**Metro interstate traffic volume**

***A project report***

***Submitted in partial fulfillment of the requirements for the award of the degree of***

## BACHELOR OF TECHNOLOGY

**IN**

**INFORMATIONTECHNOLOGY**

Submitted by

MR. K.SRI SATYA (17A91A1219)

MR.K.S.N.SWAMY (17A91A1223)

MR.P.DURGA PRASAD (17A91A1246)

Under the esteemed guidance of

MR.B.RAJESH



DEPARTMENT OF INFORMATION TECHNOLOGY

ADITYA ENGINEERING COLLEGE

## (An Autonomous Institution)

(Approved by AICTE, Permanently Affiliated to JNTUK & Accredited by NBA, NAAC with ‘A’ Grade Recognized by UGC under the sections 2(f) and 12(B) of the UGC Act 1956

Aditya Nagar, ADB Road, SURAMPALEM-533437)

**2017-2021**

# ADITYA ENGINEERING COLLEGE

## (An Autonomous Institution)

(Approved by AICTE, Permanently Affiliated to JNTUK & Accredited by NBA, NAAC with ‘A’ Grade Recognized by UGC under the sections 2(f) and 12(B) of the UGC Act 1956

Aditya Nagar, ADB Road, SURAMPALEM-533437)



## CERTIFICATE

This is to certify that it is a bonafide record of the Project work entitled **“Metro interstate traffic volume”** is being submitted by

MR. K.SRI SATYA (17A91A1219)

MR.K.S.N.SWAMY (17A91A1223)

MR.P.DURGA PRASAD (17A91A1246)

Project Guide Head of the Department

MR.B.RAJESH Mr. M. Rajababu, M.Tech,(Ph.D).,

Associate professor,

Department of IT.

**DECLARATION :**

We hereby declare that the project work which is being presented in this report entitled “**METRO INTERSTATE TRAFFIC VOLUME**”, is the work done by us in the partial fulfillment of the requirements for award of B.Tech in Information technology & Engineering of Aditya Engineering College under Jawaharlal Nehru Technological University, Kakinada is authentic record of our work carried out under the supervision of **MR.Rajesh** , technical hub trainer. The matter present in this report has not been submitted to any other university/institute for the award of B-Tech degree.

**ACKNOWLEDGEMENT**

We are very much thankful to MR.B.RAJESH,Aditya Engineering College, Surampalem for being a Project guide and for his Constant support and encouragement throughout the project. His wide knowledge and logical way of thinking had made a deep impression on me. He is the source of inspiration for innovative idea and his kind support is well known to all his students and colleagues.

We are very much thankful to Mr. M.RAJABABU, Associate Professor HOD- Dept. of IT, Aditya Engineering College, Surampalem, who has extended his valuable suggestions for the advancement of the project.

We also thank our **Management & Technical supporting staff** for making all necessary things available for us in time.

Project associates

MR.K.SRI SATYA(17A91A1219)

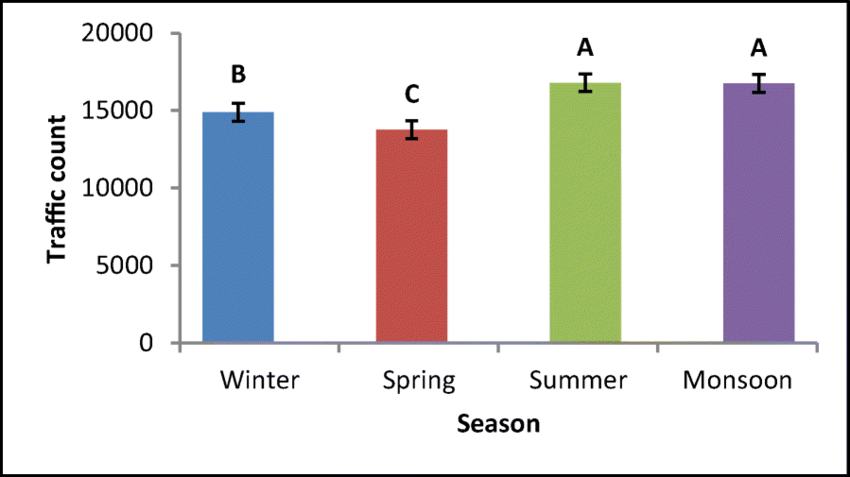
MR.K.S.N.SWAMY(17A91A1223) MR.P.DURGA PRASAD(17A91A1246)

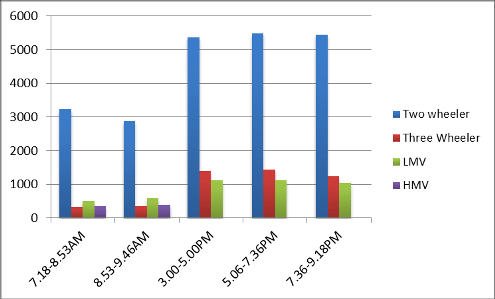
**ABSTRACT**

Traffic volume is the basic cause requiring rapid solution. The greater the [annual average daily traffic](https://www.sciencedirect.com/topics/engineering/annual-average-daily-traffic) (AADT), the greater the number of lanes with major maintenance issues due to wear and tear, leading to [structural deficiency](https://www.sciencedirect.com/topics/engineering/structural-deficiency) and eventual bridge replacement. As discussed in earlier chapters, ABC is helping with the replacement of bridges with high traffic volume. It should be noted that the traffic volume data is the most important parameter in the performance of bridges. If there is no traffic then there will be no need for a bridge. Many states have now obtained daily traffic count records during rush hour and over a 12-month period. Interstate and arterial roads in general carry heavier traffic than the local roads.

