# Traffic Prediction Using GRU Neural Network

Blessy S S
*Department of Computer Applicationsm*
*Amal Jyothi College oEngineering,Koovappally*
*Kottayam, Kerala*
blessyss2022@mca.ajce.in

Mrs. Navyamol K T
*Asst. Proffessor in Computer Science*
*Amal Jyothi College Of Engineering Koovappally*
*Kottayam, Kerala*
navyamolkt@amaljyothi.ac.in

***Abstract*- Traffic is a huge problem that affects every one of us living in cities. The expanded population in cities is a major cause for traffic. Although it can only accommodate a limited population, everyone is moving to cities in search of livelihood and opportunities.**

**Traffic jams cause an increase in fuel ignition. This causes carbon emissions that cause air pollution. The study found that people in America drop an average of 99 hours a year due to traffic congestion. From the previous years, it was found that the average time lost by drivers increased by two hours.**

**In this report, I will explore the dataset of four junctions and build a model to predict their traffic. It may help to understand the traffic patterns and solve the traffic congestion problem, which will further help in building an infrastructure to eliminate the problem.**

***Keywords- Machine learning, GRU Neural network***

## I. INTRODUCTION

Traffic jams are caused by high population density, growth of motor vehicles and expansion of delivery services. Researchers have explained congestion in a variety of ways. The concept of congestion changes when travel get better roads . when a huge volume of vehicles are there in the path other than usual traffic then congestion will occur. Congestion can be described by normal traffic congestion and expanded road users. Many reasons are there for the congestion in city areas. Based on these different causes, congestion can be divided into recurrent and non-recurrent. recurring congestion occur simultaneously, due to the large number of vehicles crowded in uncertain hours. non-recurring congestion occurs due to unforseeable events such as weather, work zones and events.

The communal, lucrative, and environmental cause of traffic congestion in past years was big. Congestion affects the urban transport system well. People in the United States (US) lost a total of $ 160 billion out of an additional 6.9 billion hours in the year 2014. It had to buy 3 billion

## II. MACHINE LEARNING

Machine learning is that the branch of Artificial Intelligence and engineering science that concentrate on the use of knowledge and algorithm. Machine learning is an important building block of arising data science. Through the use of statistics, algorithms are trained to form groups or predictions by chancing major statistics in data processing projects. These intuitions thereafter drive decision-making within operations and businesses, perfectly collide with the vital growth metrics. When big data expands and grows, the request demand for data scientists needs to increase, helping them identify best applicable business questions and also answer the data.

The system of learning in a machine learning algorithm is characterized into three

*a) A Decision process*- algorithms in machine learning are used to make predictions and classifications. Based on some labelled and unlabeled data , the algorithm builds an estimate of a pattern within data.

*b) An Error Function*- it is used to understand the prediction of the model. However, an error function can be used to make a comparison to assess the delicacy of the model, If there are known cases.

*c) Model Optimization Process*- If the model can be matched to the data points in the training set, the weight is commonly used to minimize the mismatch between the known case and the model estimate. The algorithm repeats this evaluation and optimizes the process, automatically streamlining the cargo until a certain level of exactness is achieved.

***DOI: 10.5281/zenodo.6228140***
*ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

gallons of fuel. Traffic congestion is a global malady that challenges feasible and continual development. The aim of this research paper is to contribute to the development of the society and our travel.

*b)Unsupervised Machine Learning:*

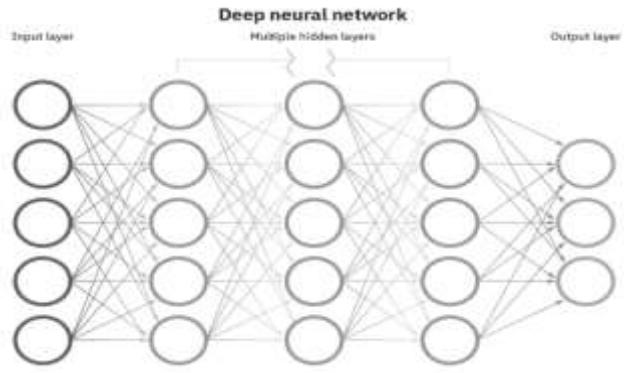
Unsupervised is also known as unsupervised machine learning, which use machine algorithm approaches to analyze and group unlabelled sets. These algorithms find out hidden patterns without the need for help. Its capability to find correspondance and differences in information makes it the best answer for exploratory data analysis, cross-selling strategies, image, and pattern recognition. It is also used to minimize the number of features in a prototype through the process dimensionality reduction. principal factor analysis and singular value decomposition are the two mainly used approaches for this.

c) *Semi-supervised learning:*

This learning is a medium between the other two techniques. it uses a tiny labeled data to conduct classification and have extraction from a bigger , unlabeled dataset. It can solve the matter of getting not enough labeled data(or not having the ability to label enough data)to train a supervised algorithm.

