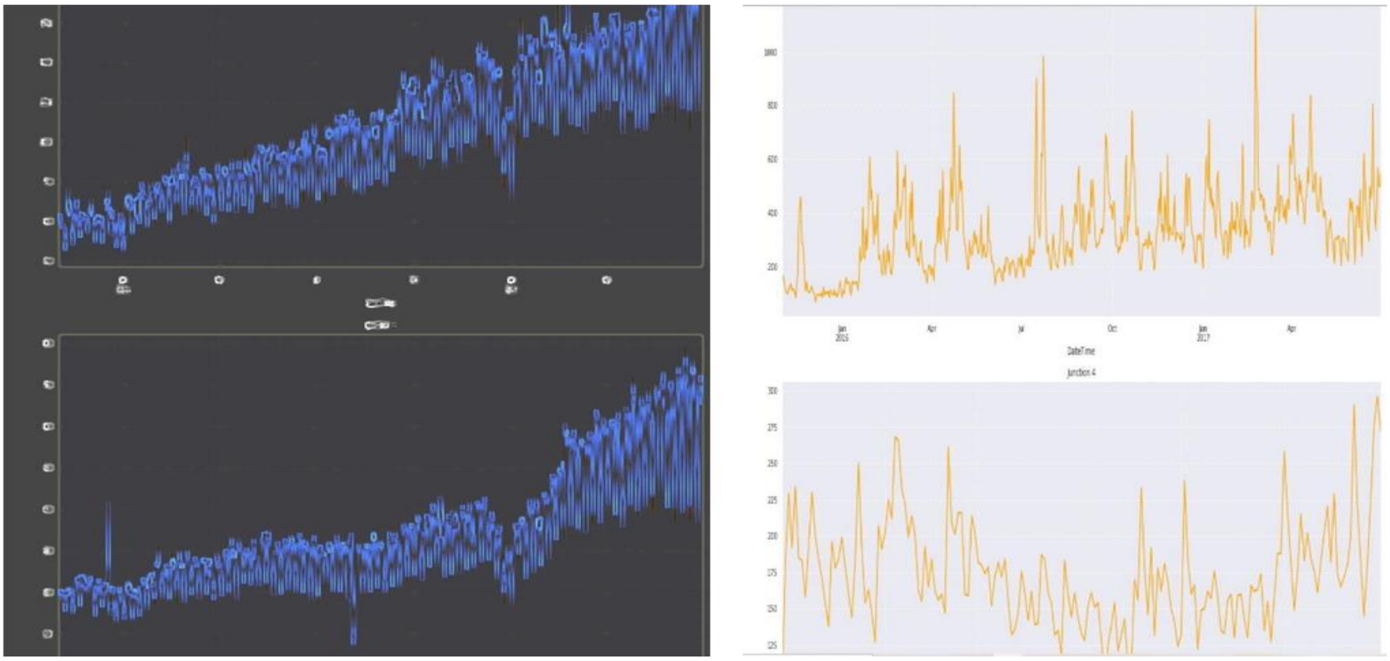


Fig 17 Show Amount of Vehicles by Junction, Each Junction by Day (24h)

If a time series lacks a pattern or seasonality, it is said to be stagnant. Nonetheless, we observeda weekly frequency and a rising tendency over time in the EDA. It is once again clear from thegraphic above that Junctions one and two are trending upward. We will be able to see the weekly seasonality more clearly if we restrict the span. At this point, I'll skip that step and continue with the appropriate dataset transforms.

- Transforming Process:
- Normalizing\differencing

In light of the aforementioned observations, the following differencing procedure should be used to remove seasonality: I'll be using the difference in weekly numbers for Junction 1. Thedifference of consecutive days is a preferable option for junction two.

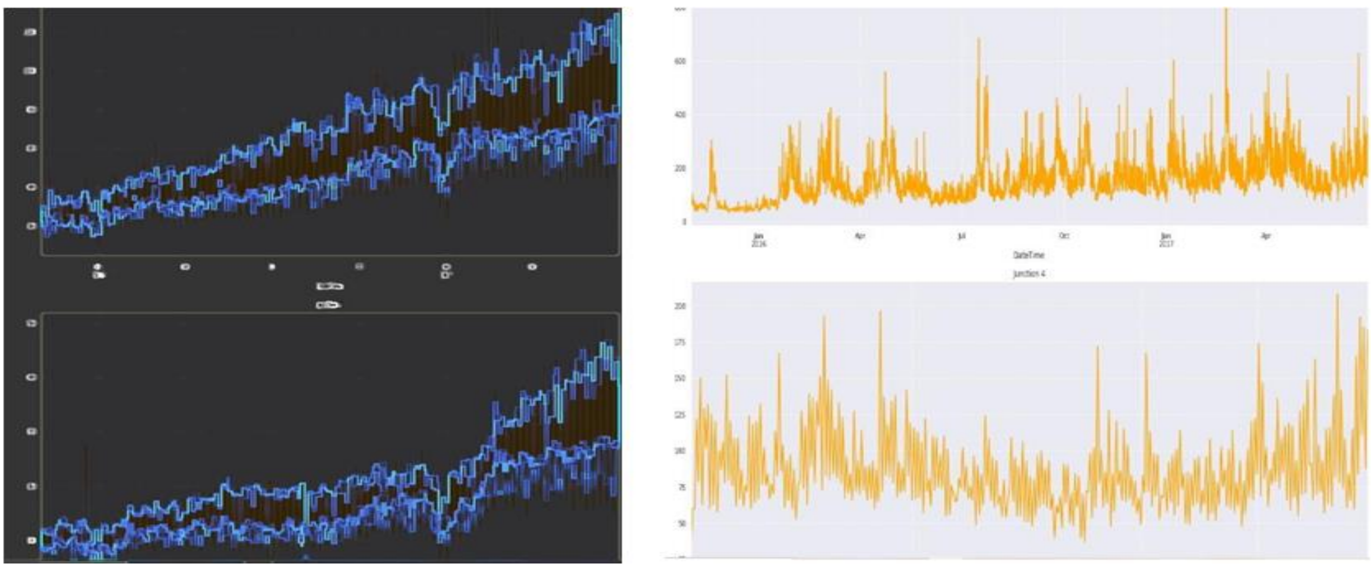


Fig 18 Show Amount of Vehicles by Junction, Each Junction by Day (12h)

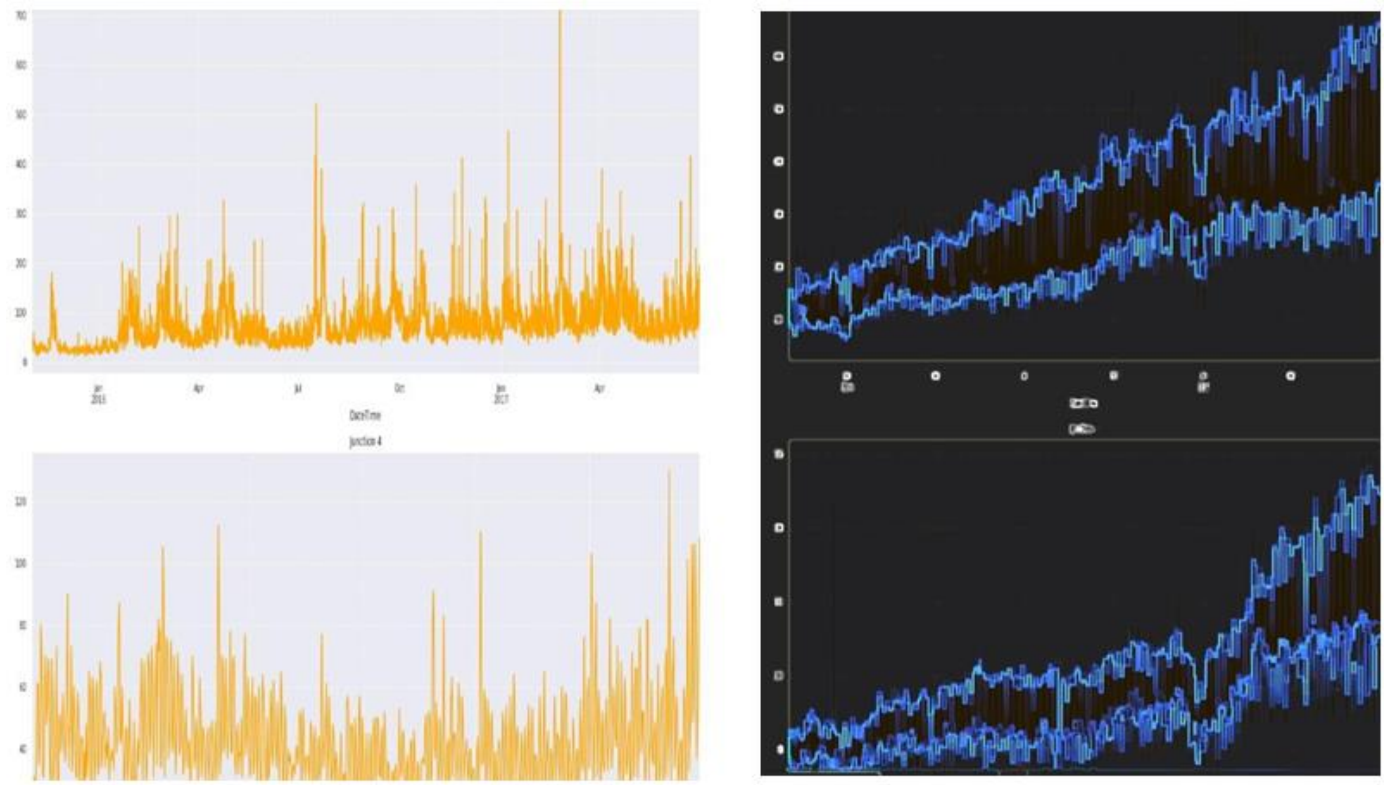


Fig 19 Show Amount of Vehicles by Junction, Each Junction by Day (6h)

