Implementation of Machine Learning to Estimate the Air Pollutants such as Carbon Dioxide, Methane, Nitrous Oxide and Total Greenhouse Gas Emissions in Bangladesh

by

Sunanda Biswas
17101449
Spandan Sarkar
17101381
MD. Manazir Islam
17101524

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sunanda Biswas

Sunanda Biswas
17101449

Spandan Sarkar

Spandan Sarkar
17101381

Manazir

MD. Manazir Islam
17101524

Approval

The thesis/project titled "Implementation of Machine Learning to estimate the measurement of air pollutants such as Carbon Dioxide, Methane, Nitrous oxide and Total Greenhouse Gas Emissions in Bangladesh" submitted by

1. Sunanda Biswas (17101449)
2. Spandan Sarkar (17101381)
3. MD. Manazir Islam (17101524)

On October 02, 2021, the summer semester of 2021 was acknowledged as acceptable in partial completion of the criteria for the degree of B.Sc. in Computer Science.

Examining Committee:

Supervisor:

Saiful

Md. Saiful Islam
Senior Lecturer
Department
Institution

Co-Supervisor:

RAFEED RAHMAN

Rafeed Rahman
Designation
Department
Brac University

Ethics Statement (Optional)

This is optional, if you don't have an ethics statement then omit this page

Abstract

Nature and technology are two different subject matter with having much dissension between each. Only a few years back, technological growth looked like a threat to nature. However, the benefit of having huge computational power and Machine Learning applications, computers now have the capability of visualizing the vital component of nature. By using the concept of machine learning, researchers have exhibited the limitless use of artificial intelligence. As a part of that process, we have identified a specific problem on air pollution to tackle by using machine learning that just the human brain is unable to determine. We have taken Bangladesh's harmful emission factors into account, then trained them by using several machine learning techniques like regression and deep learning to predict the emission level. In consequence, we have applied models such as Linear Regression, Long Short Term Memory and Multi- layer Perceptron and found highest 99.05% of accuracy rate also described how this research can be extended in the context of other countries in future years.

Keywords: Machine Learning, Deep Learning, Linear Regression, Long Short Term Memory and Multi-layer Perceptron, Emission Factors

Dedication

It is hereby declared that,

- While proceeding with a degree at Brac University, I/we presented a thesis that is our original work.
- The thesis does not include any content that has been previously published or authored by a third party unless it is properly cited with complete and correct referencing.
- The thesis does not contain any material that has been approved or submitted for any other university or other institution's degree or diploma.
- All significant sources of assistance have been acknowledged.

Acknowledgement

First and foremost, we feel grateful to the Almighty for allowing us to finish our thesis without any serious setbacks.

Second, we would really like to express our deepest gratitude to our advisor, Md. Saiful Islam, sir, for his unwavering support and guidance throughout our project. He was always eager to accommodate us anytime we needed assistance.

Finally, without our parents' unwavering support, it may not be conceivable. We are currently on the verge of graduating, and thanks to their generous support and prayers.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ϵ	Epsilon
υ	Upsilon
$AdaBoost$	Adaptive Boosting
AQI	Air Quality Index
$CADM$	Comprehensive Action Determination Model
CCS	CO_2 capture and sequestration
DT	Decision Tree
LR	Linear Regression
$LSTM$	Long Short Term Memory
MLP	Multi-layer Perceptron
$RMSE$	Root Mean Square Error
RNN	Recurrent Neural Networks
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Overview

Substances that are harmful to human health and the environment are emitted into the air are core component of air pollution. Air pollution is a mixture from human and natural sources of dangerous pollutants. Air is the most essential element of the nature. The impact of harmful particles in air, can cause an uncertain change in the environment and eventually lead to catastrophic collapse of human civilization. For a densely populated county like Bangladesh, the issue of air pollution is undoubtedly the biggest threat.

The optimal quantity of natural constituents of air is vital for the survival of living creatures on earth. Observing the information given in AQI by 'U.S. Embassy in Bangladesh' [31] which states that "Bangladesh falls in the Unhealthy level (163) in the charter of health concern, it is very certain that people of Bangladesh face lungs disease mostly because of the air pollution". If we consider that number by looking at the IQAir site [36] it stands at the very top of the list. They claim that Dhaka city is now considered for having the most tainted air in terms of human living. Bangladesh, being one of the most densely populated nations with a population of 1.26 thousand people per square kilometer, the issue of air pollution has become a life threatening one for the upcoming generation [35]. The rapid growth of industrialization coupled with an increase in public and private transportation are among the significant contributors to an increase in the concentration of hazardous substances in the air. In spite of being an agricultural country, if the ratio of toxic materials increases significantly, then the consequences may lead to a catastrophic change in future of Bangladesh. Bangladesh government is forbidding people to cut trees without any rational reason as they deliver Oxygen which is essential for the survival of living organisms. As for keeping the ecological balance perfect, a country should have at least 33 percent of its land covered by forests [33], but in Bangladesh there is only 17.49 percent forest [40] of its total land.

According to Nathanson [34], automobile emissions, fires, industrial processes, electricity generation, fossil-fuel combustion are the utmost reason of air pollution. These can eventually cause loss of biodiversity, decreased reproduction, neurological problems in vertebrates, smog formation, damage to foliage etc. Besides, acid rain is one of the most alarming point to think of as it can damage buildings, monuments and human skins. Bereitschaft and Debbage [15] figure out that metropolitan regions with greater concentrations or emissions of CO_2 while controlling population, territory and climate exhibited greater urban sprawl or sprawl-like urban morphology. Focusing on global warming, Herndon [28] states that particle pollution is the main culprit behind various human diseases like obstructive pulmonary, neurodegenerative disease, lung cancer etc. From the perspective of Bangladesh the air pollution effects are on to the next level as Susmita along with is team's [10] research shows that, in poor households, small children and inadequate training women encounter pollution exposures that are four times higher in households with higher levels of education for men. This study also suggests, this problem can be solved by increasing the children's outdoor time 5-6 hours per and it should be in cooking period. A simulation based study has done by Muntaseer Billah Ibn Azkar [14] on urban and regional air pollution in Bangladesh. They claimed that air pollution increasingly become an important anthropogenic threat for the environment in Bangladesh. The financial consequences in Bangladesh's health sector alone is estimated, according to the World Bank report for Dhaka, at US 132–583 million Dollars annually. This is for the fourth biggest towns of Bangladesh at US 200–800 million Dollars each year, taking account for 0.7 to 3.0 percent of GDP per year mentioned in "C. Brandon, Economic valuation of air and water pollution in Bangladesh: Workshop discussion draft, 1997, World Bank". Although, not much study has done to approximate the quantities of significant air pollutants. There has not been any application of a specific geographical scale meteorological, emission, or "Chemical Transport Modeling System" yet.

Therefore, in this study we are working to develop a model for forecasting the level of CO_2, CH_4, N_2O, and Total Greenhouse Gas emissions, which are the major air pollutants in Bangladesh. This research is based on implementing Long Short-Term Memory of deep learning, Linear Regression, and Multi-layer Perception for the four pollutant factors and one of these algorithms has predicted 99.05 percent accurately. The rest of the paper includes a complete overview of all the data collection, processing, model implementation and result that we obtain from our work. Overall, in this paper, machine learning models have been used on a processed dataset to get a high-accuracy estimate that will assist the government of Bangladesh to decide which steps should be taken to avoid air pollution.