Our project” **METRO INTERSTATE TRAFFIC VOLUME”Used to predict whether the traffic is increases or not in the next coming years based on the past dataset.**

Examples of traffic volumes increasing every year:





INTRODUCTION

The term **software engineering** is composed of two words, software and engineering.   
**Software** is more than just a program code. A program is an executable code, which serves some computational purpose. Software is considered to be a collection of executable programming code, associated libraries and documentations. Software, when made for a specific requirement is called software product.

**Engineering** on the other hand, is all about developing products, using well-defined, scientific principles and methods.

So, we can define **software engineering** as an engineering branch associated with the development of software product using well-defined scientific principles, methods and procedures. The outcome of software engineering is an efficient and reliable software product.

As a software engineers we are going to do a project on metro interstate traffic volume using some programming languages i.e, **python , html** etc..

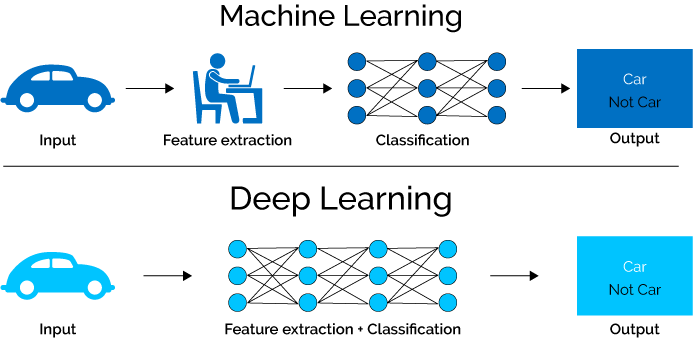
This whole project is based on machine learning and the background programming is **python** front end is flask(html code).

**Machine learning**:

Machine Learning is a sub-area of artificial intelligence, whereby the term refers to the ability of IT systems to independently find solutions to problems by recognizing patterns in databases. In other words: Machine Learning enables IT systems to recognize patterns on the basis of existing algorithms and data sets and to develop adequate solution concepts. Therefore, in Machine Learning, artificial knowledge is generated on the basis of experience.

In order to enable the software to independently generate solutions, the prior action of people is necessary. For example, the required algorithms and data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be defined. Once these two steps have been completed, the system can perform the following tasks by Machine Learning:

* Finding, extracting and summarizing relevant data
* Making predictions based on the analysis data
* Calculating probabilities for specific results
* Adapting to certain developments autonomously
* Optimizing processes based on recognized patterns



Classification of Machine Learning

machine learning can be classified into three types:

1. **Supervised learning**
2. **Unsupervised learning**
3. **Reinforcement learning**

# Supervised Machine Learning:

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.It can be further divided into two types of problems:

**1)Classification**

**2)Regression**

**Classification:**

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

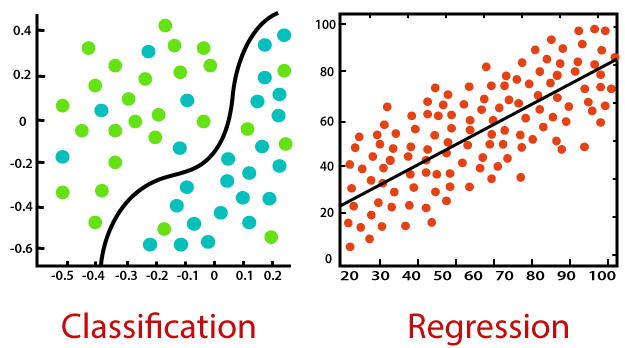
Spam Filtering,

* Random Forest
* Decision Trees
* Logistic Regression
* Support vector Machines

**Regression**

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

* Linear Regression
* Logistic Regression
* Polynomial Regression
* Stepwise Regression
* Ridge Regression
* Lasso Regression
* ElasticNet Regression



1)Linear Regression:

It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be [continuous or discrete](https://en.wikipedia.org/wiki/Continuous_and_discrete_variables), and nature of regression line is linear.

Linear Regression establishes a relationship between **dependent variable (Y)** and one or more **independent variables (X)** using a **best fit straight line** (also known as regression line).

2) Logistic Regression:

Logistic regression is used to find the probability of event=Success and event=Failure. We should use logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature.

3) Polynomial Regression:

A regression equation is a polynomial regression equation if the power of independent variable is more than 1.

## 4) Stepwise Regression:

## This form of regression is used when we deal with multiple independent variables. In this technique, the selection of independent variables is done with the help of an automatic process, which involves no human intervention.

