Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 7, April : 2024

FOUR CITY CLIMATE CHANGE ANALYSIS USING MACHINE LEARNING TECHNIQUES.

B. Anandakumar, Ph.D Scholar, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore-21.
Dr. S. Dhinakaran, Ph.D Guide, Assistant Professor, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore-21.

Abstract:

In this research paper, the temperature of four cities in Tamil Nadu is taken for temperature analysis from 2010 to 2020. In this research, different types of data analysis algorithms are used, and through those algorithms, their ability is tested, and then the rainfall prediction method is handled. For this study, the weather data of Coimbatore, Madurai, Trichy, and Chennai cities has been taken from 2010 to 2020. There are five types of algorithms used for this study, which are the KNN algorithm, the decision tree algorithm, the random forest algorithm, the Naive Bayes algorithm, and the artificial neural network. The above algorithms are used to calculate their accuracy. Then a rainfall prediction is calculated for each city. Further, this study explains how the number of vehicles is increasing for each city and how the increase in the number of vehicles in Tamil Nadu as a whole is related to the rise in temperature described.

I. Introduction

Climate change is one of the many problems that threaten the world today. Industrial growth after the 19th century caused more carbon emissions into the atmosphere due to the increased use of both fossil fuels in factories and vehicles. Due to the excessive use of fossil fuels, more greenhouse gases are being added to the atmosphere, due to which more heat is being retained in the earth, and because of this heat retention, there are more forest fires all over the world. In addition to this, the acidity of sea water is increasing, and coral reefs are continuously being destroyed; thus, the destruction of coral reefs is harmful to marine life.

There is a risk of food unavailability, and more and more storms are happening continuously, and rainfall is reaching the level of extreme rainfall, thus climate change is happening. If the current situation persists, the world is in danger of facing a major disaster. The United Nations Framework Convention on Climate Change (UNFCCC) [1] created the climate change treaty in 1992 with the goal of reducing greenhouse gases. A Conference of Parties was held by the UNFCCC in 2012 to put the Kyoto Protocol into action. The Kyoto Protocol's second part, greenhouse gas reduction, is covered by this COP from 2013 to 2020. The UNFCCC organised the summit of parties in 2015. (COP). Reduce the mean temperature here by no more than 2 degrees above pre-industrial levels. Therefore, the Intergovernmental Panel on Climate Change (IPCC) Organisation has established various conventions and agreements, among which the agreement held in Paris in 2015 is very important.

Recently, COP28 was held in the UAE in December 2023. The conference was held under the theme "Beginning of the End." The conference was attended by representatives from 154 countries who agreed to allocate funds to deal with the impacts of climate change. It was decided that a total of 100 billion US dollars should be allocated to low-income developing countries. It was further decided that the next conference will be held in Azerbaijan in November 2024 and the next in Brazil in 2025. The conference agreed to limit the temperature rise to 1.5°C and set a target to reduce greenhouse gas emissions by 43 percent by 2030 while increasing the use of renewable fuels.

Data Mining

Data mining is the process of taking information from a large collection of data. It is the result of the integration of computer science and statistics. Another goal of data mining is to identify new

informational patterns. Data mining is referred to as Knowledge Discovery in Databases, or KDD for short.In KDD, there are five steps.

A. SELECTION
The initial phase in the KDD process is information selection from a large pool of information.The relevant data has been eliminated from the overall volume of information. After filtering, there will be less data overall. It will help with the next stage of the KDD processing.
B. PREPROCESSING
After selecting the data, the next step is to obtain the desired data. Create a range of designs by converting the tools using the required data. There must be hidden information in every design.
C. TRANSFORMATION
During this technique, data is translated into an appropriate target form. There are two steps to the change. One of them is 1. Data sharing 2. Writing code. Data mapping is the process of allocating components of source and target data. The second is creating a transformation process.
D. DATA MINING
Currently, a number of techniques are being used to extract the data. Software tools are widely available for use in data mining. The new information pattern and information extraction from the target data are the tools.
E. INTERPRETATION/EVALUATION
Once the hidden pattern is found, information will be converted into a format that the user can understand. It will be represented by forms such as visual, report, and so on. Because of this, even someone who has never heard of data mining before can easily understand the concept.
F. DEPLOYMENT
The data or information will be released to the designated location upon the completion of the aforementioned stages. It's called deployment in this context.

II. Literature Review

The first journal [1] is Rainfall Prediction based on Humidity and Wind, which will be determine particular Location rainfall. In this study there are various methods are used for analysing the rainfall prediction, such as Nave Bayes, Support Vector Machine, Decision Tree, Neural Network, and Random Forest. In this study, meteorological data from Malaysia has been studied. In this research, the first process is processioning the data. Many Data are Missing is represented in the form of a question mark ("?"), and "*" is represented as negative values. So, the above data is prepossessing. After that information retrieval process, three methods are used: 1.Recall 2.Precision. 3.f-measure . In this study, decision trees and random forests performed well compared to other methods.

The next Journal[2] is rainfall forecast that use Multilayer Perception Neural Network(MLPNN) in processing rainfall data. The MLPNN architecture consist of three Layers such as Input Layer, Hidden Layer and Output Layer. There are 11 Elements are used as Input neuron . The Hidden layer use Various Element and 7 Elements used as Output neuron. In this research the data are segregated into three types 1. Training data 2. testing Data 3. Validation Data. In the training data set data are collected from 1871- 1960 or 89 years. The testing data is from 1961-2010 0r 49 years. In this study there are two methods used for processing the data. The first Method has 50 hidden neuron with Scaled Conjugate Gradient (SCG) training algorithm. The Second MLPNN model has 100 hidden neuron with SCG training algorithm and sigmoid Activation Function.

The next Journal[3] is ISMR prediction based on Causal precursor. For this study, global rainfall data from 1979 to 2004 were taken. In this study one year taken as ENSO positive and another one year taken as ENSO negative. The Data are spitted into two parts. The first part starting from the year 1979 to 2003. The second part starting from 2004 to 2016. The First part is working in the form of training hind cast model. The second Part is working in the form of test hind cast model. In this

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 7, April : 2024

study Response-Guided Causal Precursor Detection (RG-CPD) algorithm to find causal precursor without any assumptions.