1.2 Problem Statement in Context of Bangladesh

Air is one of the most important elements for survival. Consequently, air pollution is a very serious issue. Brick Ovens are responsible for 58 percent of air pollution in the cities of Bangladesh which is a major concern to nature, according to the United News of Bangladesh (UNB) [32].They also published that Forests and Climate Change Minister Md Shahab Uddin has said that, "Plans have been taken to shut [traditional] kilns currently in operation. We're working to produce eco-friendly bricks". A statement like this is surely making a mark that the government of Bangladesh is ready for the challenges that they have to face in order to keep the air free from pollution. To deal with the devastating effects of air pollution as the Government is creating some plans, the automated systems with Machine Learning can be beneficial for taking the upcoming decisions. A predictive model can help to generate measurements of air substance rates in different time periods that might be helpful to take future decisions such as what should be the next step or what should be the main concern depending on the situation from the country's perspective. In this mission, every Bangladeshi citizen needs to realize the importance of having healthy air for the purpose of breathing. Since, Bangladesh's government is creating some plans, the usefulness of Machine Learning (ML) has to be the key for taking the upcoming decisions. A predictive system, that should generate a much reliable numbers of air substances in future has become a must need thing from the country's perspective. Algorithms like Linear Regression (LR), Multi-layer Perceptron Regression (MLP), Adaptive Boosting (AdaBoost), Gradient Boosting and Decision Tree (DT) are the hot topics in terms of demand. The implementation of these algorithms for a good and qualitative dataset should produce a perfect result for the quantity of some constituents of air.

1.3 Time Series Forecasting

Forecasting a 'Time Series' is a complex problem than classification and regression problems because of the order or temporal dependence between observations. Fitting and evaluating models require specialized handling of the data, which can be challenging. By definition, temporal dependence relates to the effect of previous activity on current behavior. A study published in IEEE has done by 'Hossein Hosseini', 'Sreeram Kannan' [22], and their team on temporal dependence with Latent Variables. According to them, "Every random operation has a Hidden (latent) State that can utilize to simulate the variables' internal memory and through a random lag, each variable can be affected by its latent memory state". As a result, memory recall with varying lags at different periods is modeled. In our example, we devised an LSTM that can solve 'Time Series' problems that 'Feed Forward Networks' with fixed-size windows cannot. Recurrent neural networks can also automatically learn the 'Temporal Dependence' of the data for time series prediction that make it better from having only general benefits. We have trained our dataset with a Stacked LSTM network that comprises multiple LSTM hidden layers and tuned the model with the Dropout regularization method. The model helps to get one output per input time step, basically, a sequence of output can be achieved through the network. Henceforth, this network is considered the most stable technique for sequence prediction problems. Time series forecasting can be implemented in social economic activities as development requires competitive analysis. This process also can generate scientific evidence that can be applied to maximize the financial value. Besides, regression models aim to pick the appropriate line for a particular dataset. 'The Sum of The Squares' of the discrepancies between the observed and predicted values is minimized by the Linear Regression technique. In our paper, we have designed a Linear Regression model which generates a line by output points y based on it's input x. The dependant, or explanatory variable is the forecast variable y and the independent or explanatory variables are also known as the predictor variables x. Another deep learning approach has also done by 'Md. Shiblee', 'P. K. Kalra' and 'B. Chandra' for 'Time Deries' forecasting with Multi-layer Perceptron [11]. 'Generalized Geometric' and 'Harmonic Error' measures for time series are the focus for their study. On contrast to that, we have used the MLP model to calculate 'Root Mean Square Error (RMSE)' value. The expected value of the square of the difference between the estimator and the parameter is specified as the RMSE. Our paper is a describes these models to predict the pollutant factors according to it's annual value.

1.4 Research Objectives

Air pollution does indeed have a profound impact on human existence. According to 'REAZ HAIDER- UNB STAFF WRITER' report [7] it has been found that an average Bangladeshi person's life span shortens by 1.87 years due to air pollution. Furthermore, long term health damages are caused for example, long-term impacts of air pollution include damage to the Lungs, Liver, and Kidneys, as well as Heart Disease, Lung cancer, and Respiratory Diseases such as Emphysema. As a result, Greenhouse Gases, and air pollutants have a negative impact on both the environment and human health. Byproducts of industrialization and mass production are known to contribute most to air pollution. As a result, it is expected that strong policy will be put in place to protect the country from higher risks.

- The aim of this research is to figure out the alarming level of threatening air substances and identify their accurate measurement.
- To visualize the increasing rate of harmful air particles are making the environment of many places of Bangladesh uninhabitable for living according to WHO.
- The research focuses on the air pollutant particles as well as the factors that are responsible for the increasing the level of those substances.
- Based on this study further necessary steps should be taken in order to protect the balance of air particles.

1.5 Challenges Faced

Although air pollution is a very familiar topic all over the world, but in context of Bangladesh the resources are not up to the mark. At first, we faced a lot of difficulties to prepare a suitable data set for our research. We conducted several meetings with our supervisor and co-supervisor sir to understand what our primary goal will be.

After collecting the raw data, our main concern was to make that data ready of model implementation. Along that, choosing the appropriate machine learning algorithms was a challenging issue. For that, we have to learn several new processes and techniques of data science. Throughout this time, our supervisor and co-supervisor Sir guided us with several topic to cover for the project.

Due to the COVID-19 pandemic, even our team did not have any option to seat face to face to discuss about our work. We have conducted all our personal and formal meeting online. Although, we have cover up these issues and tried best to complete this research.

1.6 Importance of Environmental Studies

For developing sustainable strategies to protect the nature and leading healthy life, importance of environmental studies is the prominent way of concern. It is getting mandatory to create awareness how to preserve the environment for the future

generation. ‘Laura Varela-Candamio’, ‘Isabel Novo-Corti’, ‘María Teresa García-Álvarez’ have done meta analysis on the importance of environmental education in the determinants of Green Behavior [25]. Their study has intrapersonal, motivational, interpersonal, and educational as factors for the Comprehensive Action Determination Model (CADM). “Human conduct has taken on a significant role in environmental protection” is the summary of this research. Besides, our team get motivated by the research of ‘Edward S Rubin’, ‘Margaret R Taylor’ and their team on the role of environmental technology in climate policy analysis [8]. In situations without learning ’CO_2 Capture and Sequestration (CCS)’ technologies, the cost of achieving the climatic stabilization aim are much lower as they claimed. These type of studies influenced us to visualize the environmental problem in Bangladesh and pick a demanded topic to research.
Since the spread of Corona virus, number of environmental research is escalating day by day. In particular, air quality has been the prime centre of attention as mentioned by Muhammad Usman and Yuh-Shan Ho [38]. Their finding shows that,