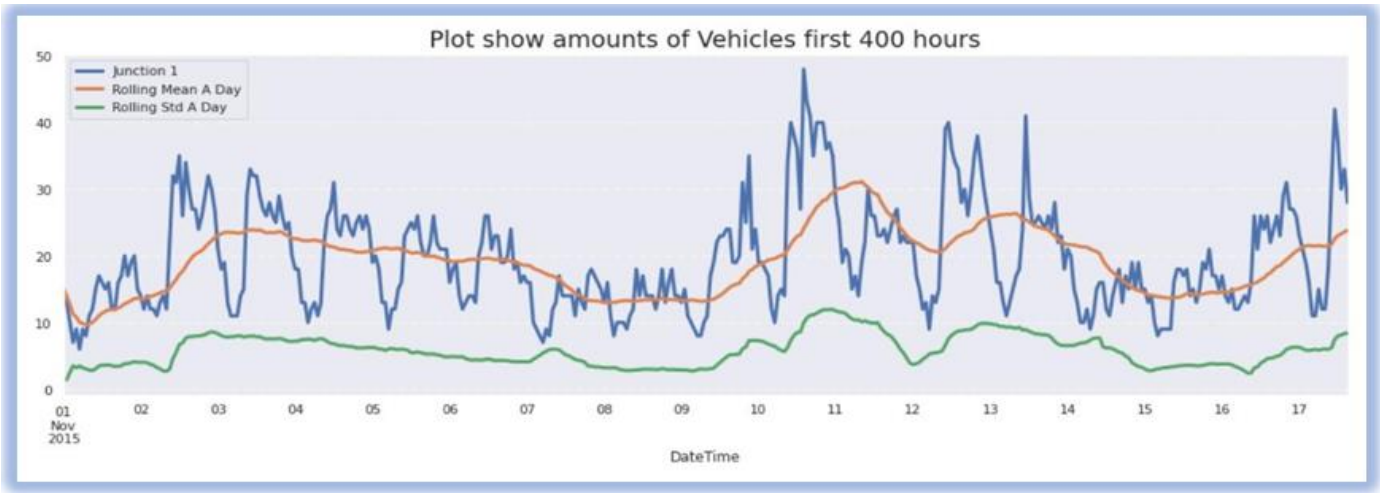


Fig 20 Show Amounts of Vehicles First 400 Hours

F. *Experimental Setting*

In order to facilitate off-campus work with ongoing development, there has been an experimental setup installed on a laptop. This is because some of the most advanced deep machine learning algorithms took longer than anticipated to estimate the best parameters and train ML models using them. Specs for the personal computer: a laptop running Windows 11 Education 64 bit from Hewlett-Packard. 7 cores, 2.40 GHz, 8 processing threads, 16 GB of memory, the user interface provided by DL platforms makes it straightforward to build deep learning architectures using ready-made, optimised libraries or components. Parallelization, cleaner coding, fewer calculations, and automatic gradient computations are some crucial characteristics of a successful deep learning platform. Large corporations like Google, Microsoft, NVidia, and Amazon are making significant investments in the creation of GPU- accelerated deep learning systems for use in large-scale, rapid computing. Of all the currently accessible platforms, Tensor Flow is the most widely utilised and well-liked by users, which iswhy we used it in our study.

G. *Experimentation Findings*

The experimental findings are reported in this section. The absolute error (MAE) and root meansquare error (RMSE), R square values, average and sum is included in the experiment's resultsin this section. The training and test data, as well as the aggregate links data, are used to calculate the MAE and RMSE. We also go over the various prediction cases that are mentionedin section The conclusions section includes a more thorough description of the performance measures referenced in both examples.

ISSN No:-2456-2165

Table 6 Regression Evaluation

	name	r2	rmse
0	LinearRegression	0.627124	0.679264
1	LinearRegression	0.526957	0.244097
2	LinearRegression	0.262216	0.426290
3	LinearRegression	0.180704	0.161093
4	average R2 and sum RMSE	0.399250	1.510744

```
metrics = make_metrics(models, z_data, 'Vehicles', test_size=0.25)
metrics
```

	name	r2	rmse
0	RandomForestRegressor	0.943684	0.268973
1	RandomForestRegressor	0.846931	0.136726
2	RandomForestRegressor	0.745795	0.255119
3	RandomForestRegressor	0.529229	0.118737
4	average R2 and sum RMSE	0.766410	0.779554

Table 7 Creation of Lag Data Regression Evaluation

	name	r2	rmse
0	LinearRegression	0.931846	5.970297
1	LinearRegression	0.861024	2.837878
2	LinearRegression	0.643595	6.207342
3	LinearRegression	0.407108	2.729944
4	average R2 and sum RMSE	0.710893	17.745461

```
metrics_lag_data = make_metrics(models, lag_data, 'Vehicles', test_size=0.25)
metrics_lag_data
```

	name	r2	rmse
0	RandomForestRegressor	0.968623	4.155620
1	RandomForestRegressor	0.887210	2.457379
2	RandomForestRegressor	0.720620	5.505176
3	RandomForestRegressor	0.500814	2.627600
4	average R2 and sum RMSE	0.769317	14.745776

ISSN No:-2456-2165

Table 8 ARIMA Model Evaluation

```
output_metric
```

	name	r2	rmse
0	ARIMA	0.935662	5.713089
1	ARIMA	0.848666	2.869207
2	ARIMA	0.625464	6.345745
3	ARIMA	0.397440	2.828067
4	average R2 and sum RMSE	0.701808	17.756108

```
summary=results.summary()
summary
```

SARIMAX Results

Dep. Variable:	y	No. Observations:	3258
Model:	SARIMAX(2, 1, 0)x(0, 0, [1], 24)	Log Likelihood	-8130.142
Date:	Sat, 14 Jan 2023	AIC	16286.285
Time:	15:02:41	BIC	16365.436
Sample:	0	HQIC	16314.637
	- 3258		
Covariance Type:	opg		

Table 9 SARIMA Model Evaluation

```
        [  1.,  19.,   9., ...,  40.,  41.,  41.],
        [  5.,   7.,  16., ...,  26.,  29.,  30.],
        [ 12.,  17.,  20., ...,  26.,  33.,  34.]]),
 (3647, 8))

X_train,X_train.shape

(array([[ 9., 25.,  2., ..., 47., 49., 46.],
        [ 2., 11., 23., ..., 56., 54., 45.],
        [ 8.,  3.,  7., ..., 23., 30., 37.],
        ...,
        [ 2.,  2.,  4., ..., 31., 34., 35.],
        [ 7., 27., 19., ..., 56., 54., 59.],
        [ 1., 21., 14., ..., 28., 24., 21.]]),
 (10940, 8))

forecast = results.forecast(steps=3647, exog=X_test)

forecast

array([ 69.32763231, 112.41892058,  67.90588468, ...,  53.88722957,
        29.613502  ,  38.18382391])

rmse = calculate_rmse(y_test, forecast)
r2 = calculate_r2_score(y_test, forecast)

rmse

5.882963826163475

r2

0.9351061481986205
```

➢ *Example 1: Test Out Various Prediction Interval*

In this instance, the predictive capabilities of three different ML models are examined. The results of the models are displayed in Table. The medium-term predicting outcomes are shown in table. Test the Effect of Including Related Factors One such variable is the time of day is time of the day that might be employed as a supplementing feature variable in addition to in a multi-variate machine learning model, other flow variables, as was explained in section 5.2. The analysis of the relationship between traffic flow and time of day in section 5.3.3 supports the notion that day of the week and time of day do affect traffic volumes and flows. For better understanding how well the deep learning top end models perform with the added extra features. Some plots have been showed

H. Summary

The investigations conducted of the models evaluations are explained in this chapter. It was thoroughly presented together with the traffic flow data correlation analysis and the comprehensive dataset analysis. The experimental findings for various circumstances are thengiven. The chapter 6 "Evaluation and Conclusion" expands on these findings.

CHAPTER SIX
ANALYSIS AND VERDICT

The experimental evaluation findings from section 5.2 is presented in this chapter. Section 6.1 evaluates the performance outcomes in more detail, and section 6.2 presents more results comments. Section 6.2 presents the study's conclusion at the end.

A. Evaluation

The forecasting outcomes of the carried out experiments are assessed in this section. Every experiment case is assessed. The performance metric is described in Section 5 for the scenario where only the flow variables are taken into account for the short-, medium-, and long-term predictions. While section 6.1.2 discusses in detail the prediction results.

➢ *Case 1: Comparison of the Outcomes of an Experiment with Several Prediction Intervals*

During experimentation, various prediction horizons were taken into account. Fivefold validations are used to compare the findings for each utilising the ECDF plots indicated section

5.1.3. On the test data, ECDF charts are created for comparative analysis. By comparing results from different domains, it is possible to gain more insightful conclusions about the performance of the model as a whole. The heat map-like presentation method used in figure 5.3, which is utilised for each table, represents the value change between the performance tables, and makes it easier to visually discern across performances.

➢ *ML Models Review:*

We test our models on given data per junction as well as on the created lagged data. Following are the models we have tested on:

- Linear Regression
- Random Forrest
- SARIMAX
- Arima

The performance evaluation suggests: Data perform better on Random Forrest Model as compare to Linear and SARIMAX on each junction with accuracy up to 94.3% for 1st Junction on normal data and it tends to increase on lagged data. The average accuracy over all the junction comes out to be 76% on normal data. Linear Regression performs poorly in the given data but tends to improve significantly in lagged data. SARIMAX model also perform better in lagged data with accuracy up to 93%. Also Junction 1 is generally giving better prediction and patterns are well understood by the model and Junction 4 because of lack of data is not performing very poorly on each model and more data would give produce a better prediction. So we can assume Random Forrest to be the best model for the set of data.

➢ *LSTM Traffic Prediction*

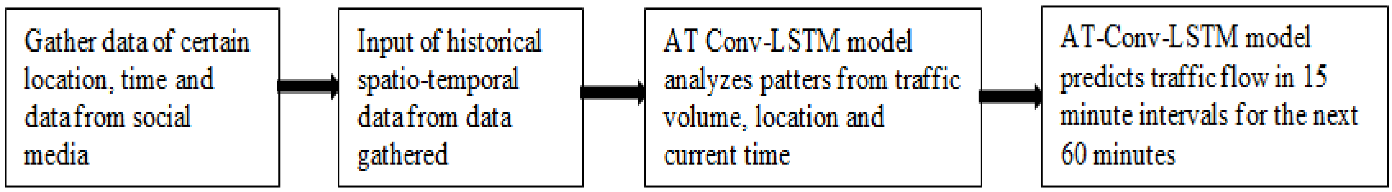


Fig 21 LSTM Traffic Prediction

The model's flow is evident.

- Input from multiple sources include social media data and historical spatiotemporal data for a specific location at various times.
- The AT-Conv-LSTM model examines trends in traffic volume, location, and time relationships.
- The AT-Conv-LSTM model estimates traffic flow every 15 minutes for the following 60 minutes.