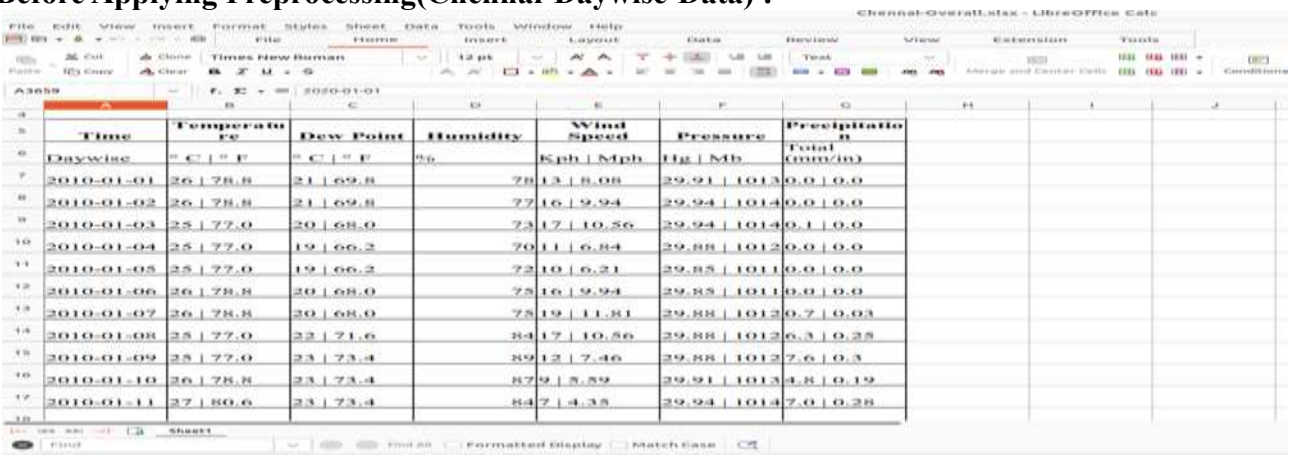
The Fourth Journal[4] is Statistical Evaluation of Combined Daily Gauge Observations and Rainfall Satellite Estimates. In this Study South American Data are taken from the year 2004 onwards. In this study a new methodology(CoSch) is used to Analysis the Waether Data. The CoSch is the Combination of satellite estimates and daily gauge Data.In this Study Cosch algorithm is Compared with two algorithm. One is called as control algorithm(3B42RT) and second one is called as 3B42V6(Which is Calculated from GPCC). There are two intermediate result will arise . These are 1. Add(Additive) 2. RAT(ratio). In the research conclude Cosch is better than control algorithm and 3B42V6.

III. Methodology

Preprocessing:

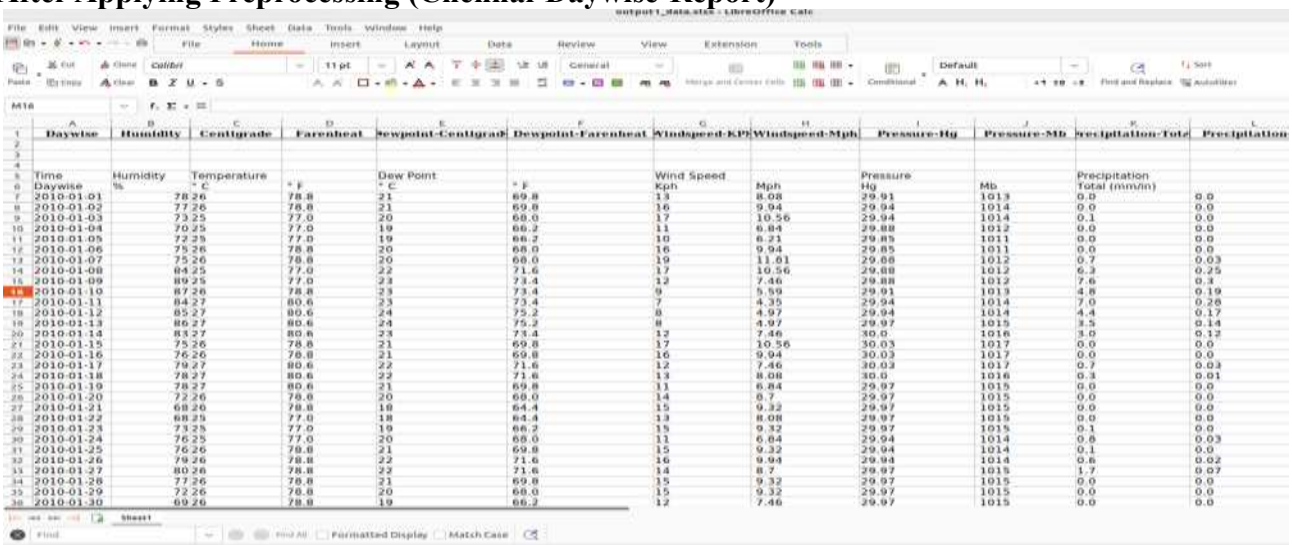
First, the weather data of four cities in Tamil Nadu, namely Chennai, Madurai, Coimbatore, and Trichy, is collected from 2010 to 2020, and then we analyse this data using different types of algorithms. This information is organised through the second standard preprocessing method of data analysis. All this information is in an Excel sheet. Among all this information, some columns include two pieces of information. It will be converted into a single column. Excel sheet is available for 7 columns before pre-processing, and after re-processing, it is converted to 12 columns.

Before Applying Preprocessing(Chennai-Daywise-Data) :



Time	Temperature	Dew Point	Humidity	Wind Speed	Pressure	Precipitation
Daywise	°C | °F	°C | °F	%	Kph | Mph	Hg | Mb	Total (mm/in)
2010-01-01	26 | 78.8	21 | 69.8	78	13 | 8.08	29.91 | 1013	0.0 | 0.0
2010-01-02	26 | 78.8	21 | 69.8	77	16 | 9.94	29.94 | 1014	0.0 | 0.0
2010-01-03	25 | 77.0	20 | 68.0	73	17 | 10.56	29.94 | 1014	0.1 | 0.0
2010-01-04	25 | 77.0	19 | 66.2	70	11 | 6.84	29.88 | 1012	0.0 | 0.0
2010-01-05	25 | 77.0	19 | 66.2	72	10 | 6.21	29.85 | 1011	0.0 | 0.0
2010-01-06	26 | 78.8	20 | 68.0	75	16 | 9.94	29.85 | 1011	0.0 | 0.0
2010-01-07	26 | 78.8	20 | 68.0	75	19 | 11.81	29.88 | 1012	0.7 | 0.03
2010-01-08	25 | 77.0	22 | 71.6	84	17 | 10.56	29.88 | 1012	6.3 | 0.25
2010-01-09	25 | 77.0	23 | 73.4	89	12 | 7.46	29.88 | 1012	7.6 | 0.3
2010-01-10	26 | 78.8	23 | 73.4	87	9 | 5.59	29.91 | 1013	4.8 | 0.19
2010-01-11	27 | 80.6	23 | 73.4	84	7 | 4.35	29.94 | 1014	7.0 | 0.28

After Applying Preprocessing (Chennai-Daywise-Report)