## II. NEURAL NETWORK

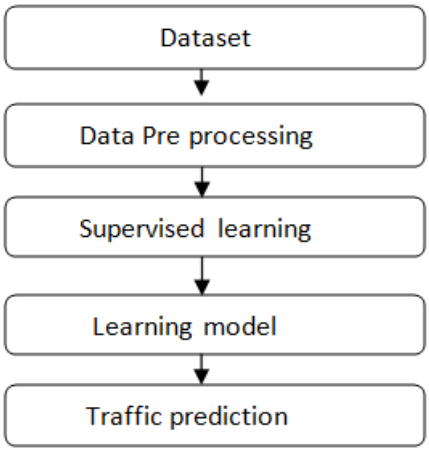
Neural networks are a subspace of machine learning and also the part of deep learning algorithms. The name and structure of the network are like a human brain. Neural networks are made up of a node, input , one or more hidden and an output layer. Each node or artificial neuron joins to a another node and it have an subsequent weight and verge value. If the outcome of any node is above the threshold value, those nodes are activated, then they send the data to the next layer of the network. Else no data will be transffered to the next layer of the network.




Neural networks train the data to make and improve their accuracy eventually. once these learning algorithms are optimized for accuracy, they're powerful tools in computing and AI, allowing us to classify and cluster data at a high velocity. Steps in speech recognition can take time in contrast to manually point out by human experts. One of

## III PROPOSED METHODOLOGY

The purpose of this paper is to explore the dataset of four junctions and build a model to predict traffic on the same. This could possibly help in solving the traffic congestion problem by providing a better understanding of traffic patterns that will further help in building an infrastructure to eliminate the problem. There I used a GRU Neural network to predict the traffic on four junctions. I used a normalization and differencing transform to achieve a stationary time series.




The steps we need to follow with the algorithm:

a) Importing libraries
b) Loading data
c) Data exploration
d) Feature Engineering
e) Exploratory Data Analysis
f) Data transformation and preprocessing
g) Modal Building
h) Fitting the model
i) Inversing the transformation of data

Table 1 : dataset attribute description

| slno | Attributes |
|---|---|
| 1 | DateTime |
| 2 | Junction |
| 3 | Vehicles |
| 4 | ID |

This dataset is a collection of numbers of vehicles at four junctions in an hour. The sensors on each of the four junctions were collecting data at different times, hence the traffic data from different time periods. Some of the junctions have provided limited data.

***DOI: 10.5281/zenodo.6228140***
*ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

the foremost well-known neural networks is Google's search algorithm.

*c) Exploring data*

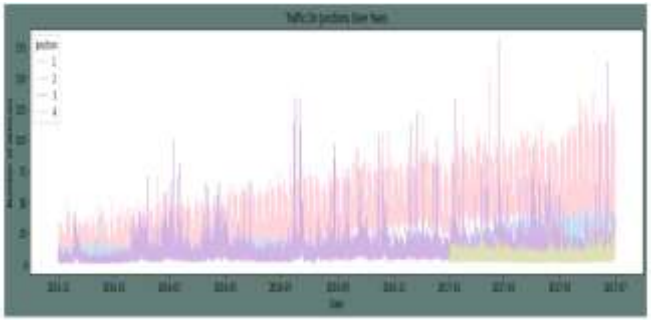


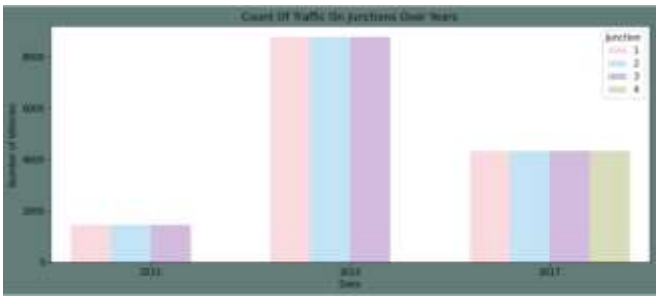
Figure1: Traffic on Junction over years

*d) Feature Engineering*

At this step, I am creating a few new features out of DateTime. Namely:

a) Year
b) Month
c) Date in the given month
d) Days of week
e) Hour

*e) Exploring data analysis*



The count plot shows that there is an increase in the number of vehicles between 2015 and 2016. However, it is inconclusive to say the same about 2017 as we have limited data for 2017 that is till the 7th month.

*f) Data transformation and preprocessing*

here, I followed the subsequent order:

a) Creating different frames for each Junction and plotting them
b) Transforming the series and plotting them
c) executing the Augmented test to examine the period of transformed series
d) Creating test and train sets

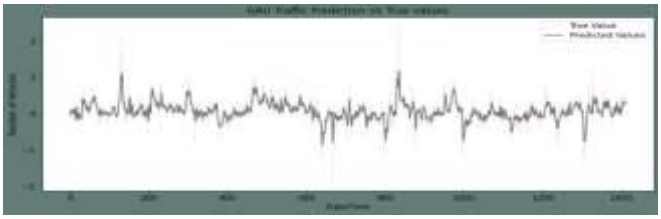


Figure 2: fitting the first junction and plotting the predictions and test data

The root mean squared error is 0.245881146563882.