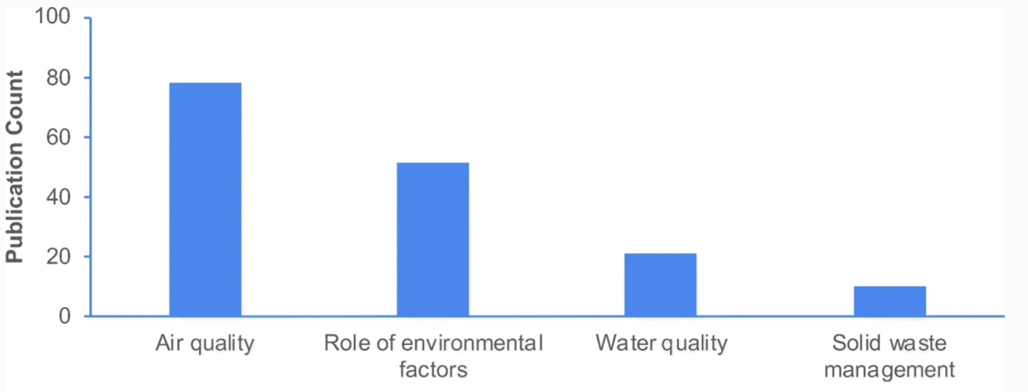

Figure 1.1: In the COVID-19 period, main subjects and the number of environment-related publications have increased [38]

emission of toxic gases like CO_2, CO, NO_2, SO_2 etc. have been significantly reduced in most of developed cities all over the world. Opposite to that, there is no evidence of reduction of these substances in Bangladesh. So, our research should be play a key role to realize the current scenario of major air pollutants in Bangladesh.

1.7 Thesis Outline

While selecting our research topic, we aimed to address a problem which is the most alarming issue for the future of Bangladesh. So as we have identified the proper machine learning approach on a processed dataset to predict the harmful substances that roaming in the air of Bangladesh. We have shown how much accurately can these models predict as well as comparisons among these algorithms. Along with it, provide visual reasoning behind the output.

At first, in (Chapter 1), we have given and overview of our topic. We have stated why ML is important in predicting air pollution. The problem statement was introduced in the perspective of Bangladesh. Also, an introduction of time series forecasting and describes the importance of environmental research.

Secondly, (Chapter 2) is the combination of related work that we have gone through to have some aspects on how air pollution factors can be categorised into machine learning. Also, analyse the results by the researchers.

Thirdly, in (Chapter 3), background analysis on out selected models have shown. The core idea behind each algorithms with graphical representation have illustrated briefly. These are the that we have implemented for obtaining our goal in this paper.

After that, (Chapter 4) contains all the research methodology which contains data collection, pre-processing. Here, it is described that how our data is fitted into the models. Next, how our team has developed the machine learning models for the processed dataset is mentioned in (Chapter 5). From setting up the parameters to using activation functions each steps have been described elaborately.

From there on, (Chapter 6) is the result section our study. Along with visual comparison of LSTM, MLP and LR, we have generated the RMSE result and concluded an experimental analysis.

Finally, we have concluded our research and mentioned the scope of further studies in (Chapter 7).

Chapter 2

Related Work

As the invention of Machine Learning (ML), it has given a new dimension to the research level. The reason is very simple and efficacious. Most of the machine learning algorithms use a huge amount of dataset to generate an outcome that is very much close to the actual result. These input sets are basically depended on some factors with following some principals. A few of the researchers use some computational approaches which is also worthwhile in that sector. 'Li F', 'Xiao X', 'Xie W', 'Ma D', 'Song Z', 'Liu K' [27] have used a quantitative approach to find out the emission factor of provincial grids and pollution transfer matrix in the 'Yangtze River Delta Region', China by interprovincial electricity transmission. They calculated that, SO_2 reached 20.50 Kiloton, NO_x reached 22.40 Kiloton, dust reached 4.30 Kiloton, CO_2 reached 39.23 Megaton were transferred to near-by regions till now from the Yangtze River Delta Region (YRDR) just because of supply of electricity.

On contrary of that, 'Z. Ghaemi', 'A. Alimohammadi' and 'M. Farnaghi' [29] have used a LaSVM-based big data learning model to build a system that has the capability of 'Dynamic Prediction of Air Pollution in Tehran'. "Data of pollution days of week and hours of day" have been their key aspects of study. This study shows that out of 21 locations the LaSVM model predicted 18 accurately. From this paper, the claim is that, overall accuracy of LaSVM algorithm is 0.71 and RMSE is 0.54. With it though, 'Traffic', 'Elevation', and 'Surface Curvature' are taken into account to monitor the spatial distribution of air pollution, as well as 'Wind Direction' and 'Speed', 'Cloudiness', 'Temperature', 'Pressure', and 'Relative Humidity', among other elements.

One of the most prominent research has done by 'Ke Hu', 'Ashfaqur Rahman', 'Hari Bhrugubanda' and 'Vijay Sivaraman' [24] in terms of predicting metropolitan air pollution estimation. Their main way of collecting these data from mobile sensors. Seven regression models have compared in this project and out of these Support Vector Regression (SVR) performance proved to be most reliable one. For the purpose of this estimation, the other regression algorithms were, 'Decision Tree Regression (DTR)', 'Random Forest Regression (RFR)', 'Extreme Gradient Boosting (XGB)', 'Multi-Layer Perceptrons (MLP)', 'Linear Regression (LR)', 'Adaptive Boosting Regression (ABR)'.

These are the Air Pollution related works that helped us to learn various procedures about how to research on air pollutant substances using machine learning. These papers have identified several techniques based on their county's perspective. But after observing our available dataset, we have come up with three algorithms that suits perfectly for our research, these are Long Short-Term Memory (LSTM), Linear Regression (LR) and Multi-layer Perceptron (MLP). While observing these models, the research by 'Wenquan Xu' [30] and his team on Time Series Forecasting based on a 'Linear Regression' model and 'Deep Learning', helped our team to select Linear Regression for our task. Although, this is a hybrid model and they claimed that the 'Hybrid Model' has a high level of predictability., but this research has given us a clear understanding about time series.

After that, a paper published in 2002 by 'Springer' [6], London has given our team an unique understanding of LSTM to 'Time Series' prediction. In this research Time-Window approaching has been applied by the researchers. The gist of this work is 'Long Short-Term Memory (LSTM)' can tackle many Time Series jobs that 'Feed-Forward Networks' with fixed size time windows cannot. This paper also suggests to use LSTM when simple conventional methods struggle.

A case study of the 'Tehran Stock Exchange' by 'Reza Ebrahimpour' [13] and his team is a perfect example of 'Time Series Forecasting' based on combination of 'MLP-experts' for trend. They worked on the Tehran stock exchange, 3 Neural Network merging techniques and an 'Adaptive Network-Based Fuzzy Inference System' which is use to forecast trends. They found the detection rate for 'Stacked Generalization', 'Modified Stacked Generalization', and 'ANFIS', respectively, are 75.97 percent, 77.13 percent, and 81.64 percent for several strategies for forecasting the 'Time Series' trend. But, the rate of recognition is enhanced to 86.35 percent when using 'The Mixture of MLP experts (ME)' method. Thus, this paper has given us to see a distinctive view of using Multi-layer Perceptron.