Daywise	Humidity	Centigrade	Farenheat	Dewpoint-Centigrade	Dewpoint-Farenheat	Windspeed-Kph	Windspeed-Mph	Pressure-Hg	Pressure-Mb	Precipitation-Total	Precipitation
Time	Humidity	Temperature		Dew Point		Wind Speed		Pressure		Precipitation	
Daywise	%	°C	°F	°C	°F	Kph	Mph	Hg	Mb	Total (mm/in)	
2010-01-01	78	26	78.8	21	69.8	13	8.08	29.91	1013	0.0	0.0
2010-01-02	77	26	78.8	21	69.8	16	9.94	29.94	1014	0.0	0.0
2010-01-03	73	25	77.0	20	68.0	17	10.56	29.94	1014	0.1	0.0
2010-01-04	70	25	77.0	19	66.2	11	6.84	29.88	1012	0.0	0.0
2010-01-05	72	25	77.0	19	66.2	10	6.21	29.85	1011	0.0	0.0
2010-01-06	75	26	78.8	20	68.0	16	9.94	29.85	1011	0.0	0.0
2010-01-07	75	26	78.8	20	68.0	19	11.81	29.88	1012	0.7	0.03
2010-01-08	84	25	77.0	22	71.6	17	10.56	29.88	1012	6.3	0.25
2010-01-09	89	25	77.0	23	73.4	12	7.46	29.88	1012	7.6	0.3
2010-01-10	87	26	78.8	23	73.4	9	5.59	29.91	1013	4.8	0.19
2010-01-11	84	27	80.6	23	73.4	7	4.35	29.94	1014	7.0	0.28
2010-01-12	85	27	80.6	24	75.2	8	4.97	29.94	1014	4.4	0.17
2010-01-13	86	27	80.6	24	75.2	8	4.97	29.97	1015	3.5	0.14
2010-01-14	83	27	80.6	23	73.4	12	7.46	30.0	1016	3.0	0.12
2010-01-15	75	26	78.8	21	69.8	17	10.56	30.03	1017	0.0	0.0
2010-01-16	76	26	78.8	21	69.8	16	9.94	30.03	1017	0.0	0.0
2010-01-17	79	27	80.6	22	71.6	12	7.46	30.03	1017	0.7	0.03
2010-01-18	78	27	80.6	22	71.6	13	8.08	30.0	1016	0.3	0.01
2010-01-19	78	27	80.6	21	69.8	11	6.84	29.97	1015	0.0	0.0
2010-01-20	72	26	78.8	20	68.0	14	8.7	29.97	1015	0.0	0.0
2010-01-21	68	26	78.8	18	64.4	15	9.32	29.97	1015	0.0	0.0
2010-01-22	68	25	77.0	18	64.4	13	8.08	29.97	1015	0.0	0.0
2010-01-23	73	25	77.0	19	66.2	15	9.32	29.97	1015	0.1	0.0
2010-01-24	76	25	77.0	20	68.0	11	6.84	29.94	1014	0.8	0.03
2010-01-25	76	26	78.8	21	69.8	15	9.32	29.94	1014	0.1	0.0
2010-01-26	79	26	78.8	22	71.6	16	9.94	29.94	1014	0.6	0.02
2010-01-27	80	26	78.8	22	71.6	14	8.7	29.97	1015	1.7	0.07
2010-01-28	77	26	78.8	21	69.8	15	9.32	29.97	1015	0.0	0.0
2010-01-29	72	26	78.8	20	68.0	15	9.32	29.97	1015	0.0	0.0
2010-01-30	69	26	78.8	19	66.2	12	7.46	29.97	1015	0.0	0.0

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 7, April : 2024

Feature Selection:

1. Filter Method

A.Chi Square Method:

Output:

Selected Features: ['Windspeed-Mph', 'Pressure-Hg', 'Pressure-Mb', 'Precipitation-Total', 'Precipitation-in']

The above is the Output of one of the Filter method of Chi Square Method. In this method it take 5 features from the above table column. The remaining data is eliminated.The above is Output of the Trichy city data filteraion method.Likewise the remaining city data's are Filtered.

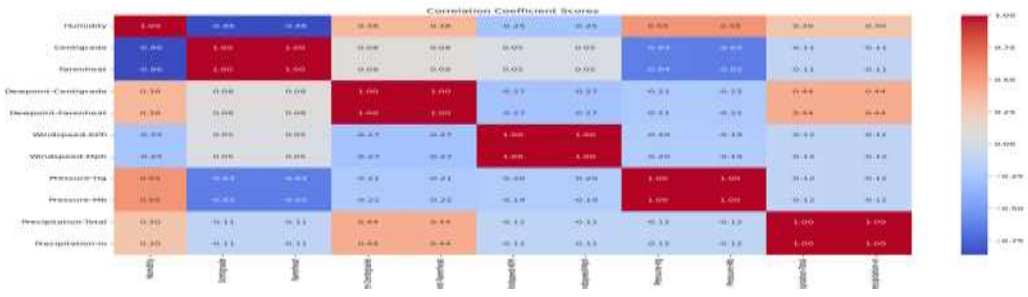
B. Information Gain

Information Gain Scores:
Precipitation-Total 3.512067
Precipitation-in 2.858856
Dewpoint-Farenheat 0.357675
Dewpoint-Centigrade 0.356927
Humidity 0.097663
Pressure-Mb 0.084328
Pressure-Hg 0.073497
Windspeed-Mph 0.039175
Windspeed-KPh 0.038650
Farenheat 0.035390
Centigrade 0.035387
dtype: float64

The above is the Output of Second filter method which is Information gain.In this method it take all data from the dataset.Then it will give the score to all column. Based on the above output the Two column is taken as target variable. The precipitation-Total and Precipitation-in are the two Column which got high Score. The remaining data's are low Score.Likewise the algorithm applied to remaining cities.

C. Correlation -Coefficient Method

	Feature	Score
0	Daywise	0.000000
10	Precipitation-Total	-4018.000000
8	Pressure-Hg	-4018.033759
9	Pressure-Mb	-4018.035828
5	Dewpoint-Farenheat	-4028.439204
3	Farenheat	-4036.448366
4	Dewpoint-Centigrade	-4055.589998
2	Centigrade	-4064.137311
1	Humidity	-4167.591727
7	Windspeed-Mph	-4559.575139
6	Windspeed-KPh	-4559.666384



The above is the third method which is called Correlation Co efficent Method. The above is the output which gives the result as the Precipitation-total and Precipitation-in got higest Result. The same program is applied to other cities which gives the same result.

D) Variance Threshold Method

Output:

	Humidity	Centigrade	Farenheat	...	Windspeed-KPh	Windspeed-Mph	Pressure-Mb
0	80	26	78.8	...	12	7.46	1013
1	81	26	78.8	...	11	6.84	1014
2	82	25	77.0	...	11	6.84	1014
3	78	25	77.0	...	8	4.97	1013
4	76	26	78.8	...	8	4.97	1011

[5 rows x 8 columns]

The above is the output of the Variance threshold method. In this method set the threshold value as 0.1 for target variable. In this program target variable as Precipitation-in column. After applying the threshold method got the result.

2.Wrapper Method

A. Recursive Feature Elimination(RFE)

Output:

Selected Features: ['Centigrade', 'Farenheat', 'Dewpoint-Centigrade', 'Dewpoint-Farenheat', 'Windspeed-KPh', 'Windspeed-Mph', 'Pressure-Hg', 'Precipitation-Total']

The above is the output of the Recursive Feature Elimination(RFE) Method which is implemented into the dataset. It will remove least important features. It gives the rank based on their Imporatance.