ISSN No:-2456-2165

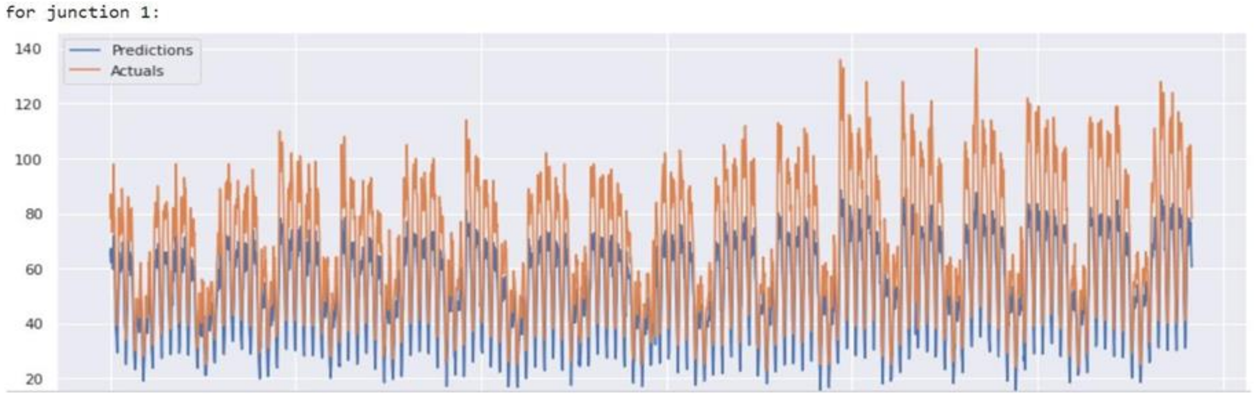


Fig 22 For Junction 1

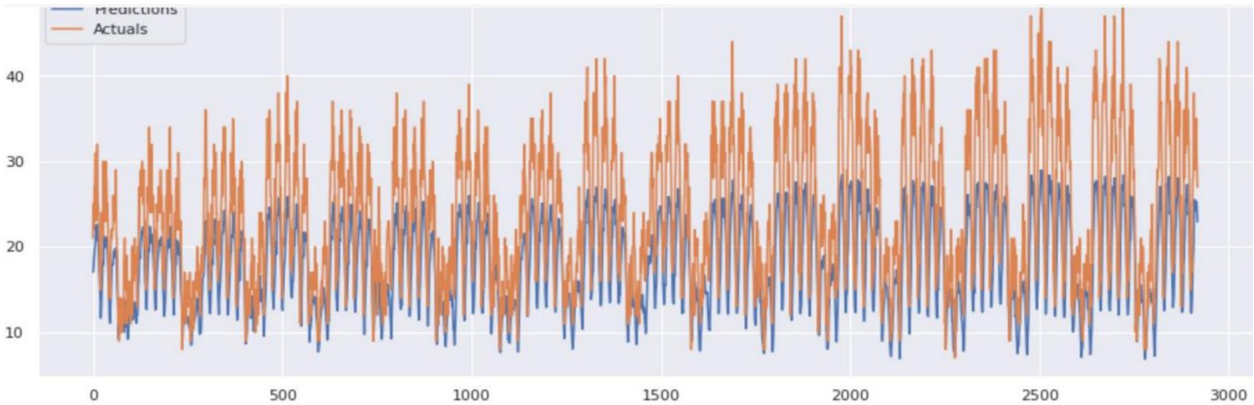


Fig 23 For Junction 2

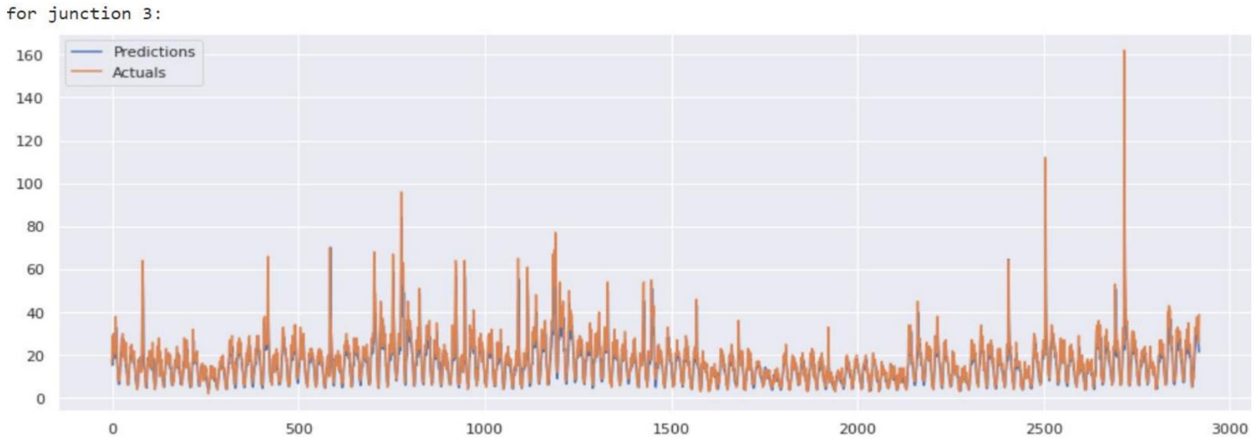


Fig 24 For Junction 3

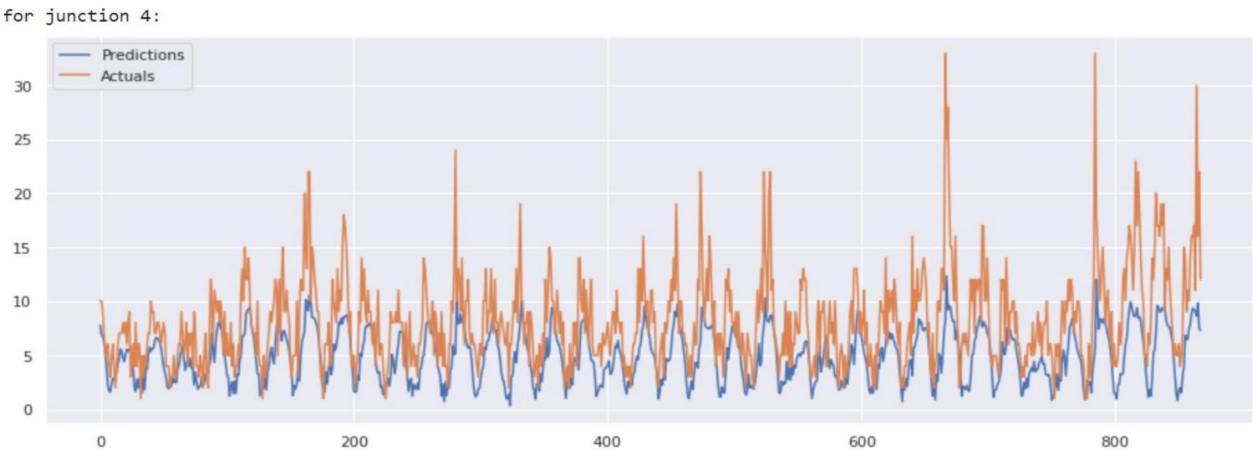
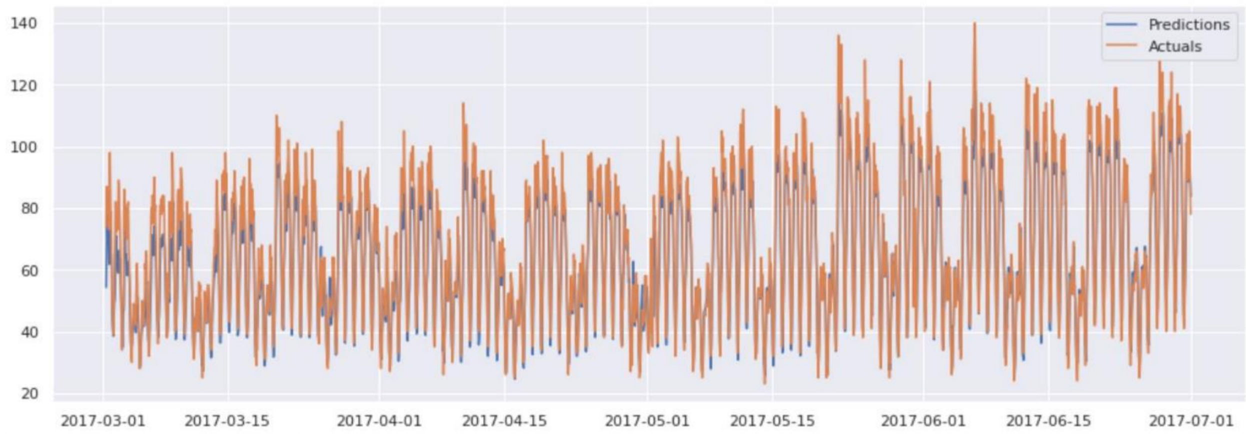


Fig 25 For Junction 4

ISSN No:-2456-2165

- *Final Prediction for Testing Data after DataTransformation*

```
plot_predictions(y_test, y_pred)
```




```
(                    Predictions  Actuals
```

Fig 26 LSTM Predictions for All Junctions

```
r2 = calculate_r2_score(y_test, y_pred)
r2
```

```
0.877141919678073
```

➢ *GRU Traffic Prediction*

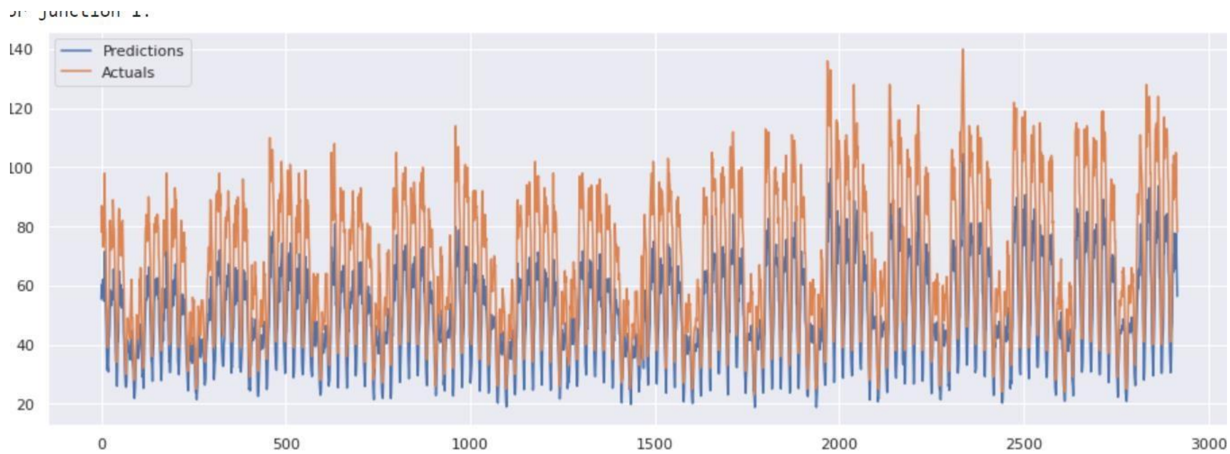


Fig 27 For Junction 1

```
for junction 2:
```

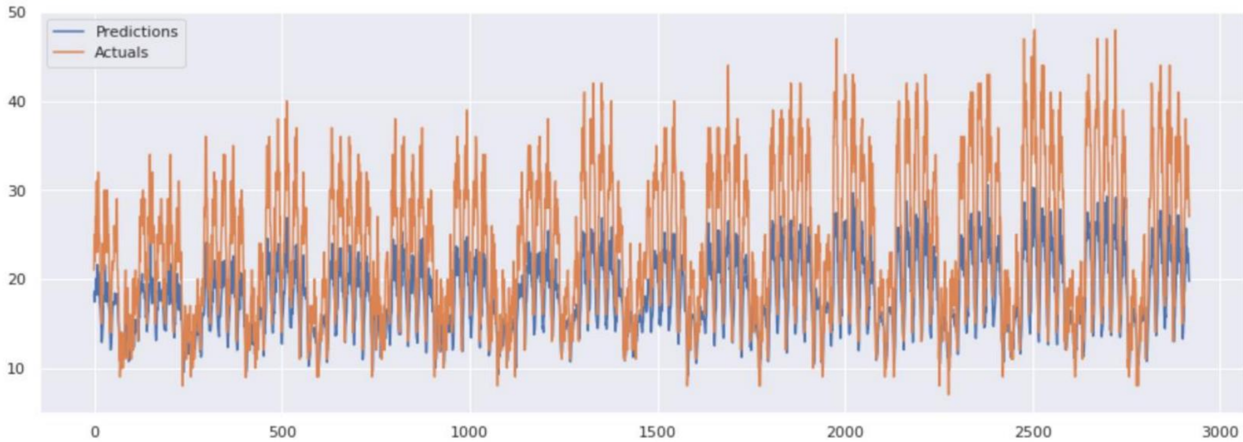


Fig 28 For Junction 2

ISSN No:-2456-2165

for junction 3:

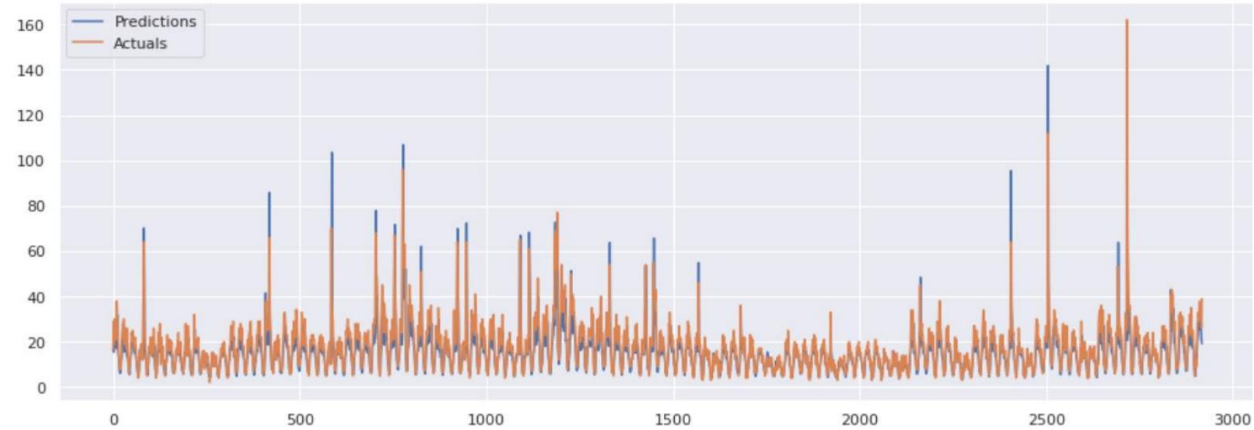


Fig 29 For Junction 3

for junction 4:

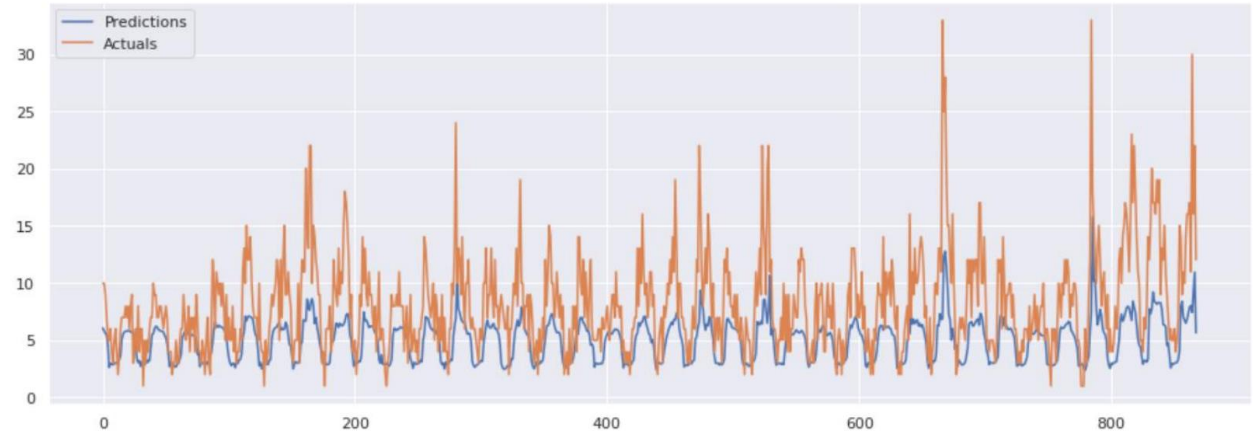


Fig 30 For Junction 4

- *Final Prediction for Testing Data after Data Transformation*

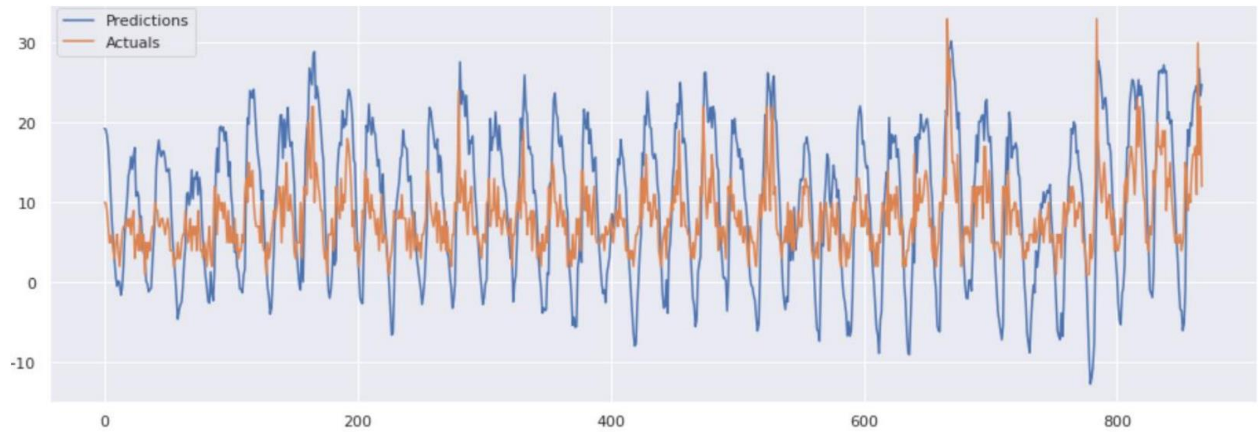


Fig 31 GRU Prediction for All Junctions

```
[ ]  r2 = calculate_r2_score(y_test, y_pred)
     r2

     0.9003486676045338
```

➢ *Deep Learning Models Review*

In this study, a GRU and LSTM neural network was trained to forecast traffic at four intersections, and a normalisation and differencing technique were employed to produce a stationary time series. It requires a different technique for each junction to make it stationary because the junctions vary in trends and seasonality. used the root mean squared error as the model's evaluation metric. Plotted the Predictions along with the first test values in addition. Conclusions drawn from the data analysis Compared to junctions two and three, junction one is seeing a faster increase in the number of vehicles. Junction Four has very little data, therefore could not draw any conclusions from it. The traffic on the Junction One has a stronger hourly and weekly seasonality Compared to other. The Gated Recurrent Units, or GRU, were anotherimportant technique employed in the study to identify traffic flow since it was equally effective with sequential data as the LSTM. High detection rates with minimal false alarm rates are possible with GRU. It was determined that GRU might outperform LSTM in terms of detectionrate while performing similarly in terms of false alarm rates. It also does a great job of capturingboth short- and long-term dependence.

B. Limitations

➢ *Common Pre-Processing Assumptions*

It is important to note some of this thesis' restrictions and restrictions. First, the original raw dataset collection period lasted for eight months. Many complex individual linkages data had to be prepared for additional pre-processing by manipulating the raw dataset using presumptions. While using the fundamental principles of fluid flow, the flow data aggregations for parallel highway road links and link subtraction for anti-parallel flows were two of the assumptions used.

➢ *Lack of Access to Common ITS Data Across Writings*

Second, the seasonal variations were not present throughout the entire year in the training and test data. The trained model primarily saw data from the spring, summer, and early winter seasons. This did not generalise the traffic's annual pattern as a whole. The dataset could be generalised more in upcoming efforts. Regrettably, none of the datasets indicated in the literature research were examined using the models described in this study since the information needed to reproduce the dataset or the experimental setting was unavailable. The existing ITS requires a shared dataset for comparative research components.

➢ *Mechanism for Each Model's Hyper Parameter Tuning*

Even if pre-training was spent looking for the best hyper parameter. But, given the time limits,a far more thorough search can be done by expanding the domains covered by the hyper parameter search inputs. Appendix A lists the hyper parameters for each model that perform the best. Hyper parameters based on the model's top results were used to train the final model.

C. Conclusions

This section responds to the research queries posed in section 1.2:

➢ *RQ1: What potential difficulties can prevent the effective application of traffic roadparameter in forecasting systems?*

➢ *RQ2: What are the most recent machine learning architectures of traffic prediction forforecasting traffic flow, and what impact does the suggested approach have on the performanceof the selected model?*

The thorough literature study in sections 2.3 and 2.4 provides answers to research questions one and two (RQ1 & RQ2). Several prediction models took into account distinct traffic datasets, which was one of the primary difficulties revealed by the research review, and there were no shared datasets throughout the literature. In a perfect world, the merits of the model performance would be assessed using shared datasets. As a result, traffic networks data, it is becoming more and more difficult to determine which model is the most advanced. Recent developments in deep learning algorithms have allowed them to define specified the new parameters for the state of the art, and the experimental findings in this thesis have demonstrated that this is also true for the thesis's domain. In-depth research has been done on Deep Learning Networks, which developed from several neural network-based forecasting models.

➢ *RQ3: What cutting-edge machine learning architectures are there for anticipating traffic flow?*

Modern deep machine learning approaches are currently being seriously taken into consideration in the field of ITS, according to the most recent literature assessment. In order to address of the data’s dynamic nature, the newly proposed methodologies in this research use aone kind of approach on a traffic road junctions level. This leads to the researchers' proposal of a collection of hybrid data-driven algorithms, most of which revolve on RNNs and ANNs.

- *RQ4: What advantages do deep machine learning techniques have over traditional or machine learning methods when taking traffic flow data into account?*

Deep learning models are a good fit for big data issues because of their added benefits of adaptability and ongoing model training. Deep learning models take the lead where machine learning techniques like linear regression and RFR are constrained, as in this thesis. The majority of ITS research focuses on spatial-temporal transport data. LSTMs handle the learning of temporal data, whereas ANNs or CNNs handle the learning of spatial data for deep learning models is used to provide the two side directional flow function of individual roadways while taking into account the net inflows and outflows. Additionally, utilising statistical and neural based machine learning models, the suggested Optimized objective function is compared for limitations. while taking into account various loss functions and training optimization techniques. Lastly, in order to improve forecast accuracy, we present the machine learning model that best fit the suggested flow goal function. The attributes that are time-dependent in the trials are also assessed using the deep learning models in a different experiment situation. The hyper parameter tuning of each individual model, which took a lot of effort for investigation is what drives deep modular learning models. But this is the secret to using deep learning models to make the best predictions. Without this, shallow ML models may outperform deep learning techniques in terms of effectiveness.

The experiment's findings are, in cases where the data is sufficiently sparse to allow for categorical prediction, as in the case of SVR and RFR, shallow machine learning techniques can be used. However, in cases where the data is not sufficiently sparse to allow for categorical prediction, deep learning techniques, such as GRU and LSTM, must be used because they outperformed the former in highly correlated sparse data conditions.

D. Contributions

In this thesis, there are two significant contributions made.