Besides, time series data analysis has long been a major topic of interest in domains other than Deep Learning, such as Economics, Engineering, and Medicine. 'James D. Hamilton' [3] describes typical data manipulation approaches, while 'E. Michael Azoff' [2] shows the use of traditional ANN techniques to this type of data. Major attempts were made to recognize which input stimulates a neuron optimally in order to describe human comprehensible characteristics or notions to which the neuron is responding mentioned by 'Ross Girshick', 'Jeff Donahue', 'Trevor Darrell' and 'Jitendra Malik's' work [17]. The information barrier concept has also gained traction as a means of describing the Deep Learning model's adaptability and intellectual abilities. Despite its novelty,Deep learning in recent years has gained much importance as mentioned by 'Jürgen Schmidhuber' [19] since it provides a comprehensive overview of the field's history of accomplishments that lead to its current state. Also, Vast influences recent innovations by deep learning has been described by 'I. Arel', 'Derek C. Rose' and 'T. Karnowski' in their work "*Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]*" [12]. We next move on to a discussion of Deep Learning's applications to Time-Series Data.

All of these findings use various dataset on various algorithms, but the foremost reason behind this is to make people aware of how much dramatic changes the air particles will have in the upcoming years. Along that, these works guided us to visualize the implementation of ML in various working fields. All the procedures described in these papers put significant amount of value to determine path of our own work. Finally, we are being motivated to design our own model using a suitable dataset for the sake of predicting overall air substances.

Chapter 3

Background Analysis

3.1 Linear Regression

To find out how the ‘Dependent’ and ‘Independent’ variables are related, one of the most used statistical analysis is ’Linear Regression’. [23]. The probability distribution function that represents the Regression is in Eq. 3.1.

$$Y = f(X, ß) \tag{3.1}$$

In this equation Y denotes the ‘Dependent’ variable and X is the ‘Independent’ variable also known as input for the equation. ß is known as an unknown parameter.

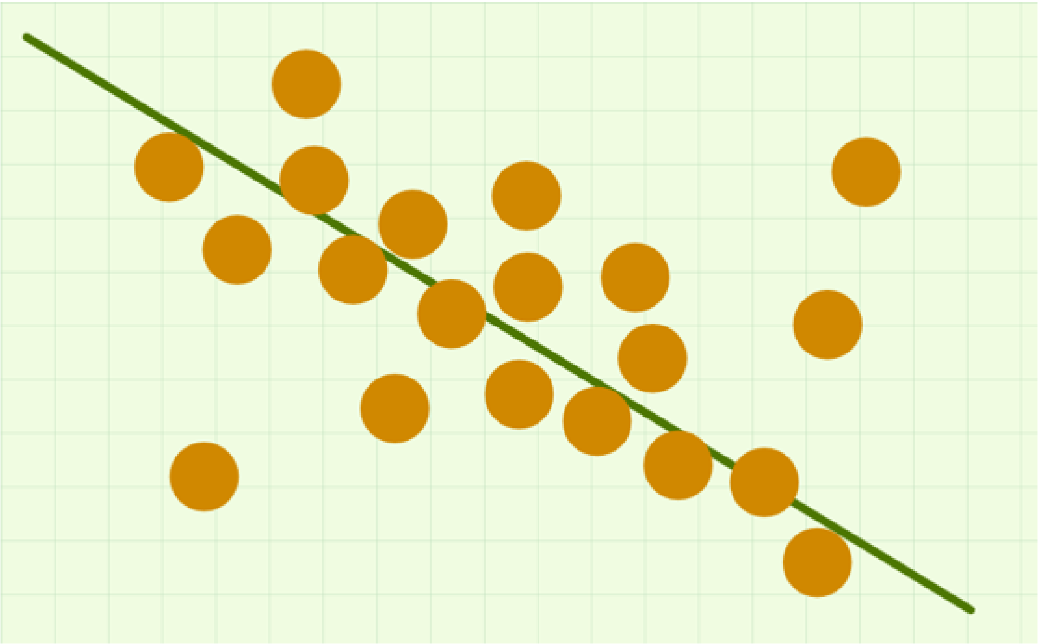

Figure 3.1: Regression Line

Univariate and Multivariate Regression are the two types of variable dependencies. Univariate Regression is used to figure out dependency among single variables shown in Eq. 3.2.

$$y = a + bx + \epsilon \tag{3.2}$$

Here y and x both are represented as dependent and independent variables with coefficient b and a is the constant. On the other hand, Multivariate Regression identifies the connection among several variables simultaneously shown in Eq. 3.3.

$$y = a + b_1x + b_2x + \ldots.. + b_nx + \epsilon \tag{3.3}$$

Another adaptive way of Linear Regression is to make it Polynomial Regression by making nonlinear relationships between variables. This can be achieved using data conversions based on basis functions. The technique can be implemented in making of pipeline used in Hyperparameters and Model Validation and Feature Engineering. Here, the 'Independent' variable x and the 'Dependent' variable y is modelled as an *nth* Degree Polynomial in x. This is nothing but the method of making multidimensional linear model given in Eq. 3.4.

$$y = a + b_1x + b_2x^2 + b_3x^3 + \ldots.. + b_nx^n + \epsilon \tag{3.4}$$

It is still a linear model as linearity attributed to the idea that AA coefficients never divide or multiply each other. In Linear Regression we use an effective one dimensional x value as input of the function. On the other hand, to fit more complicated relationship between x and y we upgrade the dimension of input x for the function in Polynomial Regression.

For learning Linear Regression, it is must to estimate the values of coefficients utilized with the data available to us for the representation. There are four types of techniques to design a Linear Regression model and they are *Simple Linear Regression*, '*Ordinary Least Squares*', '*Gradient Descent*', and '*Regularization*'. Besides, as the model has been well researched, there are many additional techniques but these four are the most useful among all. '*Ordinary Least Squares*' is the most widely used and '*Gradient Descent*' is best approach for machine learning.

Simple Linear Regression
In practical, this is not a very useful application. If a single input is available, a statistical way to estimate the coefficients applied in Simple Linear Regression. Means, standard deviations, correlations and covariance are essential to calculate statistical properties.

Ordinary Least Squares
When there are several inputs, this method can be used to predict the values of the coefficients. The residual sum of squares (RSS) is minimized using the Ordinary Least Squares approach. The process is a Regression line through the data to calculate the distance from each data point to the Regression line then square it. After that, summation of all the Squared Errors is performed. This is the amount the Ordinary Least Squares attempts to decrease as possible. Concept of matrix is needed here to manipulate the dataset. As well as understanding of Linear Algebra

procedures aid in predicting the best Coefficient Values. The availability of all the data and enough memory are essential to fit the data into the model and perform the matrix operations. Quick calculation is a feature of this process.

Gradient Descent
Gradient Descent is applicable when there are one or more inputs than that. It is an useful process of optimizing the values of the coefficients. Also, minimizing the model's error for the dataset continuously. At first, the procedure puts random values for each coefficient. Following that, The residual sum of squares (RSS) is determined for each pair of input and output values.. For minimizing the error, coefficients get updated in such that way. Addition to that, as a Scale Factor, a learning rate is applied. A Minimum Sum Squared Error is the main achievement of this process and it requires repetition of the process again and again. Alpha is a Learning Rate Parameter which sets the magnitude of the improvement step to be taken on each Procedural Iteration. This technique is particularly useful when dealing with large datasets with a large number of rows or columns that may not fit in memory.