B. Sequential Feature Selection

Output:

Selected features: ('5', '6', '7', '8', '9')

MSE of the selected features: -5.615343698816957e-06

Performance of different feature combinations:

{1: {'feature_idx': (9,), 'cv_scores': array([-5.94128439e-06, -5.53597943e-06, -5.67252727e-06, -5.67292727e-06,
-5.28588976e-06]), 'avg_score': -5.621721623126191e-06, 'feature_names': ('9',)}, 2: {'feature_idx': (5, 9), 'cv_scores': array([-5.93967142e-06, -5.53189562e-06, -5.68220793e-06, -5.67487182e-06,
-5.27655957e-06]), 'avg_score': -5.6210412706420764e-06, 'feature_names': ('5', '9')}, 3: {'feature_idx': (5, 6, 9), 'cv_scores': array([-5.94184282e-06, -5.53208390e-06, -5.68531729e-06, -5.67488719e-06,
-5.27675434e-06]), 'avg_score': -5.622177106651132e-06, 'feature_names': ('5', '6', '9')}, 4: {'feature_idx': (5, 6, 8, 9), 'cv_scores': array([-5.94237847e-06, -5.53158979e-06, -5.69007953e-06, -5.67922204e-06,
-5.27602202e-06]), 'avg_score': -5.623858369851605e-06, 'feature_names': ('5', '6', '8', '9')}, 5: {'feature_idx': (5, 6, 7, 8, 9), 'cv_scores': array([-5.92731293e-06, -5.53491211e-06, -5.67648064e-06, -5.67815520e-06,
-5.25985762e-06]), 'avg_score': -5.615343698816957e-06, 'feature_names': ('5', '6', '7', '8', '9')}, 6: {'feature_idx': (3, 5, 6, 7, 8, 9), 'cv_scores': array([-5.92738731e-06, -5.54128157e-06, -5.67996316e-06, -5.68158155e-06,
-5.26071411e-06]), 'avg_score': -5.618185540439349e-06, 'feature_names': ('3', '5', '6', '7', '8', '9')}, 7: {'feature_idx': (3, 4, 5, 6, 7, 8, 9), 'cv_scores': array([-5.93205433e-06, -5.54330181e-06, -5.67997298e-06, -5.67915658e-06,

-5.26071620e-06]), 'avg_score': -5.619040379164889e-06, 'feature_names': ('3', '4', '5', '6', '7', '8', '9')}, 8: {'feature_idx': (0, 3, 4, 5, 6, 7, 8, 9), 'cv_scores': array([-5.98259820e-06, -5.52600700e-06, -5.67734199e-06, -5.66544352e-06,
-5.26687964e-06]), 'avg_score': -5.623654068462531e-06, 'feature_names': ('0', '3', '4', '5', '6', '7', '8', '9')}, 9: {'feature_idx': (0, 1, 3, 4, 5, 6, 7, 8, 9), 'cv_scores': array([-5.98872859e-06, -5.53035048e-06, -5.67774733e-06, -5.66926454e-06,
-5.27314383e-06]), 'avg_score': -5.627846955043493e-06, 'feature_names': ('0', '1', '3', '4', '5', '6', '7', '8', '9')}, 10: {'feature_idx': (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), 'cv_scores': array([-5.98890157e-06, -5.53090858e-06, -5.67783542e-06, -5.66927251e-06,
-5.27314310e-06]), 'avg_score': -5.628012236302277e-06, 'feature_names': ('0', '1', '2', '3', '4', '5', '6', '7', '8', '9')}}

The above is the output of the Sequetial Feature selection method.The Dataset is stored in the form of Excel Sheet. So, First it read the Excel Sheet. After that it exclude Non numeric column ‘Daywise’. Then set the target variable as ‘Precipitation-in’. The dataset is Split into two types. One is called Traning dataset and another one is called testing Set. Then create Linear regression model.The next step is implement Sequential feature Selector into the dataset.Then it will find best set of features with minimum MSE(Mean Squared Error).

C. Genetic Algorithm

Output:

gen	nevals
0	50
1	27
2	27
3	29
4	27
5	33
6	24
7	33
8	34
9	26
10	28
11	35
12	28
13	26
14	29
15	30
16	30
17	31
18	29
19	31
20	30
21	23
22	33
23	26
24	31
25	20
26	28
27	32
28	30

29 27
30 36
31 31
32 20
33 25
34 33
35 37
36 25
37 32
38 36
39 27
40 28

Selected Features: ['Humidity', 'Centigrade', 'Farenheat', 'Dewpoint-Centigrade', 'Windspeed-Mph', 'Pressure-Hg']
Mean Squared Error on Test Set: 8.249879015509502e-34
First Trichy Climate Dataset taken as input from 2010 to 2020.The next step is split the dataset into two parts one is X as features and y as target variable.Next step is to implement genetic selection algorithm. After that fit the genetic algorithm into training data set and then select best features. After that evaluate the performance of the selected features.

3.Embedded Methods:

A. Lasso Regression (L1 regularization)

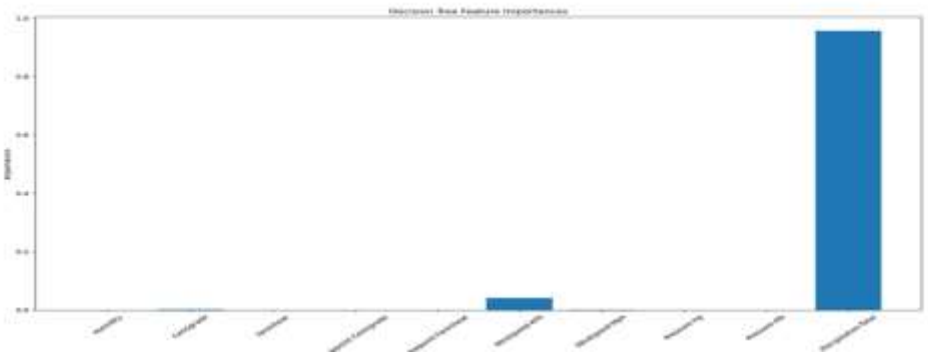
Output:

Selected Features: Index(['Precipitation-Total'], dtype='object')
Mean Squared Error on Test Set: 8.504234907370433e-05
The data set which consist of 11 years of data from the year of 2010 to 2020.Then the data set is splitted into two parts. One is training data and another one is testing data.After that standardscaler method is used to implement which execute and standardise the dataset.After that fit a Lasso regression model to the training data. The next step is selecting non zero coefficient from the dataset. Finally evaluate the model performance using the test data set.

B. Decision Tree and Random Forest

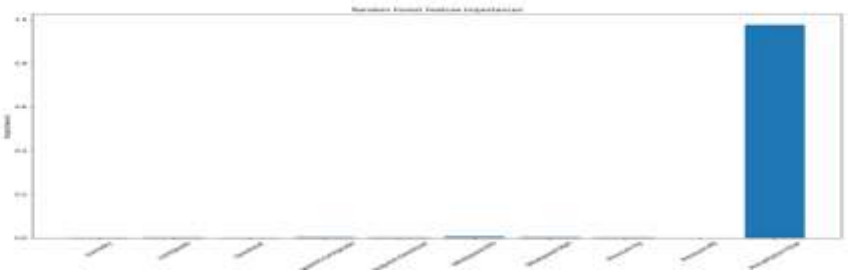
Output

Decision Tree Feature Importances: [4.41390011e-05 1.99645861e-03 9.01887814e-07 2.06364161e-06
2.75152214e-06 4.07300875e-02 6.94705839e-04 2.98081566e-06
1.96428109e-06 9.56523947e-01]



Random Forest Feature Importances: [6.34978601e-04 2.40014203e-03 9.64835915e-04 3.30279538e-03
2.65981314e-03 7.01504697e-03 4.05797060e-03 2.33591925e-03