## 5)Ridge Regression:

## Ridge Regression is a technique used when the data suffers from multicollinearity (independent variables are highly correlated). In multicollinearity, even though the least squares estimates (OLS) are unbiased, their variances are large which deviates the observed value far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors.

## 6) Lasso Regression:

## Similar to Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) also penalizes the absolute size of the regression coefficients.

## 7)  ElasticNet Regression: ElasticNet is hybrid of Lasso and Ridge Regression techniques. It is trained with L1 and L2 prior as regularizer. Elastic-net is useful when there are multiple features which are correlated.

## Examples of regression:

## img11.png

## img33.png

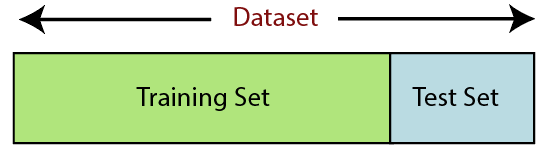
## About DataSet:

When we are doing a project with machine learning the data set is very much important because machine learning algorithms are trained by those data sets only.So that data set set is important

In machine learning data preprocessing, we divide our dataset into a training set and test set. This is one of the crucial steps of data preprocessing as by doing this, we can enhance the performance of our machine learning model.

Suppose, if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models.

If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:



**Training Set:** A subset of dataset to train the machine learning model, and we already know the output.

**Test set:** A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

Before applying the data to the algorithm we need to apply some preprocessing techniques to the data set.

**PREPROCESSING TECHNIQUES :**

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

1. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.The preprocessing techniques are:

1) Handling Null Values

2)Standardization

3)Handling Categorical Variables

4)One-Hot Encoding

5)Multicollinearity

# Handling Null Values:

In any real world dataset there are always few null values. It doesn’t really matter whether it is a regression,classfication or any other kind of problem, no model can handle these NULL or NaN values on its own so we need to intervene.

In python NULL is reprsented with NaN. So don’t get confused between these two,they can be used interchangeably.

Standardization:

It is another integral preprocessing step.In Standardization we transform our values such that the mean of the values is 0 and the standard deviation is 1.Formula for standardization:

# https://miro.medium.com/max/432/1*qwpzqVh-PVENHa6q1qZ8kQ.png

# Handling Categorical Variables:

Handling categorical variables is another integral aspect of Machine Learning. Categorical variables are basically the variables that are discrete and not continuous. Ex — color of an item is a discrete variables where as its price is a continuous variable.

# One-Hot Encoding:

So in One-Hot Encoding what we essentially do is that we create ’n’ columns where n is the number of unique values that the nominal variable can take.

**Multicollinearity:**

Multicollinearity occurs in our dataset when we have features which are strongly dependent on each other.

About Modules:

The whole project is done by python as background so that we use many modules like pandas ,numpy, metrics, pickle , etc…

**The below modules are used in the current project:**

**Pandasmodule :**

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data.

Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

**Key Features of pandas**

* Fast and efficient DataFrame object with default and customized indexing.
* Tools for loading data into in-memory data objects from different file formats.
* Data alignment and integrated handling of missing data.
* Reshaping and pivoting of date sets.
* Label-based slicing, indexing and subsetting of large data sets.
* Columns from a data structure can be deleted or inserted.
* Group by data for aggregation and transformations.
* High performance merging and joining of data.
* Time Series functionality.

**Numpy module:**

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

* **Mac**and **Linux**users can install NumPy via pip command:
* Pip install numpy
* **Windows** does not have any package manager analogous to that in linux or mac.Please download the pre-built windows installer for NumPy from here (according to your system configuration and Python version).  
  And then install the packages manually.

**Stacking:** Several arrays can be stacked together along different axes.

* **np.vstack:** To stack arrays along vertical axis.
* **np.hstack:** To stack arrays along horizontal axis.
* **np.column\_stack:** To stack 1-D arrays as columns into 2-D arrays.
* **np.concatenate:** To stack arrays along specified axis (axis is passed as argument).

**Label encoder module:**

In machine learning, we usually deal with datasets which contains multiple labels in one or more than one columns. These labels can be in the form of words or numbers. To make the data understandable or in human readable form, the training data is often labeled in words.

**Label Encoding** refers to converting the labels into numeric form so as to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way on how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning.

**METRICS MODULE :**

The metrics that you choose to evaluate your machine learning algorithms are very important.

Choice of metrics influences how the performance of machine learning algorithms is measured and compared. They influence how you weight the importance of different characteristics in the results and your ultimate choice of which algorithm to choose.

**PICKLE MODULE :**

Pickling is useful for applications where you need some degree of persistency in your data. Your program's state data can be saved to disk, so you can continue working on it later on. It can also be used to send data over a Transmission Control Protocol (TCP) or socket connection, or to store python objects in a database. Pickle is very useful for when you're working with machine learning algorithms, where you want to save them to be able to make new predictions at a later time, without having to rewrite everything or train the model all over again.

## Sample Code:

## img01.jpg

## img03.jpg

## Output:image55.jpg

## 2020-05-27.png

## GitHub Link:

## https://github.com/Traffic-prediction/miniproject.git