- *Comprehensive Review of Prediction Model Literature*

The literature study is given in a fairly organised manner, covering everything from some machine learning and deep learning models to in-depth discussions of well-liked statistical forecasting approaches. This study examined Linear regression, Random forest, Arima , Sarimax, Linear Regression , Random forest Regression and RNN GRU , LSTM models for ITS traffic flow forecasts using real data from the UK with Regression models, Random forestand RNN based models.

- *Suggested Flow Forecast Objective Function at Junction Level*

Topological junction-based modular road traffic network is the second contribution breakdown that can be used by models to extract spatial-temporal data and forecast traffic flow into and out of road links.

E. Future Works

Create a deep learning model for predicting traffic flow that takes into account the spatial and temporal characteristics of time series data. Because of the inability to adequately handle the spatial temporal aspects of times series data, there is currently no good traffic flow prediction. More research is required to develop a better prediction model that can predict traffic flow more accurately. This thesis demonstrates how traffic flow data from the Highway England can be processed using Machine learning and deep learning-based approaches (HE). While numerous traffic characteristics have been acquired for the MIDAS dataset (see section 4.4.1),They are compatible with the flow features and can be utilised together. The use of new feature vectors enhances deep learning models performances (such weather conditions, average speed etc.) which can then be used to develop more accurate Real-time traffic forecast systems for the general public. According to Appendix C, this would further alter the goal functions and make them more detailed representations of the network path, enabling forecasting of flow and the identification of emerging bottleneck spots and their effects on specific links.

REFERENCES

[1]. J. V. W. L. Anthopoulos, "A United Smart City Model USCM for Smart City Conceptualization and Benchmarking," Int. J. Electron. Gov. Res., Vols. 12, 2, p. 77–93, April2016.
[2]. Berwick, "Smart cities in Europe enabling innovation," borne Clarke International, 2014.
[3]. J. F. J. Alma M, "Introduction to intelligent transportation systems," cham, p. pp 1–17,2016.
[4]. H. G. D.A.C. Kevin Doper" https://www.its.dot.gov/research_archives/cicas/index.htm, Cooperative Intersection Collision Avoidance Systems," Official website, United States Department of Transportation, 2017. "Official website path," 2020.
[5]. M. O. K. Yamamoto T., "Official web site of vehicle information and communication system (VICS), https://trid.trb.org/view/574444," Japan, 2017.
[6]. K. Osama Mohammed, "A Machine Learning Approach to Short-Term Traffic Flow Prediction: A Case Study of Interstate 64 in Missouri," IEEE International Smart Cities Conference (ISC2), 04 March 2019.
[7]. Y. a. Y. L. M. Chen, "Mining moving patterns for predicting next location," Inf. Syst., vol.54, p. 156–168, Dec. 2015
[8]. Z. X. Y. a. S. V. U. X. Zhan, "Citywide traffic volume estimation using trajectory data," IEEE Trans. Know. Data Eng., Vols. 29, no. 2, p. 272–285, Feb. 2017.
[9]. X. L. J. J. J. O. a. L. S. Y. Tan, "A study of best practices in promoting sustainable urbanization in China," J. Environ. Manage.," vol. 193, p. 8–18, May 2017.
[10]. S. F. G. Z. Y. a. B. Y. B. Yu, "k-nearest neighbour model for multiple-time-step predictionof short-term traffic condition," J. Transp. Eng., Vols. 142, no. 6, 2016.
[11]. A. a. E. K. E. Koi, "3D Markov process for traffic flow prediction in real-time Sensors,"Vols. 16, no. 2, 2016.
[12]. J. a. S. Sun, "Neural network multitask learning for traffic flow forecasting," in in Proc. IEEE Int. Joint Conf. Neural Newt. (IJCNN), Jun. 2008.
[13]. Chen, "Research on traffic flow prediction in the big data environment based on the improved RBF neural network," IEEE Trans. Ind. Informant, Vols. 13, no. 4, p. 2000–2008, Aug. 2017
[14]. S. H. H. a. K. X. W. Huang, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," IEEE Trans. Intel. Transp. Syst., Vols. 15, no. 5, p. 2191– 2201, Oct. 2014
[15]. D. W. K. Z. L. a. F.-Y. W. Y. Lb, "Traffic flow prediction with big data: A deep learningapproach," IEEE Trans. Intel. Transp. Syst., Vols. 16, no. 2, p. 865–873, Apr. 2015.
[16]. L. Y.-L. L. a. F. W. Y. Duane, "An efficient realization of deep learning for traffic a imputation," Transp. Res. C, Emerge. Technol., vol. 72, p. 168–181, Nov. 2016.
[17]. G. P. a. V. O. Socolow, "Deep learning for short -term traffic flow prediction vol," Transp.Res. C, Emerge. Technol., vol. 79, p. 1–17, Jun. 2017.
[18]. T. Y. T. K. T. a. T. B. H. Liu, "Visualization of driving behaviour based on hidden feature extraction by using deep learning," IEEE Trans. Intel. Transp. Syst., Vols. vol. 18, no.9, p. 2477–2489, Sep. 2017.
[19]. Z. e. al., "LSTM network: A deep learning approach for short-term traffic forecast," IETIntel. Transp. Syst., Vols. 11, no. 2, p. 68–75, 2017.
[20]. T. Y. W. H. Y. a. Y. W. X. Ma, "Long short-term memory neural network for traffic speedprediction using remote microwave sensor data," Transp. Res. C, Emerge. Technol., vol. 54, p.187_197, May 2015.
[21]. L. Z. C. a. Y. W. R. Key, "Two-stream multi-channel convolutional neural network (TM-CNN) for multilane traffic speed prediction considering traffic volume impact arXiv:1903.01678. [Online]. Available: http://arxiv.org/abs/1903.01678," 2019.
[22]. W. Yunxiang Liu, "Prediction of Road Traffic Congestion Based on Random Forest," 10th International Symposium on Computational Intelligence and Design, pp. 361-364, 2017.
[23]. J. C. F. C. A. C. E. H.-V. Jos A. Moral-Munoz, "Analysing Highly Cited Papers in Intelligent Transportation Systems," IEEE Trans. Intel. Transp. Syst., Vols. Vol.17, no.4, pp. 993-1001, 2016.
[24]. a. R. L. Lunge, "Reconfigurable Controller for Active Fault-tolerant Control Systems withApplicability to Flight Control," in Proceedings of the Romanian Academy, 2014. http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0001:0013:EN:PDF".
[25]. T., "War Rooms of the Street: Surveillance Practices in Transportation ControlCentres," The Communication Review, pp. 367-389, 2007.
[26]. Shin, "Ubiquitous city: Urban technologies, urban infrastructure and urban informatics," Journal of Information Science, vol. 35 (5), pp. 515-526., 2009.
[27]. "The Intelligent Transportations Systems Joint Program Office (ITS JPO),https://www.its.dot.gov/index.htm".
[28]. Mohan, "Nariel: Rich Monitoring of Road and Trac Conditions using MobileSmartphones," 2008.
[29]. A.-Y. A. E. T. J. H. S. H. K Ali, "Crowd ITS: Crowdsourcing in intelligent transportationsystems," IEEE Wire. Common. Newt. Conf. WCNC, p. 3307–3311, 2012
[30]. J. J. A. T. M. J. R. F. a. J. L. M. d. C. L. Figueiredo , "Towards the development of intelligent transportation systems," 2001 IEEE Intel. Transp. Syst. Proc., vol. 81, p. 1206–1211,2001.
[31]. N. a. W. Z. E. Hamada, "Security of Cooperative Intelligent Transport Systems: Standards, Threats Analysis and Cryptographic Countermeasures," Electronics, Vols. 4, no. 3,p. 380–423, 2015.
[32]. S. A. a. A. R. Cook, "Analysis of freeway traffic time-series data by using Box–Jenkinstechniques," Transp. Res., vol.

722, pp. 1–9, 1979

[33]. S. A. a. A. R. Cook, "Analysis of freeway traffic time-series data by using Box Jenkinstechniques," Transp. Res., vol. 722, p. 1–9, 1979.

[34]. B. A. M. M. M. O. Ghosh B., "Bayesian Time-Series Model for Short-Term Traffic Flow Forecasting, "Journal of transportation engineering, p. 180–189., 2007.

[35]. Y. J. Octane I., "Dynamic prediction of traffic volume through Kalan filtering theory,"Transportation Research, p. 1–11, 1984.

[36]. Y. S. J. M. K. L. D. Castro-Neto M., "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," Expert Systems with Applications, p. 6164–6173,2009.

[37]. M. J. Smith B. L., "Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches,". In: Systems, Man, and Cybernetics, Information and Technology'., IEEE International Conference, vol. 2, p. 1706–1709., 1994.

[38]. H. L. W. H. Qipao F., "Intelligent simulation and prediction of traffic flow dispersion," Transportation Research, p. 843–863, 2001.

[39]. L. a. Y.-D. Tao, "On forecasting freeway occupancies and volumes," Transp. Res. Rec., vol. 773, pp. 47–49, 1980.

[40]. W. Fan and R. B. Machemehl, *Characterizing Bus Transit Passenger Waiting Times*, vol.SWUTC/99/1, no. 1. 1999.

[41]. R. Fernández, "Modelling public transport stops by microscopic simulation," *Transp. Res.Part C Emerg. Technol.*, vol. 18, no. 6, pp. 856–868, 2010.

[42]. National Research Council (U.S.) *et al.*, "Guidelines for the design and location of Bus Stops," *Transit Coop. Res. Progr.*, 1994.

[43]. J. B. D.B.Hess, "Waitingfor the bus," *J. Public Transp.*, vol. 7, no. 4, pp. 67–84, 2004.

[44]. P. G. Furth and T. H. J. Muller, "Service Reliability and Hidden Waiting Time: Insights from Automatic Vehicle Location Data," *Transp. Res. Board*, vol. 1955, 2006.

[45]. F. McLeod, "Estimating bus passenger waiting times from incomplete bus arrivals data," *Oper. Res. Soc.*, vol. 58, no. 11, pp. 1518–1525, 2007.

[46]. N. E. Myridis, "Probability, Random Processes, and Statistical Analysis, by H. Kobayashi, B.L. Mark and W. Turin," *Contemp. Phys.*, vol. 53, no. 6, pp. 533–534, Nov. 2012.

[47]. O. C. Ibe, O. A. Isijola, O. A. Isijola-Adakeja, and O. C. Ibe, "M/M/1 multiple vacation queueing systems with differentiated vacations and vacation interruptions," *IEEE Access*, vol.2, pp. 1384–1395, 2014.

[48]. G. Xin and W. Wang, "Model Passengers' Travel Time for Conventional Bus Stop," *J. Appl. Math.*, vol. 2014, pp. 1–9, Apr. 2014.

[49]. D. A. Wu and H. Takagi, "M/G/1 queue with multiple working vacations," *Perform. Evil.*,vol. 63, no. 7, pp. 654–681, Jul. 2006.

[50]. H. Yu, Z. Wu, D. Chen, and X. Ma, "Probabilistic Prediction of Bus Headway Using Relevance Vector Machine Regression," *IEEE Trans. Intel. Transp. Syst.*, vol. 18, no. 7, pp. 1772–1781, Jul. 2017.