Regularization
This method reduces the model's complexity while minimizing the model's sum of The Squared Error on the training data, which may be achieved using Ordinary Least Squares. This reduction complexity is the number or absolute size of the sum of all coefficients in the model. There are two types of 'Regularization' process, '*Lasso Regression*' and '*Ridge Regression*'. If, 'Ordinary Least Squares' overfit the training data and collinearity in input values make these two methods useful for application.
Ridge Regression which is also known asL_2 is the most common way of Regularization. Tikhonov regularization is another name of this approach. This is accomplished by penalizing the model coefficients' sum of squares (2-norms). The model fit penalty has given in Eq. 3.5.

$$P = \alpha \sum_{n=1}^{N} \theta_n^2 \tag{3.5}$$

Another way of Regularization is *Lasso Regression* which indicates penalizing the sum of absolute values (1-norms) of regression coefficients given in Eq. 3.6.

$$P = \alpha \sum_{n=1}^{N} |\theta_n| \tag{3.6}$$

Lasso Regression is also know as L_1. The strength of the penalty is monitored by α.

3.2 Multi-layer Perceptron

Artificial Neural Networks
Human brain is the most significant organ that has the capability to take decision

based on its previous experiences. The concept of ANN was build based on interconnected brain cells by programming computers. The phrase is derived from 'Biological Neural Networks,' which are responsible for the development of the human brain's structure. At first, idea of making layers inside network was created.

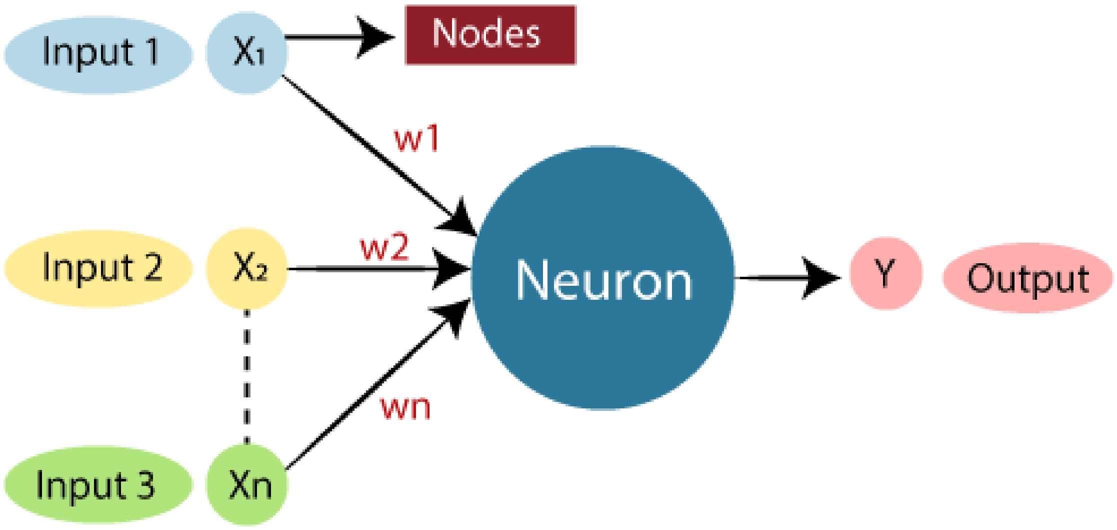

Figure 3.2: Artificial Neural Networks [20]

Figure 3.2 describes some input gets multiplied with assigned weights then these inputs get inserted into neuron to generate output. This process is very similar to Biological Human Neural Network shown in Figure 3.3.

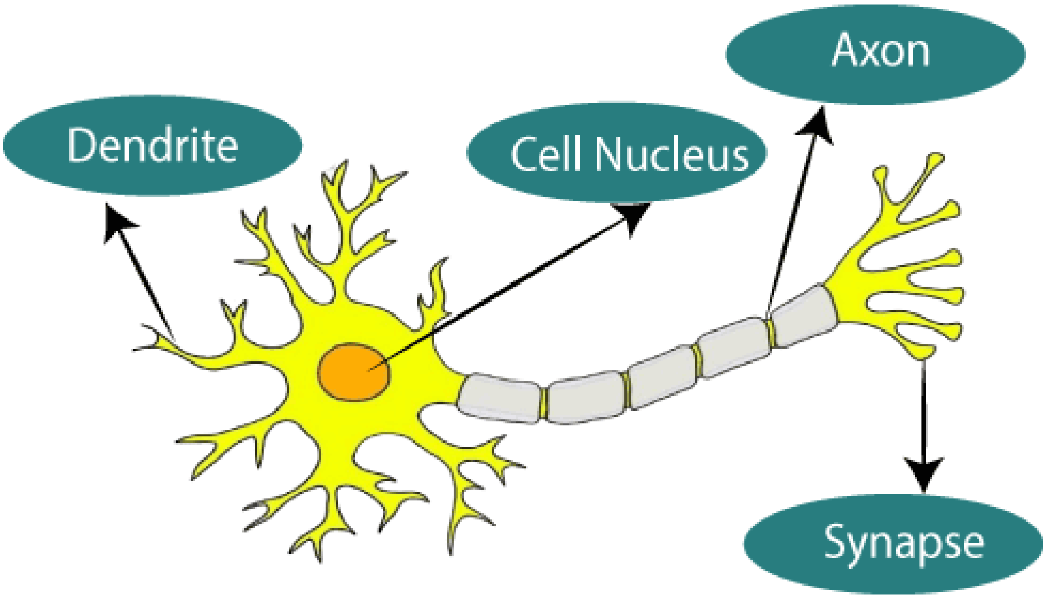

Figure 3.3: Biological Neural Network [20]

‘Biological Neural Network’	‘Artificial Neural Network’
‘Dendrites’	‘Inputs’
‘Cell nucleus’	‘Nodes’
‘Synapse’	‘Weights’
‘Axon’	‘Output’

Table 3.1: Relational terms between Biological Neural Network and ANN [20]

According to Wang SC’s consideration artificial neural networks are universal function approximators [7]. The application of artificial neural networks are widely considerable such as text classification, information extraction, semantic parsing, speech and character recognition etc. This method has given computer the power of thinking like human do.
The perceptron was presented in 1958 which is recognized for the first practical artificial neural network [1]. Neural Networks have grown increasing popularity since 1986. Neural Networks are mathematical structures built as the simplified version of the human nervous system [9]. To achieve expected output, the neural network transforms the inputs according to the process. Multi-Layer Perceptron is the most commonly used Neural Network model [21]. A Perceptron is a Supervised Learning method for binary classifiers. This algorithm makes it possible for neurons to learn. Also, one by one, the components in the training set are processed. Single layer and Multi-layer are two types of Perceptrons. Only linearly separable patterns can be learned by Single Layer Perceptrons on the other hand, the processing capability of Multilayer Perceptrons or Feedforward Neural Networks with two or more layers is larger. In order to build a linear decision limit, the Perceptron algorithm computes the weights for the input signals.