2.89704636e-04 9.76338793e-01]
Mean Squared Error for Decision Tree on Test Set: 0.00042259259259259236
Mean Squared Error for Random Forest on Test Set: 1.8767864197530247e-05

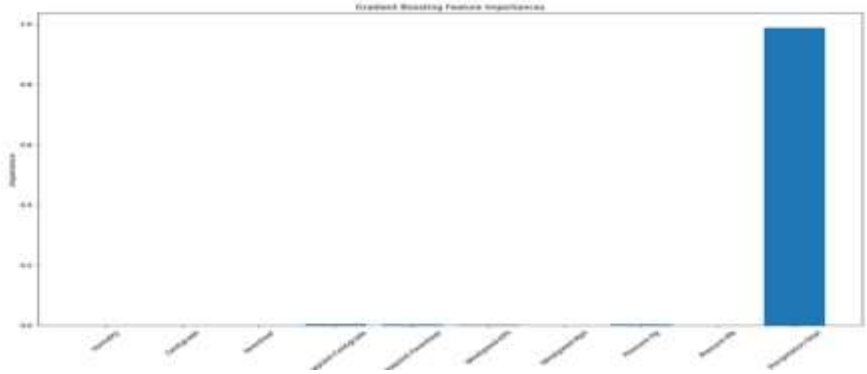


This method is come under Decision tree and Random Forest algorithm. First all data's are stored into excel File. After that Data's are splitted into training set and Testing set. Then "StandardScaler" Method is used for standarized the Dataset. After that Decision Tree and Random Forest Method is applied into data set or the data's are fit into Decision tree or Random Forest.Then i got the result of important features. Then the next step will be visualize the importance of each feature. Finnaly Evaluate the model performance of on the testing set using mean squared error.

Gradient Boosting Machines

Output

Gradient Boosting Feature Importances: [3.94265827e-07 5.54404323e-05 1.80287881e-04
4.16002771e-03
3.26040132e-03 1.23443714e-03 7.17699915e-05 3.02037591e-03
5.24687918e-07 9.88016341e-01]
Mean Squared Error for Gradient Boosting on Test Set: 4.017685226633661e-05



In this method first Data's are loaded into Excel file and read the Excel file. The next step is data's are splitted into training and testing set. Then using the method which is used to standarised the data called StandardScaler. Next step is fit the gradient Boosting model to the training data. After applying gradient boosting model, then get the imporatnt features. The next step is visualize the model and finally evaluate the model performance using mean squared error.

After the Preprocessing and feature selection the following algorithms are used to implment into the data and get the result.

1.Supervised Machine Algorithm
2.Naive Bayes Algorithm
3.Decision Tree Algorithm
4.Random Forest Algorithm

After Implementing the above algorithm, finally calculate rain prediction.

1.Supervised Machine Algorithm

After the preprocessing, it is then applied to the Supervised Machine Learning Algorithm. In this algorithm, Chennai City got a good result in terms of accuracy. Trichy, Mudurai, and Coimbatore got overall accuracy of 0.26, 0.22, and 0.25, respectively.

Parameter	City	Accuracy	Macro Avg	Weighted Avg	Overall Accuracy
Precision	Trichy		0.97	1.00	0.26789345
Recall	Trichy		0.92	1.00	
f1-Score	Trichy	1.00	0.91	1.00	
Support	Trichy	804	804	804	
Precision	Madurai		0.81	0.49	0.22950819672131148
Recall	Madurai		0.08	0.23	
f1-Score	Madurai	0.23	0.05	0.14	
Support	Madurai	366	366	366	
Precision	Coimbatore		0.88	0.58	0.25
Recall	Coimbatore		0.07	0.25	
f1-Score	Coimbatore	0.25	0.05	0.17	
Support	Coimbatore	804	804	804	
Precision	Chennai		1.00	1.00	1.0
Recall	Chennai		1.00	1.00	
f1-Score	Chennai	1.00	1.00	1.00	
Support	Chennai	1	1	1	

2.Naive Bayes Algorithm

After applying the SVM algorithm to all four city data sets,next move to the Naive Bayes algorithm. The overall accuracy in Trichy,Madurai, Coimbatore, and Chennai was 0.23, 022, 0.25, and 0.23, respectively. The remaining data exists in the table.

Parameter	City	Accuracy	Macro Avg	Weighted Avg	Overall Accuracy
Precision	Trichy		0.97	1.00	0.23209876543209876
Recall	Trichy		0.92	1.00	
f1-Score	Trichy	1.00	0.91	1.00	
Support	Trichy	804	804	804	
Precision	Madurai		0.81	0.49	0.22950819672131148
Recall	Madurai		0.08	0.23	
f1-Score	Madurai	0.23	0.05	0.14	
Support	Madurai	366	366	366	
Precision	Coimbatore		0.88	0.58	0.25
Recall	Coimbatore		0.07	0.25	
f1-Score	Coimbatore	0.25	0.05	0.17	
Support	Coimbatore	804	804	804	

Precision	Chennai		0.87	0.61	0.23209876543209876
Recall	Chennai		0.08	0.23	
f1-Score	Chennai	0.23	0.05	0.15	
Support	Chennai	810	810	810	

3.Decision Tree Algorithm

After applying the SVM algorithm and the Naive Bayes algorithm, we moved on to the Decision Tree algorithm. In this algorithm, Trichy,Madurai, Coimbatore, and Chennai result in 0.99, 0.99, and 1, respectively. This algorithm gives good results and good accuracy.

Parameter	**City**	**Accuracy**	**Macro Avg**	**Weighted Avg**	Overall Accuracy
Precision	Trichy		0.97	1.00	0.9987562189054726
Recall	Trichy		0.92	1.00	
f1-Score	Trichy	1.00	0.91	1.00	
Support	Trichy	804	804	804	
Precision	Madurai		0.96	1	0.9972677595628415
Recall	Madurai		0.92	1	
f1-Score	Madurai	1.00	0.90	1	
Support	Madurai	366	366	366	
Precision	Coimbatore		0.97	1.00	0.9987562189054726
Recall	Coimbatore		0.94	1.00	
f1-Score	Coimbatore	1.00	0.93	1.00	
Support	Coimbatore	804	804	804	
Precision	Chennai		1.00	1.00	1.0
Recall	Chennai		1.00	1.00	
f1-Score	Chennai	1.00	1.00	1.00	
Support	Chennai	810	810	810	

4.**Random Forest Algorithm**

After applying the SVM,naive bayes, and decision tree algorithms, we next move into the random forest algorithm. This algorithm also gives good results. The overall results of Trichy,Madurai, Coimbatore, and Chennai were 0.99, 0.99, and 1.0, respectively.