[51]. Z. Yu, J. S. Wood, and V. V. Gayah, "Using survival models to estimate bus travel times and associated uncertainties," *Transp. Res. Part C Emerg. Technol.*, vol. 74, pp. 366–382, 2017.

[52]. H. Yu, D. Chen, Z. Wu, X. Ma, and Y. Wang, "Headway-based bus bunching prediction using transit smart card data," *Transp. Res. Part C Emerg. Technol.*, vol. 72, pp. 45–59, 2016.

[53]. B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, "Bus travel time prediction using a time-space discretization approach," *Transp. Res. Part C Emerg. Technol.*, vol. 79, pp. 308– 332, 2017.

[54]. M. Meng, A. Rau, and H. Mahardhika, "Public transport travel time perception: Effects of socioeconomic characteristics, trip characteristics and facility usage," *Transp. Res. Part APolicy Pact.*, no. xxxx , pp. 0–1, 2018.

[55]. A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Inf. Syst.*, vol. 64, pp. 266–280,2017.

[56]. A. Comi, A. Nuzzolo, S. Brinchi, and R. Verghini, "Bus travel time variability: Some experimental evidences," *Transp. Res. Procedia*, vol. 27, pp. 101–108, 2017.

[57]. Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparsetrajectories," *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '14*, no. 5,pp. 25–34, 2014.

[58]. B. Yang, C. Guo, and C. S. Jensen, "Travel cost inference from sparse, spatio-temporally correlated time series using markov models," *Proc. VLDB Endow.*, vol. 6, no. 9, pp. 769–780,2013.

[59]. T. Liebig, N. Piatkowski, C. Bockermann, and K. Morik, "Dynamic route planning with real-time traffic predictions," *Inf. Syst.*, vol. 64, pp. 258–265, 2017.

[60]. L. Gasparyan, E. Bullet, F. Calabrese, O. Verscheure, B. O'Brien, and M. O'Donnell, "System and analytics for continuously assessing transport systems from sparse and noisy observations: Case study in Dublin," *IEEE Conf. Intel. Transp. Syst. Proceedings, ITSC*, no. April 2015, pp. 1827–1832, 2011.

[61]. B. Sun *et al.*, "An improved k-nearest neighbours method for traffic time series imputation," ©*IEEE CAC 2017*, vol. 10, no. October, pp. 7346–7351, 2017.

[62]. M. Moniruzzaman, H. Mao, and W. Anderson, "Short-term prediction of border crossing time and traffic volume for commercial trucks: A case study for the Ambassador Bridge," *Transp. Res. Part C Emerg. Technol.*, vol. 63, pp. 182–194, 2016.

[63]. Y. Duane *et al.*, "An efficient realization of deep learning for traffic data imputation," *Transp. Res. Part C Emerg. Technol.*, vol. 72, no. 10, pp. 168–181, 2016.
[64]. O. D. Cardozo, J. C. García-Palomares, and J. Gutiérrez, "Application of geographically weighted regression to the direct forecasting of transit ridership at station-level," *Appl. Georg.*,vol. 34, no. 4, pp. 548–558, 2012.
[65]. Q. V. Le I.lSutskever, OV.inyals, "Sequence to Sequence Learning with Neural Networks," in *Neural Information Processing Systems Conference*, 2016, pp. 1–9.
[66]. L. Deng and N. Jointly, "Deep Discriminative and Generative Models for PatternRecognition," pp. 1–26, 2015.
[67]. G. B. Zhou, J. Wu, C. L. Zhang, and Z. H. Zhou, "Minimal gated unit for recurrentneural networks," *Int. J. Auto. Compute.*, vol. 13, no. 3, pp. 226–234, 2016.
[68]. V. Sze, Y. H. Chen, T. J. Yang, and J. S. Emer, "Efficient Processing of Deep NeuralNetworks: A Tutorial and Survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.
[69]. K. Yin, W. Wang, X. Bruce Wang, and T. M. Adams, "Link travel time inference using entry/exit information of trips on a network," *Transp. Res. Part B Methodology.*, vol. 80, pp. 303–321, 2015.
[70]. F. N. Saves, "Forecast Comparison of Models Based on SARIMA and the Kalan Filterfor in action," 2013.
[71]. P. J. Brock well and R. A. Davis, *Introduction to Time Series and Forecasting, Second Edition Springer Texts in Statistics*. 2003.
[72]. V. N. Vanir, *Statistical learning theory*. 1998.
[73]. Geoffrey E. Hinton, "Training Products of Experts by Minimizing ContrastiveDivergence," London, 2002.
[74]. S. Hoch Reiter and J urgent Schmidhuber, "LONG SHORT-TERM MEMORY," *NeuralCompute.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[75]. T. M. Units, "National Traffic Information Service DATEX II Service," 2018.
[76]. Daft, "Road traffic statistics," pp. 1–13, 2014.
[77]. Highways England, "Highways England – Data.gov.uk – Journey Time and Traffic FlowData April 2015 onwards – User Guide," no. April, pp. 1–14, 2015.
[78]. A. Rahim and S. Ram lingam, "Empirical Formulation of Highway Traffic Flow Prediction Objective Function Based on Network Topology," *Int. J. Adv. Res. Sci. Eng.Technol.*, vol. 5, no. November, 2018.
[79]. D. Zhang and M. R. Kabuki, "Combining Weather Condition Data to Predict Traffic Flow: A GRU Based Deep Learning Approach," in *2017 IEEE 15th Intl Cone on Dependable*

ISSN No:-2456-2165

APPENDIX A:
HYPER PARAMETERS TUNING RESULTS

While fresh observations were later supplied to the model without going through the separate fitting phase for forecasting for different horizons, ARIMA unlike other models, only requiredone grid search (shown in figure A.1).

```
   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
  13     12      15      1     0     0   1.131D-04   3.338D+00
  F =   3.3382735898034337

CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(1,0,1)[24] intercept   : AIC=82355.555, Time=3.11 sec
 ARIMA(0,1,0)(0,0,0)[24] intercept   : AIC=82353.721, Time=0.26 sec
 ARIMA(1,1,0)(1,0,0)[24] intercept   : AIC=79304.141, Time=2.64 sec
 ARIMA(0,1,1)(0,0,1)[24] intercept   : AIC=inf, Time=24.97 sec
 ARIMA(0,1,0)(0,0,0)[24]             : AIC=82351.721, Time=0.11 sec
 ARIMA(1,1,0)(0,0,0)[24] intercept   : AIC=79303.054, Time=0.28 sec
 ARIMA(1,1,0)(0,0,1)[24] intercept   : AIC=79304.139, Time=2.74 sec
 ARIMA(1,1,0)(1,0,1)[24] intercept   : AIC=79306.144, Time=2.94 sec
 ARIMA(2,1,0)(0,0,0)[24] intercept   : AIC=77990.318, Time=0.62 sec
 ARIMA(2,1,0)(1,0,0)[24] intercept   : AIC=77992.119, Time=3.69 sec
 ARIMA(2,1,0)(0,0,1)[24] intercept   : AIC=77992.120, Time=3.53 sec
 ARIMA(2,1,0)(1,0,1)[24] intercept   : AIC=77994.119, Time=4.14 sec
 ARIMA(2,1,1)(0,0,0)[24] intercept   : AIC=inf, Time=12.46 sec
 ARIMA(1,1,1)(0,0,0)[24] intercept   : AIC=inf, Time=8.81 sec
 ARIMA(2,1,0)(0,0,0)[24]             : AIC=77988.318, Time=0.30 sec
 ARIMA(2,1,0)(1,0,0)[24]             : AIC=77990.119, Time=1.45 sec
 ARIMA(2,1,0)(0,0,1)[24]             : AIC=77990.120, Time=1.35 sec
 ARIMA(2,1,0)(1,0,1)[24]             : AIC=77992.119, Time=1.68 sec
 ARIMA(1,1,0)(0,0,0)[24]             : AIC=79301.054, Time=0.17 sec
 ARIMA(2,1,1)(0,0,0)[24]             : AIC=inf, Time=3.29 sec
 ARIMA(1,1,1)(0,0,0)[24]             : AIC=inf, Time=2.50 sec

Best model:  ARIMA(2,1,0)(0,0,0)[24]
Total fit time: 81.087 seconds
```

Fig A1 ARIMA Model Fitting for Searching Grid

Creation of Lag data

```
lag_df = data.copy()
for i in range(1, 6):
    lag_df[f'Vehicles_lag_{i}'] = data.Vehicles.shift(i)
```

```
lag_df
```

	DateTime	Junction	Vehicles	Year	Month	Day	hour	Vehicles_lag_1	Vehicles_lag_2	Vehicles_lag_3	Vehicles_lag_4	Vehicles_lag_5
0	2015-11-01 00:00:00	1	15	2015	11	1	0	NaN	NaN	NaN	NaN	NaN
1	2015-11-01 01:00:00	1	13	2015	11	1	1	15.0	NaN	NaN	NaN	NaN
2	2015-11-01 02:00:00	1	10	2015	11	1	2	13.0	15.0	NaN	NaN	NaN
3	2015-11-01 03:00:00	1	7	2015	11	1	3	10.0	13.0	15.0	NaN	NaN
4	2015-11-01 04:00:00	1	9	2015	11	1	4	7.0	10.0	13.0	15.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
48115	2017-06-30 19:00:00	4	11	2017	6	30	19	17.0	16.0	16.0	14.0	10.0
48116	2017-06-30 20:00:00	4	30	2017	6	30	20	11.0	17.0	16.0	16.0	14.0
48117	2017-06-30 21:00:00	4	16	2017	6	30	21	30.0	11.0	17.0	16.0	16.0
48118	2017-06-30 22:00:00	4	22	2017	6	30	22	16.0	30.0	11.0	17.0	16.0
48119	2017-06-30 23:00:00	4	12	2017	6	30	23	22.0	16.0	30.0	11.0	17.0

ARIMA and SARIMAX

```
pip install pmdarima
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting pmdarima
  Downloading pmdarima-2.0.2-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.manylinux_2_28_x86_64.whl (1.9 MB)
     |████████████████████████████████| 1.9 MB 620 kB/s eta 0:00:01
Requirement already satisfied: scipy>=1.3.2 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (1.7.3)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (0.29.28)
Requirement already satisfied: scikit-learn>=0.22 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (1.0.2)
Requirement already satisfied: joblib>=0.11 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (1.1.0)
Requirement already satisfied: statsmodels>=0.13.2 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (0.13.2)
Requirement already satisfied: urllib3 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (1.26.9)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (61.2.0)
Requirement already satisfied: numpy>=1.21.2 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (1.21.5)
Requirement already satisfied: pandas>=0.19 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pmdarima) (1.4.2)
Requirement already satisfied: python-dateutil>=2.8.1 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pandas>=0.19->pmdarima) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from pandas>=0.19->pmdarima) (2021.3)
Requirement already satisfied: six>=1.5 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from python-dateutil>=2.8.1->pandas>=0.19->pmdarim
a) (1.16.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from scikit-learn>=0.22->pmdarima) (2.
2.0)
Requirement already satisfied: patsy>=0.5.2 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from statsmodels>=0.13.2->pmdarima) (0.5.2)
Requirement already satisfied: packaging>=21.3 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from statsmodels>=0.13.2->pmdarima) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/T7820S1/Anaconda/lib/python3.9/site-packages (from packaging>=21.3->statsmodels>
=0.13.2->pmdarima) (3.0.4)
Installing collected packages: pmdarima
Successfully installed pmdarima-2.0.2
Note: you may need to restart the kernel to use updated packages.
```