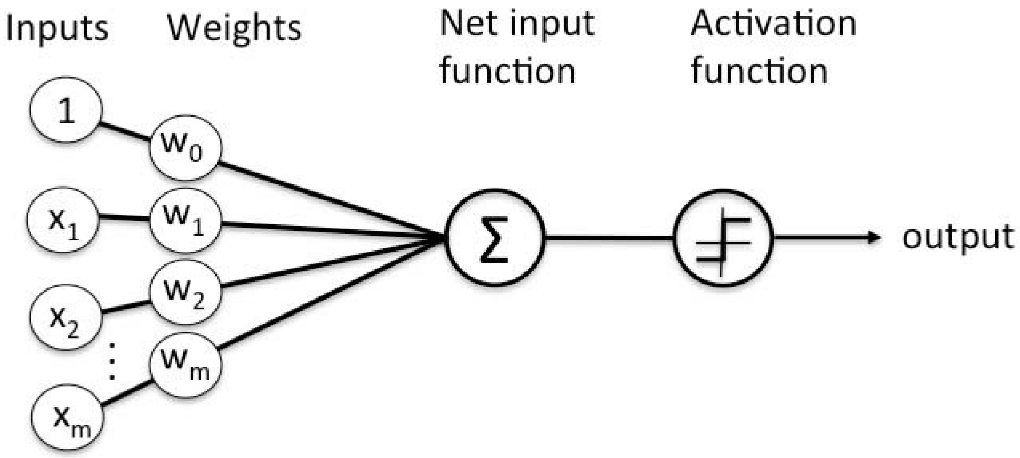

Figure 3.4: Working procedure of Perceptron [39]

According to the Perceptron Learning Rule, the system will automatically discover the best weight coefficients. To check the neurons are working ability, Input Features are then multiplied with these weights. If the input signals exceed a specific threshold, the Perceptron will receive several input signals, either signal will be produced or the output is not returned.

Perceptron Function
Perceptron maps its input and then it is multiplied by a weight coefficient that has been learned previously. From this an output value gets generated which is shown in Eq 3.7.

$$f(x) = \begin{cases} 1 & if\ w * x + b > 0 \\ 0 & otherwise \end{cases} \tag{3.7}$$

In this equation, w is vector for real valued weights, x is vector of inputs and b is bias which is an element that modifies the input boundary without being dependent on the input value. According to number of Perceptrons, the total summation is given in Eq 3.8.

$$\sum_{i=1}^{m} w_i x_i \tag{3.8}$$

Here, m is the number of inputs to the Perceptrons. Depending on the activation fuction, the output can be represented as '1' or '0' otherwise '1' or '-1'.

Activation Functions of Perceptron
Conversion of the numerical output into '+1' or '-1' is known as step rule. The activation function uses this step rule to determine whether the weighting function's output is larger than zero. The Sign function outputs '+1' or '-1' depending on whether Neuron output is larger than zero. Opposite to that, the Sigmoid Function is an S-curve that produces a value between "0" and "1". Figure 3.5 shows these three activation funtion.

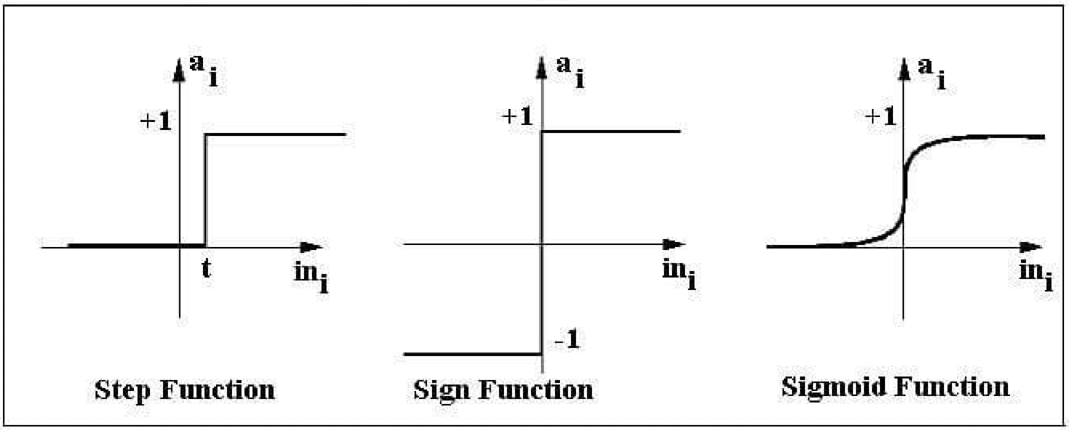

Figure 3.5: Perceptron Activation Function [39]

Multi-layer Perceptron
More than two layers of neurons are required to design the Multi-layer Perceptron as shown in Figure 3.6. The Input Layer and the Output Layer are the main point of focus although it is the hidden layer that generates probability score accurately. Eq 3.9 and Eq 3.10 expresses the output of the MLP network. In this equation [9], h indicates the elements of Hidden Layer, o denotes the elements of Output Layer. $w_{\mathrm{ji}}^{\mathrm{h}}$

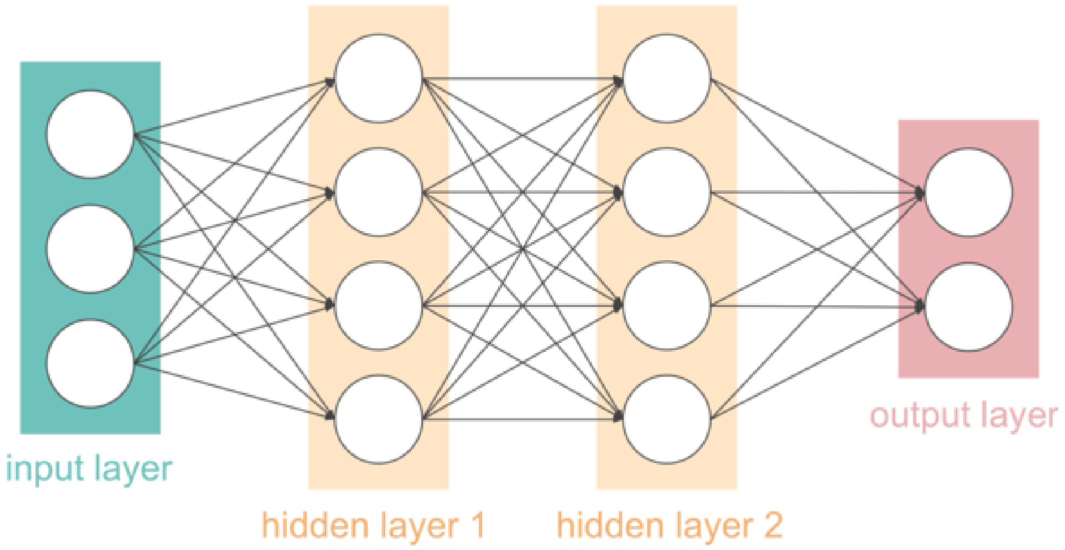

Figure 3.6: Multi-Layer Perceptron [9]

points the weight of the connection that neuron j of the Input Layer with the neuron i of the Hidden Layer. After that, $w^{\mathrm{o}}_{\mathrm{ik}}$ is the weight which relates the Neuron i of the Hidden Layer and the output layer neuron k. $f^{\mathrm{h}}_{\mathrm{i}}$ indicates the transfer function of neuron i of the Hidden Layer, Neuron k of the Output Layer's transfer function is $f^{\mathrm{o}}_{\mathrm{k}}$. $b^{\mathrm{h}}_{\mathrm{i}}$ is the bias of Neuron i of the Hidden Layer, $b^{\mathrm{o}}_{\mathrm{k}}$ is the bias of the Neuron k of the Output Layer.

$$y_k^o = f_k^o(b_k^o + \sum_{i=1}^{S} w_{ik}^o y_j^h) \tag{3.9}$$

$$= f_k^o(b_k^o + \sum_{i=1}^{S} w_{ik}^o f_j^h(b_j^h + \sum_{j=1}^{N} w_{ji}^h x_j)) \tag{3.10}$$

The equation of excitation level shown as $n^{\mathrm{h}}_{\mathrm{i}}$ of Neuron i of the Hidden Layer and $n^{\mathrm{o}}_{\mathrm{k}}$ points the Neuron k of the Output Layer is figured out in Eq. 3.11 and Eq. 3.12.

$$n_i^h = b_i^h + \sum_{i=1}^{N} w_{ik}^h x_j \tag{3.11}$$

$$n_k^o = b_k^o + \sum_{i=1}^{N} w_{ik}^h x_j \tag{3.12}$$

The Training Set, The Validation Set and The Test Set are the three early stopping techniques of data separation that avoid overtraining which is responsible for MLP not to detect the new situations [26]. Backpropagation algorithm and derived algorithms from this are the most applicable error minimizing function among all.