Parameter	**City**	**Accuracy**	**Macro Avg**	**Weighted Avg**	**Overall Accuracy**
Precision	Trichy		0.97	1.00	0.996268656716418
Recall	Trichy		0.91	1.00	
f1-Score	Trichy	1.00	0.90	1.00	
Support	Trichy	804	804	804	
Precision	Madurai		0.93	1.00	0.9972677595628415
Recall	Madurai		0.93	1.00	

f1-Score	Madurai	1.00	0.86	1.00	
Support	Madurai	366	366	366	
Precision	Coimbatore		0.97	1.00	0.9987562189054726
Recall	Coimbatore		0.94	1.00	
f1-Score	Coimbatore	1.00	0.93	1.00	
Support	Coimbatore	804	804	804	
Precision	Chennai		1.00	1.00	1.0
Recall	Chennai		1.00	1.00	
f1-Score	Chennai	1.00	1.00	1.00	
Support	Chennai	810	810	810	

5.Drought Prediction Algorithm

Parameter	City	Accuracy	Macro Avg	Weighted Avg	Overall Accuracy
Precision	Trichy		0.78	0.83	0.8296296296296296
Recall	Trichy		0.78	0.83	
f1-Score	Trichy	0.83	0.78	0.83	
Support	Trichy	810	810	810	
Precision	Madurai		0.79	0.81	0.8121890547263682
Recall	Madurai		0.78	0.81	
f1-Score	Madurai	0.81	0.78	0.81	
Support	Madurai	804	804	804	
Precision	Coimbatore		0.82	0.86	0.8669154228855721
Recall	Coimbatore		0.79	0.87	
f1-Score	Coimbatore	0.87	0.81	0.86	
Support	Coimbatore	804	804	804	
Precision	Chennai		0.79	0.81	0.8098765432098766
Recall	Chennai		0.77	0.81	
f1-Score	Chennai	0.81	0.78	0.81	
Support	Chennai	810	810	810	

6.RainFall Prediction

Trichy

Mean Squared Error: 25.944891973942884
Predicted Rainfall: [1.20806647e+12]

Madurai

Mean Squared Error: 97.4151618858831
Predicted Rainfall: [1.36301877e+14]

Coimbatore

Mean Squared Error: 29.67517560209207

Predicted Rainfall: [6.7065323e+12]

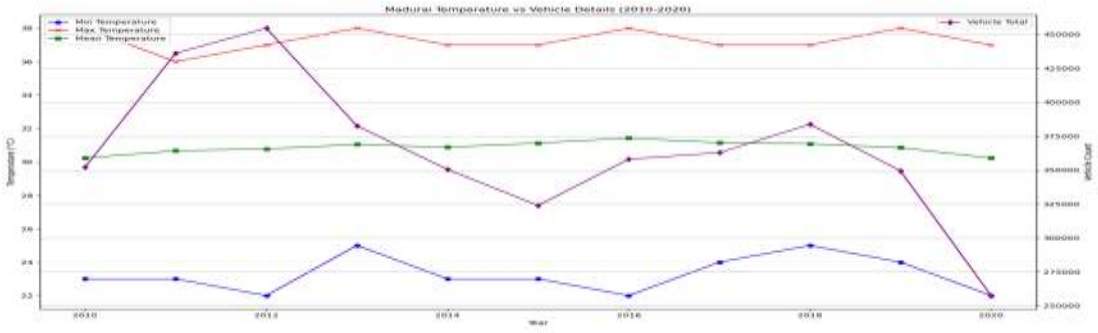
Chennai

Mean Squared Error: 2.7931519458936566e-30
Predicted Rainfall: [10.29305637]

Trichy Temperature vs Vehicle Details

Year	mean	max	min	Vehicle-Total
2010	30.232044198895	38	23	352130
2011	30.6676557863502	36	23	436110
2012	30.7772151898734	37	22	454520
2013	31.0465753424658	38	25	382650
2014	30.8821917808219	37	23	350450
2015	31.1260273972603	37	23	323960
2016	31.4266304347826	38	22	357970
2017	31.149863760218	37	24	362850
2018	31.0849315068493	37	25	383720
2019	30.8547945205479	38	24	349430
2020	30.2267759562842	37	22	257510

The above table shows the comparison between temperature and vehicle. The vehicle count increased from 2010 to 2012. In 2013, it decreased from the previous year. It decreased continuously from 2013 to 2016. Then the years 2016–2018 will continuously increase, and again, they will decrease. The below charts show the comparison between temperature and total number of vehicles. In the above table, the mean temperature has increased continuously from 2010 to 2012. The mean temperature decreased from 2017 to 2020, while the vehicle count also decreased continuously. In the year 2013, the vehicle count decreased, but the temperature increased. If the vehicle count increases or decreases, it will affect the mean temperature. The maximum temperature value is between 36 and 38. Moreover, the minimum temperature value is 22 to 25. The maximum and minimum temperature values vary from one year to another. It will not be used for any kind of prediction.

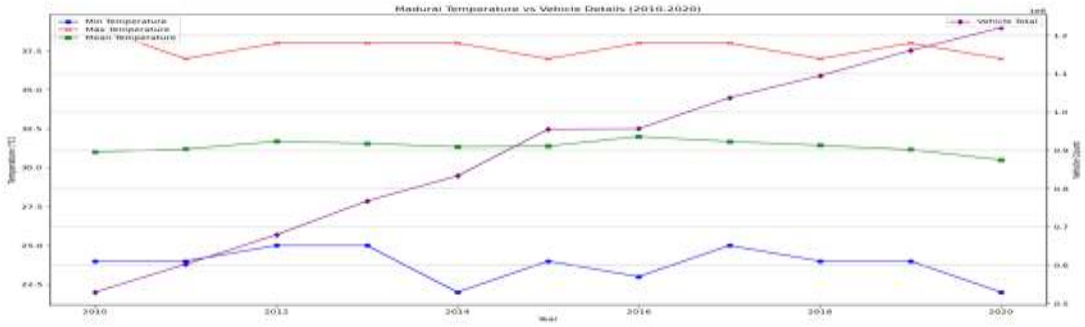


Madurai Temperature vs Vehicle Details

Year	mean	max	min	Vehicle-Total
2010	29.1417721518987	36	24	1179765

2011	28.5643835616438	35	23	1241096
2012	28.9617486338798	36	23	1386129
2013	28.9561643835616	36	23	1528042
2014	28.9287671232877	36	23	1648789
2015	29.1068493150685	36	22	1901277
2016	30.0463215258856	38	24	1903910
2017	29.1377245508982	36	23	2058614
2018	28.9397260273973	35	23	2163946
2019	28.7561643835616	35	22	2277814
2020	27.8879781420765	35	22	2377904

The above table shows the temperature versus the total number of vehicles. The vehicle count increased from 2010 to 2020. The mean value of the temperature also increased from 2010 to 2016. In the years 2013, 2014, and 2015, there was a slight variation. Exist, that is decreased, but overall it increased. From 2017 to 2020, the temperature value decreased. The maximum temperature in Madurai remained constant from 2010 to 2015 and increased by 38 degrees Celsius.