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
import pmdarima as pm
```

```
import pmdarima as pm
```

```
#splitting lag data into training and testing
for i in range(1,5):
    Xtrain, Xtest, ytrain, ytest=split_data(lag_data[i], 'Vehicles',test_size=0.25)
```

```
Xtrain, Xtest, ytrain, ytest
```

```
(array([[ 6., 17., 22., ...,  7.,  7.,  8.],
        [ 5., 13., 21., ...,  7.,  6.,  6.],
        [ 5., 15., 21., ...,  8.,  9., 12.],
        ...,
        [ 1., 17., 21., ..., 11., 13., 10.],
        [ 5.,  2.,  1., ...,  7.,  9., 10.],
        [ 3., 18., 20., ...,  7.,  8.,  6.]]),
 array([[ 2.,  7.,  7., ...,  6.,  4.,  6.],
        [ 6., 29.,  1., ..., 22., 17., 23.],
        [ 6., 15., 21., ...,  6.,  9.,  7.],
        ...,
        [ 2.,  9.,  9., ...,  8.,  2.,  3.],
        [ 3., 27., 18., ..., 13.,  8., 10.],
        [ 4.,  8., 17., ...,  6.,  4.,  8.]]),
 array([13,  8,  5, ...,  7,  4, 10]),
 array([ 5, 11,  5, ..., 12, 11,  4]))
```

```
#metrics for evaluation
def model_metrics_out(y_test,forecast):
    rmse = calculate_rmse(y_test, forecast)
    r2 = calculate_r2_score(y_test, forecast)
    return r2,rmse
```