3.3 Long Short Term Memory

Recurrent Neural Networks
Internal memory makes RNN one of the most powerful and widely used neural network. RNNs can recall vital details about the input they received due to their internal memory. This enables them to be extremely exact in anticipating what will occur next. That is why this the most preferable application for time series forecasting, audio, video analysis, speech recognition, weather prediction and many more. Compared with other algorithms, RNNs have deeper understanding of sequence.

RNN vs Feed-Forward Neural Networks
The information in a Feedforward Neural network only flows in one way from the Input Layer to the Output Layer, passing through the Hidden Layers. The information never goes though a same node more than once and moves straight through the network. As Feed-Forward Network does not have memory to receive from data it performs bad for larger amount of dataset. Except its training it can not remember anything about the past. The present input is all that matters to a feed-forward network, it has no concept of temporal order. On contrary to that the information in an RNN cycles via a loop which makes it superior. When the method makes a choice, it examines the current input as well as what RNN has learnt from prior inputs. Figure 3.7 illustrates the difference between the concept of these two models.

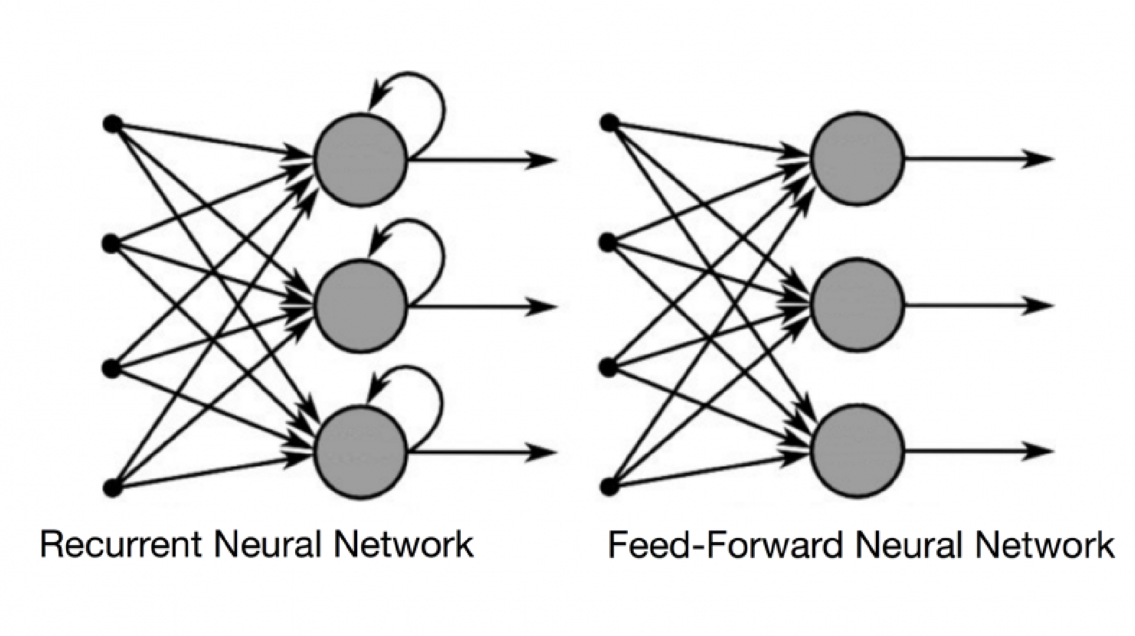

Figure 3.7: RNN vs Feed-Forward Neural Network [37]

Feedforward is a method of transferring information from one place to another. This is common in all other deep learning algorithms where 'Feed-Forward' applies a weight matrix to its inputs before producing the output. But the key for RNN is to apply weights are applied to both the current and previous inputs by RNNs. Through gradient descent and backpropagation through time, a Recurrent Neural Network will adjust the weights. Also, RNNs can map one to many, many to many, and many to one, whereas Feed-Forward Neural Networks map one input to one output as given in Figure 3.8.

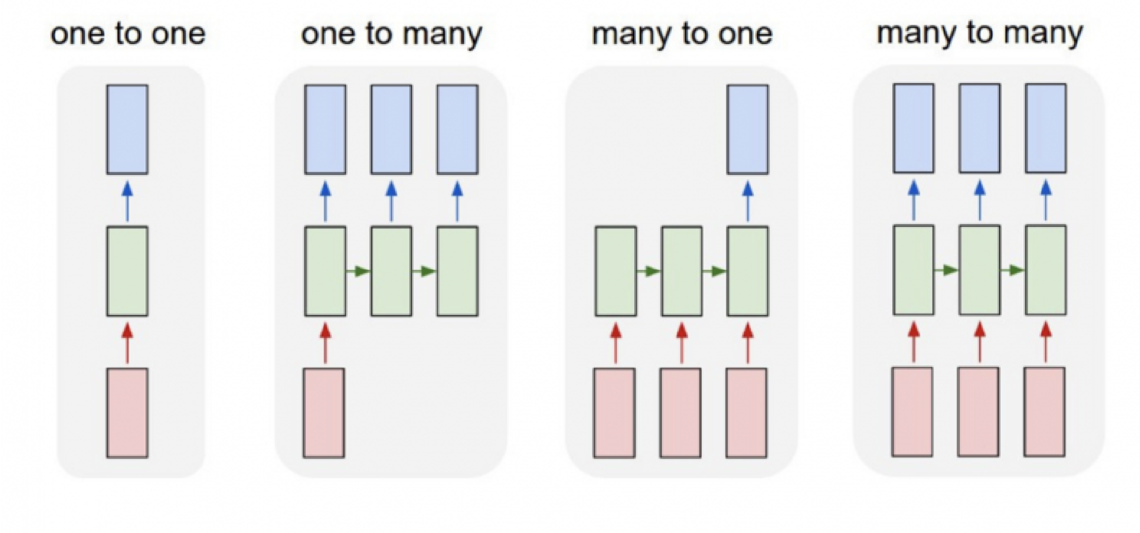

Figure 3.8: Mapping Function [37]

Backpropagation

Backpropagation is a technique for determining The Tradient of an Error Function in relation to the weights of the Neural Network. To calculate the partial derivative of the errors with respect to the weights, the algorithm goes backwards through the several levels of gradients. These weights are then used to reduce error margins during training. Backpropagation is the process of travelling backwards through the Neural Network in order to identify the Partial Derivatives of the error with respect to the weights, also allowing this value to be subtracted from the weights. Then the derivatives are employed by descent of the gradients, a process that can reduce a certain function to a minimum. To find the error, forward-propagation is used to get the model's output and determine if it's right or wrong. Figure 3.9 is the demonstration of propagation in Neural Networks.