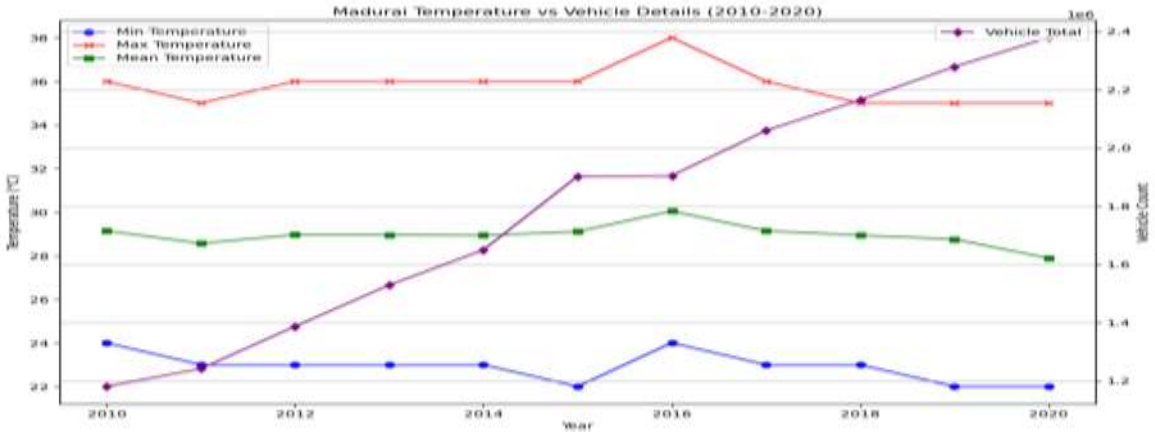


Coimbatore Temperature vs Vehicle Details

Year	mean	max	min	Vehicle-Total
2010	29.1417721518987	36	24	1179765
2011	28.5643835616438	35	23	1241096
2012	28.9617486338798	36	23	1386129
2013	28.9561643835616	36	23	1528042
2014	28.9287671232877	36	23	1648789
2015	29.1068493150685	36	22	1901277
2016	30.0463215258856	38	24	1903910
2017	29.1377245508982	36	23	2058614
2018	28.9397260273973	35	23	2163946

2019	28.7561643835616	35	22	2277814
2020	27.8879781420765	35	22	2377904

The above table shows the Coimbatore city temperature and vehicle total. The Coimbatore City vehicle fleet has continuously increased from 2010 to 2020. The mean temperature also increased from 2010 to 2016. After that, from 2017 to 2020, the mean temperature is decresed.The maximum temperature also increased from 2010 to 2016, and after 2016, the maximum temperature was 2 degrees. The minimum temperature decreased continuously from 2010 to 2020.

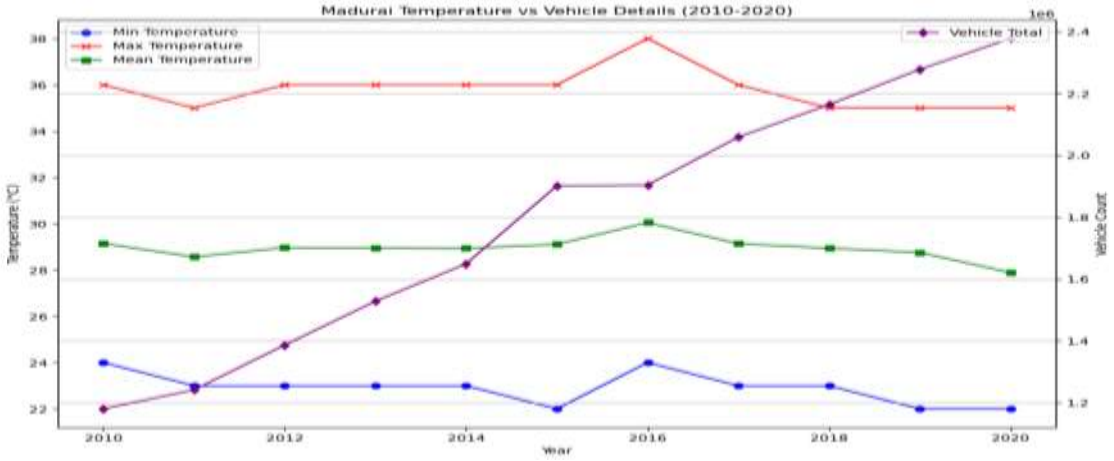


Chennai Temperature vs Vehicle Details

Year	mean	max	min	Vehicle-Total
2010	29.5178082191781	35	22	3148700
2011	29.3972602739726	35	25	3455789
2012	29.7841530054645	36	24	3767294
2013	29.5643835616438	36	25	4071979
2014	29.586301369863	35	25	4354231
2015	29.5780821917808	35	25	4934412
2016	29.6830601092896	35	24	4937988
2017	29.6493150684932	35	25	5298883
2018	29.4575342465753	34	24	5641519
2019	29.1753424657534	34	25	5996624
2020	29.0025188916877	36	25	6351729

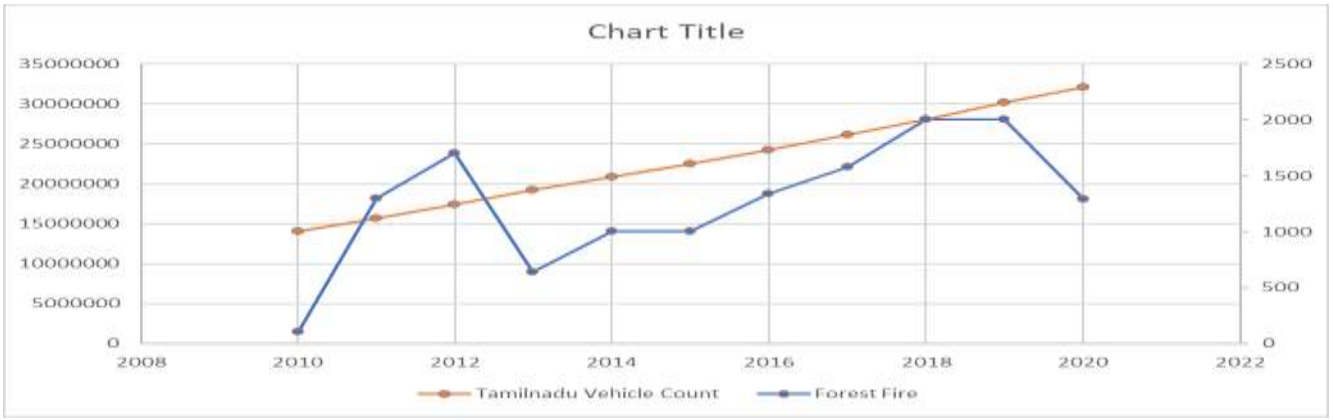
The above table shows the data on vehicle count versus minimum, maximum, and mean temperature in Chennai. The vehicle count has continuously increased from 2010 to 2020. The overall minimum temperature also increased from 22 degrees Celsius to 25 degrees Celsius from 2010 to 2020. But the minimum temperature was one degree Celsius in 2016 and 2018. Moreover, the maximum temperature has increased from 35 degrees Celsius to 36 degrees Celsius. But in the years 2018 and 2019, the temperature decreased from 35 degrees Celsius to 34 degrees Celsius. The mean temperature increased by 0.27 degrees Celsius in 2012 compared to 2010. The mean temperature increased by 0.17 degrees Celsius in the year 2016. The mean temperature has continuously decreased from 2017 to 2020.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, No. 7, April : 2024