```
#SARIMAX modelling
metrics=[]
for i in range(1,5):
    X_train, X_test, y_train, y_test=split_data(lag_data[i], 'Vehicles',test_size=0.25)
    results = pm.auto_arima(y_train, #data
```

➤ *Time Series Data Reduction*

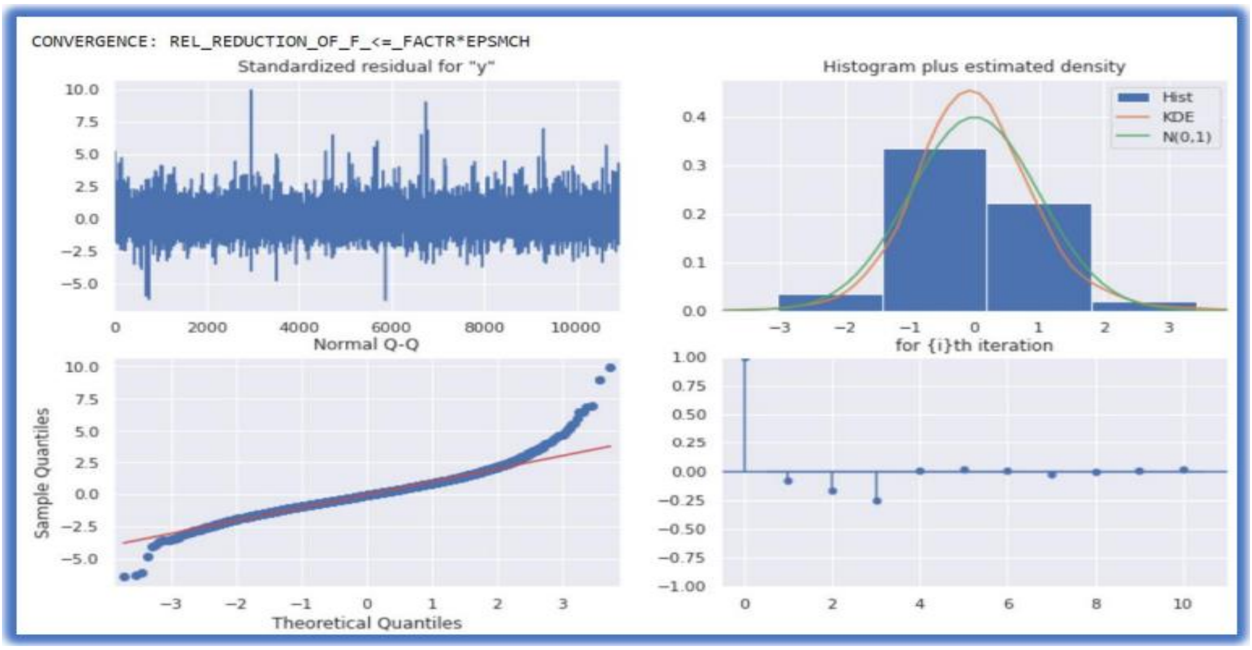
SARIMAX Results

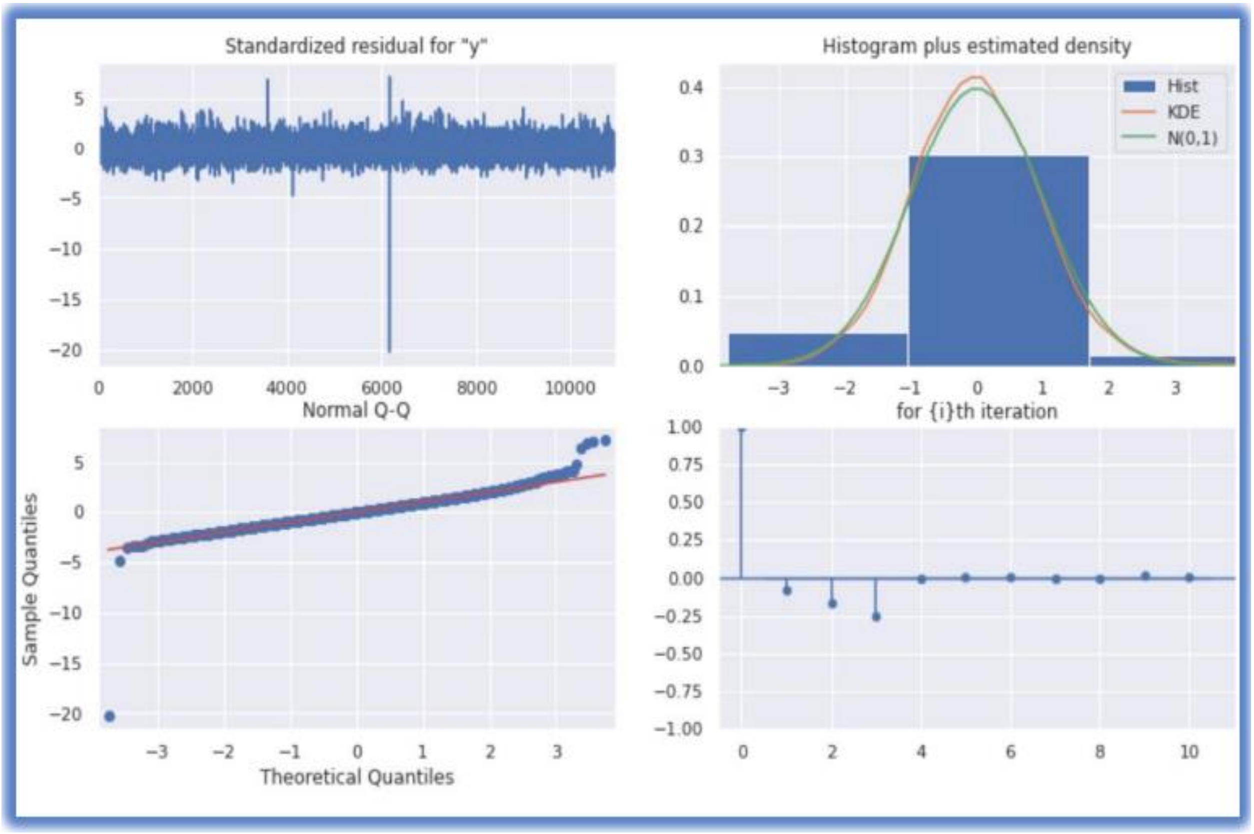
Dep. Variable:	y	**No. Observations:**	3258
Model:	SARIMAX(2, 1, 0)x(0, 0, [1], 24)	**Log Likelihood**	-8130.142
Date:	Sat, 14 Jan 2023	**AIC**	16286.285
Time:	15:02:41	**BIC**	16365.436
Sample:	0	**HQIC**	16314.637
	- 3258		
Covariance Type:	opg		

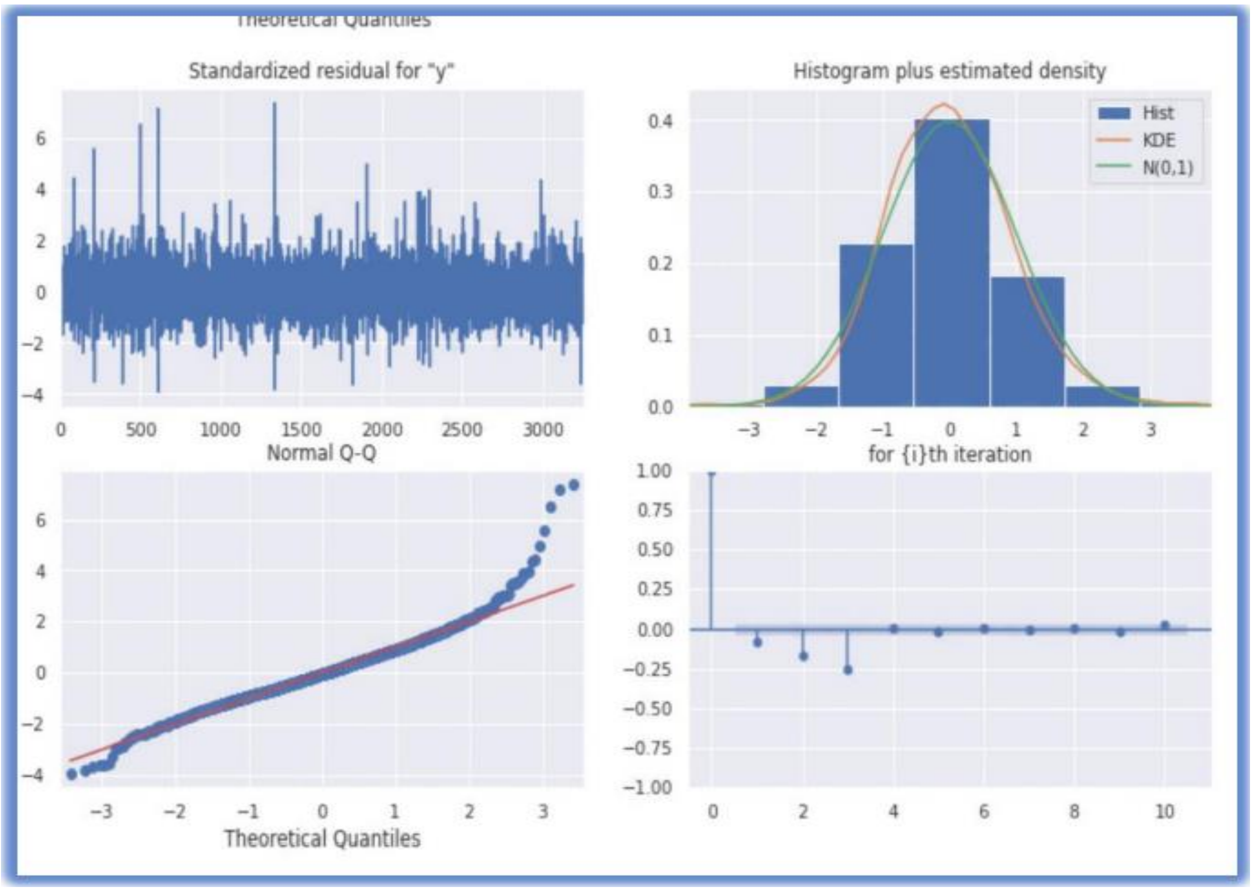
	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
intercept	0.0001	0.054	0.003	0.998	-0.106	0.106
x1	-0.0051	0.027	-0.193	0.847	-0.057	0.047
x2	0.0013	0.005	0.269	0.788	-0.008	0.011
x3	0.0994	0.008	13.171	0.000	0.085	0.114
x4	0.4092	0.016	26.210	0.000	0.379	0.440
x5	0.1832	0.015	12.178	0.000	0.154	0.213
x6	0.0942	0.018	5.240	0.000	0.059	0.129
x7	0.0014	0.017	0.081	0.935	-0.032	0.035
x8	-0.0826	0.016	-5.193	0.000	-0.114	-0.051
ar.L1	-0.6763	0.013	-50.108	0.000	-0.703	-0.650
ar.L2	-0.3226	0.013	-24.204	0.000	-0.349	-0.296
ma.S.L24	-0.0099	0.019	-0.523	0.601	-0.047	0.027

Fig A2 SARIMAX Model for Searching Grid

➤ *Time Series Data Reduction*




Standardized residual for "y"
Histogram plus estimated density
Hist
KDE
N(0,1)
Normal Q-Q
Sample Quantiles
Theoretical Quantiles
for {i}th iteration


Standardized residual for "y"
Histogram plus estimated density
Hist
KDE
N(0,1)
Normal Q-Q
Theoretical Quantiles
for {i}th iteration

ISSN No:-2456-2165

APPENDIX B:
CONTINUATION OF DISCUSSION OF SELECTED MODELS

A. *Long Short-Term Memory (LSTM)*

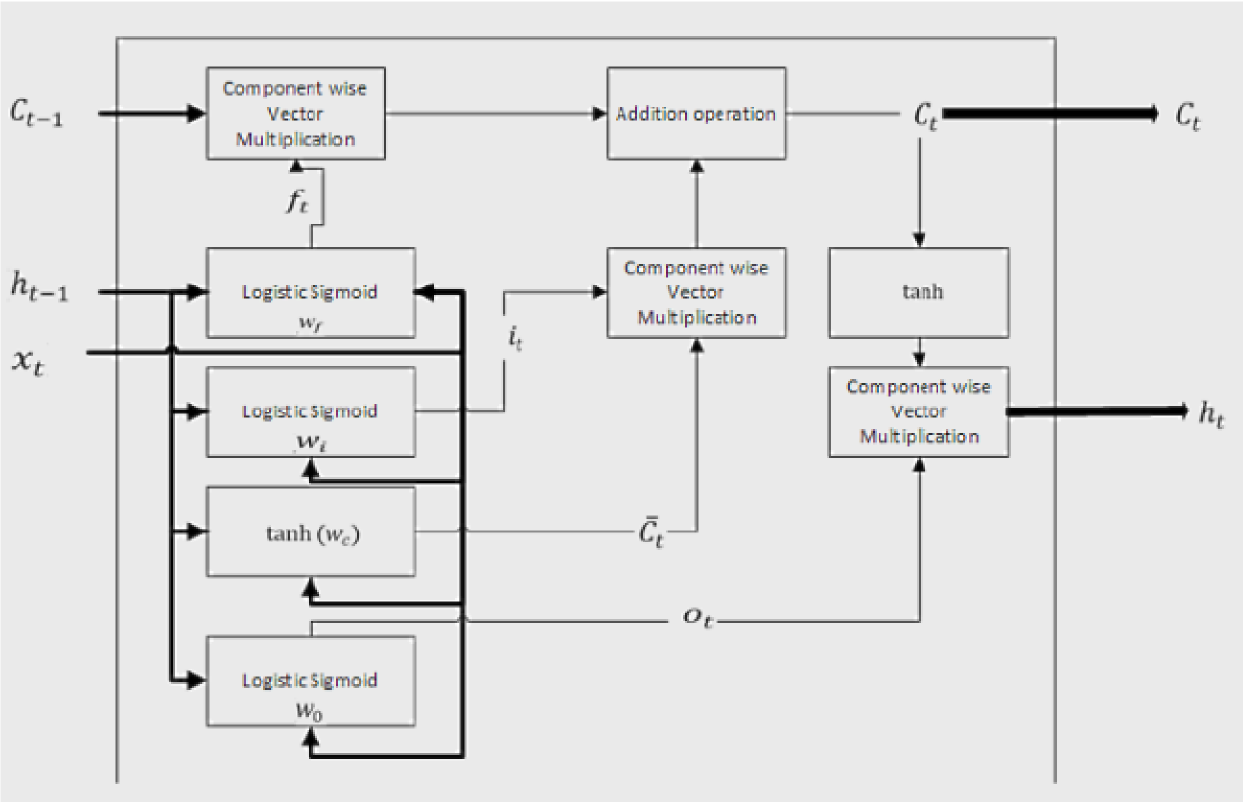
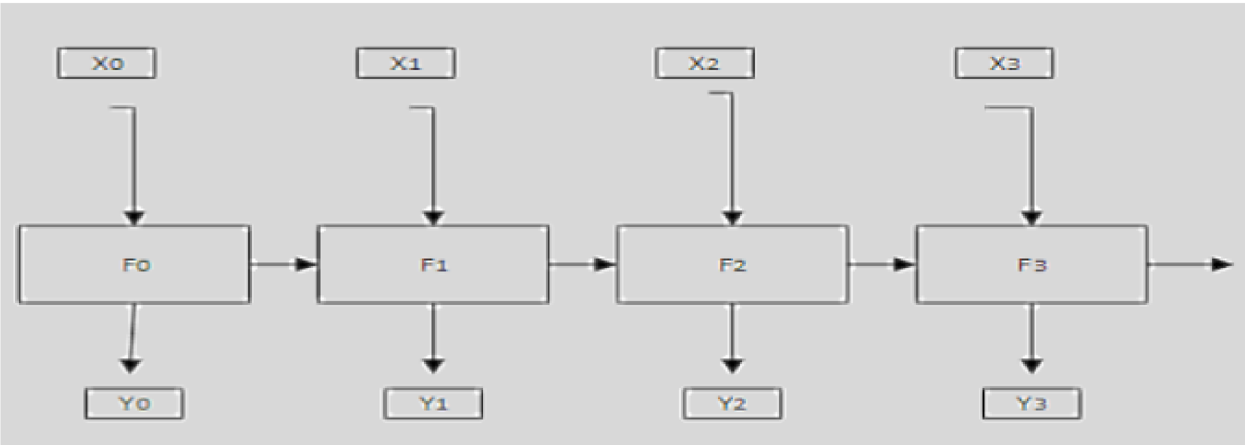


Fig B.1 Shows the Data Flow and Operations in an LSTM unit Structure that Includes the Forget, Input, Output, and Update Gates.

B. *Recurrent Neural Networks vs. Traditional Neural Networks RNN:*

During the learning phase, traditional neural networks treat each input as an isolated piece of data that has no connection to the information that comes after it. If we are to explore a more sequential type of data learning using NN, it is sort of a flaw that the machine learning model must encounter. A standard NN's level of reasoning for the prior event or value is insufficient to anticipate for the occurrences that will follow, leading to lower categorization and predictionaccuracy. All machine learning (ML) algorithms have a cost function that is fixed in purpose and value, but the structurally inherent ability of LSTM-NN and GRU-NN solves this issue bymemorising the cost function.



ISSN No:-2456-2165

Generic Recurrent Neural Network Structure is shown in Figure B.2. In contrast to a feed- forward neural network, each neuron sends its output to the next neuron in the same layer as well as the next neuron in line in the next layer. In order to assess how each neuron will react to new data, the most recent and most recent sources of input are integrated for each neuron.

A trained feed forward network can be trained further for as many data structures for one class,but the essential issue is that it won't necessarily change the classification accuracy for the othertraining data class. This is where a trained feed forward network differs from a recurrent neuralnetwork. RNNs absorb data from a sequence into their own memory (F (n-1)). RNN make decisions that a feedforward network is unable to make by using this historical data.

➢ *Why are GRU and LSTM Being used as Modified RNN?*

While learning long sequences during gradient descent, larger basic RN networks display the gradient expanding phenomenon. This problem is resolved by LSTM and GRU employing various gates to regulate the information flow in RNN. Figure B.1 illustrates the fundamental constructions of an LSTM and GRU. Here, just the mathematically based GRU structure is discussed. However, the ultimate strategy is to evaluate the performance of the based RNN algorithm in our suggested model against the LSTM and GRUs and, if possible, suggest a change in the gated recurrent units. A gated recurrent unit (GRU) is similar to an LSTM in terms of construction and operation, with the exception that it lacks an output gate.

➢ *Node Links Flow Rate Probability Estimations Based on GRU-NN Prediction:*

The next step after developing the GRU-NN model is to train the model using the dataset for individual links flow rate that we already filtered in the Data preliminary analysis (refer section 3.6). The many instances of the GRU-NN model are trained individually on the time series flow rate dataset (refer to table 5.5) for all the linkages with distinct data developed models forthe inflows and outflows, in accordance with the overall suggested model (refer to section 5.7methodology framework). The idea is to identify the bottleneck link at each node while takinginto account the flowrate likelihood probability of all other links on a single node. This is doneby using the flow rate probability prediction for the link whose unidirectional flow rate data is used for the training considering. The link that goes around the bottleneck

➢ *Estimation of the Gaussian Mixture Model Distribution (GMM) Using Past Links Flow Rate Data*

The likelihood of the anticipated flow rate. The probability of the connection 1 is given by P (fi,, t) L1. It is necessary to account for the impact of other linkages within a node by estimatingthe likelihood flow rate probabilities of each instance one at a time. The cause links and the impact connections are two types of Node Links. The Bayesian network is referred to as a Gaussian mixture model (GMM) in which proposes a methodology that is substantially similar. We choose to utilise the GMM model to estimate the flowrate for each individual link using historical data for various time occurrences.

ISSN No:-2456-2165

APPENDIX C:
FUTURE WORKS

Finding the Network Bottleneck for Flow Rate

Future uses of this thesis's work might combine deep learning with statistical methods, including—but not limited to—estimations of each link's flow rate likelihood based onhistorical flow rate data from models using recurrent neural networks with Gaussian Bayes (RNNGB). Every patch has patches with flow rate restrictions. To identify which link patch iscurrently operating as a bottleneck and impeding flow over it, it is important to carefully inspecteach link patch. For instance, a link in one direction may act as a flow limiter at one point, butact as a non-restricting flow link later. The obstruction has been located.