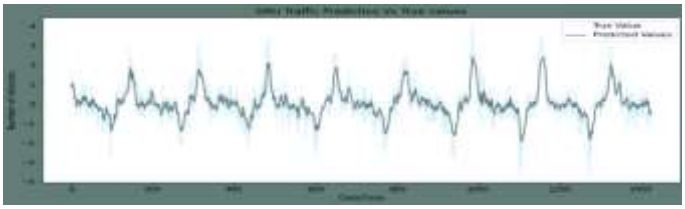


Figure 3- Fitting the second junction and plotting the predictions and test data

The root mean squared error is 0.5585970393765944

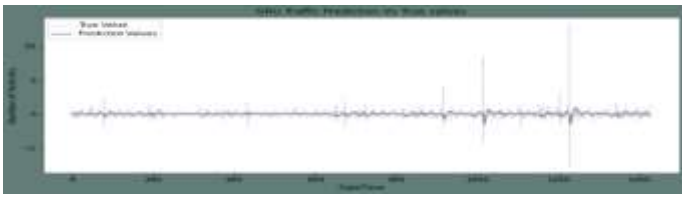


Figure 4: Fitting the third junction and plotting the predictions and test data

The root mean squared error is 0.606136678

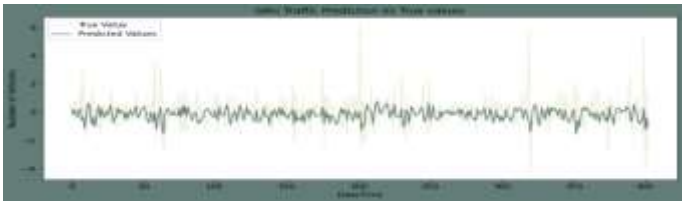


Figure 5: Fitting the fourth junction and plotting the predictions and test data

The root mean squared error is 1.0241982484501175.

***DOI: 10.5281/zenodo.6228140***
*ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

*g) Model Building*

For this paper, I used Gated Recurrent Unit (GRU). In this sectio the data frames for all four junctions.

*h) Fitting the model*

Now, fit the transformed training sets of four junctions to the model created and analyze them to the transformed test sets.

The result of the modal is:

| | Junction | RMSE |
|---|---|---|
| 0 | Junction1 | 0.245881 |
| 1 | Junction2 | 0.558597 |
| 2 | Junction3 | 0.606137 |
| 3 | Junction4 | 1.024198 |

The Root Mean Square Error is an individual marker for measuring the production. Thus, in this project, I am including the outcome plots as well.

*i) Inversing and transforming the data*

In this section, I will be inversing transforms that I applied to the datasets to remove the seasonality and trends. Performing this step will make the predictions get back on the accurate scale.

*Predictions and originals for the junction*

Figure6: the inverse transform on the first junction

Figure7: the inverse transform on the second junction

Figure8: the inverse transform on the third junction

Figure9: the inverse transform on the fourth junction

## V CONCLUSION

In this paper, I trained a GRU Neural network to predict the traffic on four junctions. I used a normalization and differencing transform to achieve a stationary time series. As the Junctions vary in trends and seasonality, I took different approaches for each junction to make it stationary. I applied the root mean squared error as the evaluation metric for the model. In addition to that, I plotted the Predictions with the original test values. Take away from the data analysis:

The Number of vehicles in Junction one is rising more rapidly compared to junction two and three. The sparsity of data in junction four bars me from making any conclusion on the same.

The traffic of junction one has a stronger weekly seasonality as well as hour seasonality. Whereas other junctions are linear

## VI REFERENCES

[1] Anand Paul, Naveen Chilamkurti, Alfred Daniel, Seungmin Rho,Chapter 8 - Big Data collision analysis framework,Editor(s): Anand Paul, Naveen Chilamkurti, Alfred Daniel, Seungmin Rho,Intelligent Vehicular Networks and Communications,Elsevier,

[2] A. R. Abdellah and A. Koucheryavy, "Deep learning with long short-term memory for IoT traffic prediction," Lecture Notes in Computer Science, vol. 12525, pp. 267–280, 2020.View at: Publisher Site | Google Scholar

[3] sciencedirect.com

[4] www.kaggle.com/karnikakapoor/traffic-prediction-gru

[5] Hadi Habibzadeh, Tolga Soyata, Burak Kantarci, Azzedine Boukerche, Cem Kaptan,Sensing, communication and security planes: A new challenge for a smart city system design,Computer Networks,Volume 144

***DOI: 10.5281/zenodo.6228140***
*ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*