Forest Fire vs Vehicle Count

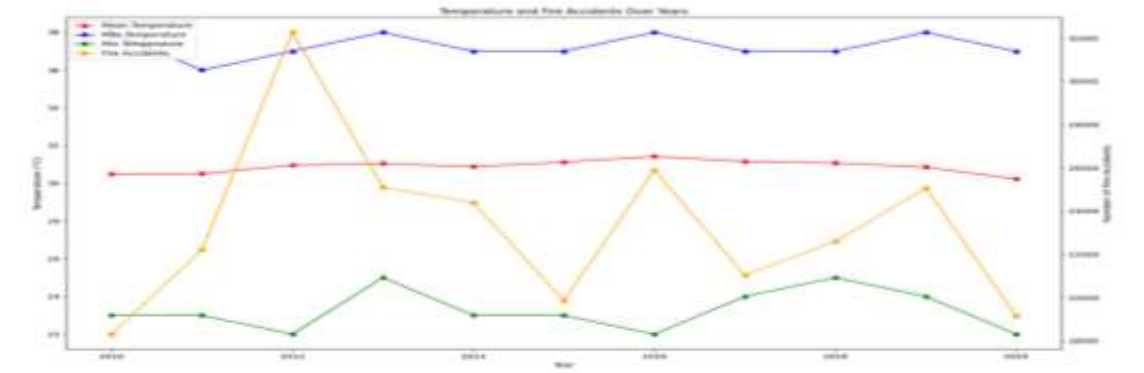
Year	Forest Fire Area	Tamilnadu Vehicle Count
2010	103	14060000
2011	1300	15640000
2012	1700	17410000
2013	643	19230000
2014	1002	20860000
2015	1002	22520000
2016	1340	24200000
2017	1580	26110000
2018	2003	28040000
2019	2006	30180000
2020	1290	32100000



In the above table, Tamil Nadu vehicle count versus forest fire area loss. The total number of vehicles continued to increase from 2010 to 2020. Meanwhile, fire area loss has also increased continuously in the years 2010–2012. In the year 2013, forest area loss decreased from the previous year. In the year 2014 onwards, the forest fire area loss has continuously increased. It was extended up to 2019. Again, in 2020, it decreased from the previous year.

Temperature Versus Fire Accident

Year	mean	max	min	Number of Fire Accident
2010	30.4860050890585	38	23	18311
2011	30.5123287671233	36	23	22219
2012	30.9562841530055	37	22	32273
2013	31.0465753424658	38	25	25109
2014	30.8821917808219	37	23	24393
2015	31.1260273972603	37	23	19866
2016	31.4266304347826	38	22	25897
2017	31.1498637602180	37	24	21041
2018	31.0849315068493	37	25	22601
2019	30.8547945205479	38	24	25068
2020	30.2267759562842	37	22	19142



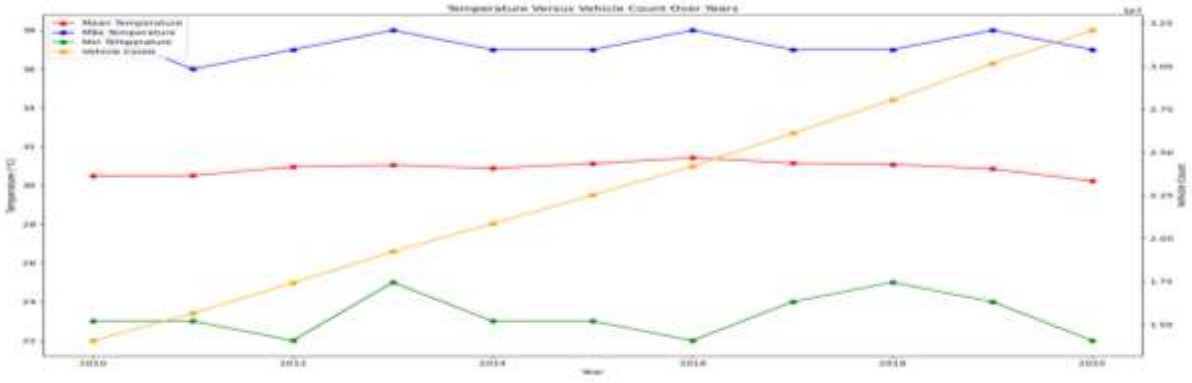
The above table shows the temperature versus fire accident details. The temperature is represented in the form of minimum, maximum, and mean values. The number of fire accidents continued to increase from 2010 to 2012. Mean While the mean temperature value is also increasing continuously, In the years 2013–2015, the number of fire accidents has continuously decreased from the previous year. But the mean temperature value has increased. Moreover, in 2016, the highest number of fire accidents occurred from 2010 to 2020. The mean temperature value was also high compared to other years. From 2017 to 2019, the number of fire accidents increased steadily. The mean value of temperature is also increased. In the year 2020, fire accidents decreased from the previous year. Meanwhile, the mean value of the temperature also decreased.

Temperature Versus Vehicle Count

Year	mean	max	min	Tamilnadu Vehicle Count
2010	30.4860050890585	38	23	14060000
2011	30.5123287671233	36	23	15640000
2012	30.9562841530055	37	22	17410000
2013	31.0465753424658	38	25	19230000
2014	30.8821917808219	37	23	20860000

2015	31.1260273972603	37	23	22520000
2016	31.4266304347826	38	22	24200000
2017	31.1498637602185	37	24	26110000
2018	31.0849315068493	37	25	28040000
2019	30.8547945205479	38	24	30180000
2020	30.2267759562842	37	22	32100000

The above table shows temperature versus vehicle count. The Tamil Nadu vehicle count has continuously increased from 2010 to 2020. The mean temperature also increased from 2010 to 2018. After 2018, it will decrease. The temperature difference between 2010 and 2018 was 1.40 degrees Celsius. The minimum temperature has increased from 23 to 25. The maximum temperature remained the same in 2010 and 2020.



IV. Conclusion:

In the above study reveals Mean,Max and Minimum Temperature of all is increased from 0.5 to 1 degree Celcisious. Meanwhile the vehicle count is continuosly increased from 2010 to 2020. There is direct relationship between vehicle count and temperature. Moreover Forest fire area loss is happen during the year 2010 to 2020. The temperature increased was effected the forest fire. Every year Fire accident is also increased. So, The temperature is increase will lead to fire accident. So the above study the vehicle which are running in Fossil Fuel increase. It increase the temperature of the city. The Temperature increase result to Fire Accident and Forest Fire. Moreover, this Research find the Temperaure increase will lead to Drought.

References

1.Suhaila Zainudin#, Dalia Sami Jasim#, and Azuraliza Abu Bakar , “Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction” International Journal Advanced Science Engineering Information technology(2016).

2.Lemuel Clark Velasco, Ruth Serquiña,Mohammad Shahin Ali Abdul Zamad and Bryan Juanico, “Week-ahead Rainfall Forecasting Using Multilayer Perceptron Neural Network” The fifth Information Systems international conference[2019].

3.G. Di Capua,M. Kretschmer,J. Runge,A. Alessandri,R. V. Donner,B. Van Den Hurk,R. Vellore,R. Krishnan,D. Coumou, “Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors”[2019].

4. Daniel a. Vila,luis gustavo g. De goncalves,david l. Toll,jose roberto rozante, “Statistical Evaluation of Combined Daily Gauge Observations and Rainfall Satellite Estimates over Continental South America”[2008].