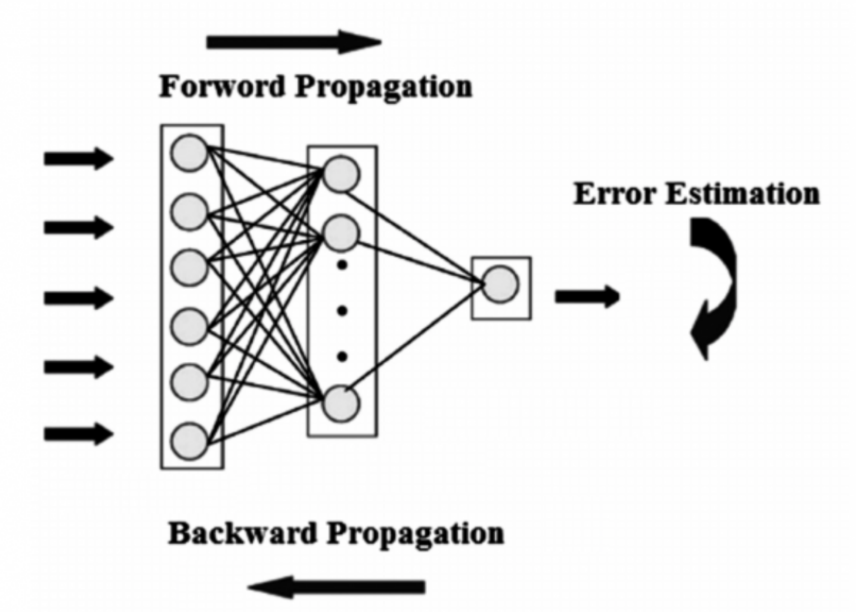

Figure 3.9: Propagation in Neural Networks [37]

Gradient the major issue of standard RNNS
In relation to its input, the gradient is a partial derivative which measures measures how much the output of a function changes with a slide change in input. In short, gradient is the slope of a function. For fast leaning of the model, the gradient should be higher as it steeper the slope. Vanishing Gradient is the major concern for typical RNN as it occurs when the values of a Gradient are too small. This can result in stopping the learning of the model or takes more time to generate result. To solve this problem the concept of LSTM was invented by 'Sepp Hochreiter' and 'Juergen Schmidhuber' [4].

Concept of LSTM
Long Short Team Memory can be explained by using its gates. At first, the cell that works like a path that passes relevant information all the way down the Sequence Chain. This the memory of the Neural Network. Even earlier time information can therefore lead to later times which reduce short-term memory effects. There are also several gates that select which data on the Cell State is permitted. Gates can work out what information is critical to keep and what information to discard while training the data.

Activation Function
The Sigmoid and the Tanh are two Activation Function of the LSTM model. The Tanh Activation is used to regulate the flow of data across the network. The Tanh function compresses values so that they are always between '-1' and '1'. The Sigmoid Activation is comparable to Tanh Activation but in this case, the Sigmoid Activation squishes values between '0' and '1'. The Sigmoid Activation Function is useful for updating or forgetting data because any number multiplied by 0 equals 0. Unwanted values get disappears by Sigmoid Activation Function.

Gates of LSTM
The Input Gate is used to update the Cell State. The former Hidden State as well as the Current Input into a Sigmoid Function are communicated. Inside Input Gate the Tanh function's output gets multiplied with Sigmoid Function's output. Based on the Sigmoid Function's output it is decided that which values of Tanh Function's output should kept. Then The Forget Vector is Pointwise Mltiplied by the Cell State. On the output from the Input Gate, a Pointwise Addition is performed. This technique changes the state of the Cells to new values that the Neural Network deems appropriate. The output gate finally decides what should be the next Hidden State. After that, the Tanh Function's output multiply with the Sigmoid Function's output to choose the data that should be provided by the Hidden State. The Hidden State is the output. After that, the new Cell State and Hidden are carried over to the next time step. The Forget Gate determines whether information should be discarded or saved. The Sigmoid Function passes information from the previous Hidden State as well as information from the Current Input. The results are between '0' and '1'. The closer the value to '0', the more value to forget, and the closer value to '1', the more value to keep.

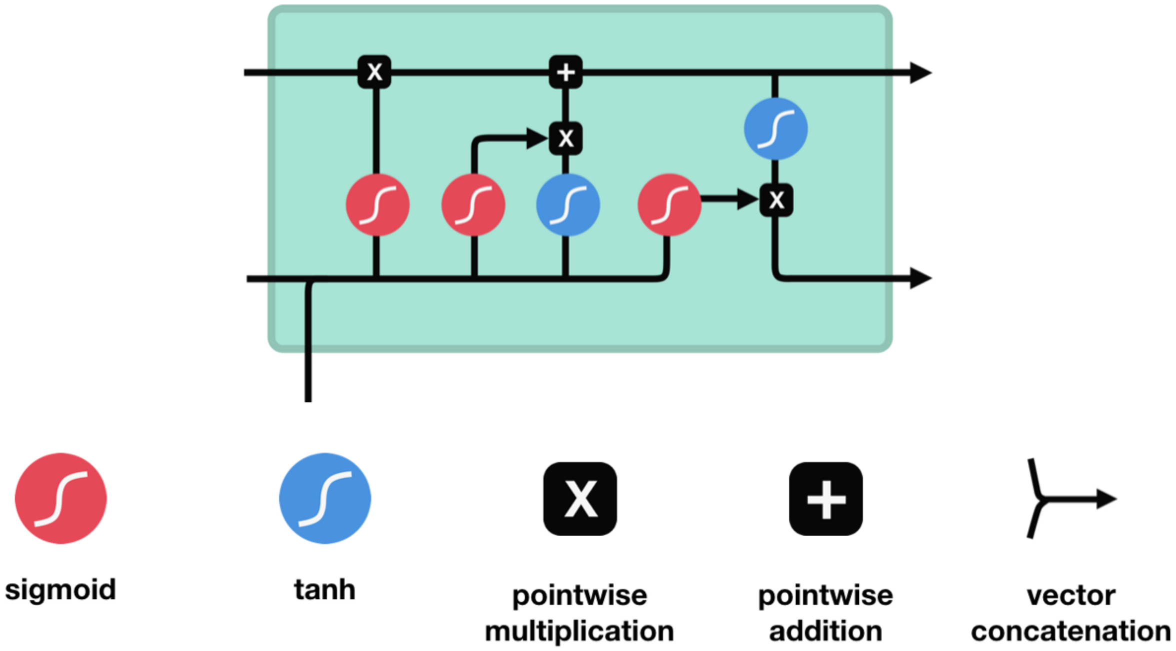

Figure 3.10: Long Short Term Memory Architecture [37]

Chapter 4

Research Methodology

4.1 Dataset

For our research, the dataset about the emission of harmful air substances in the air of Bangladesh has been retrieved from the source World Bank Data Bank. Unfortunately, there was insufficient information in the dataset with lots of missing values Figure 4.1. Therefore, we have chosen a reliable and relevant dataset with 61 entries.

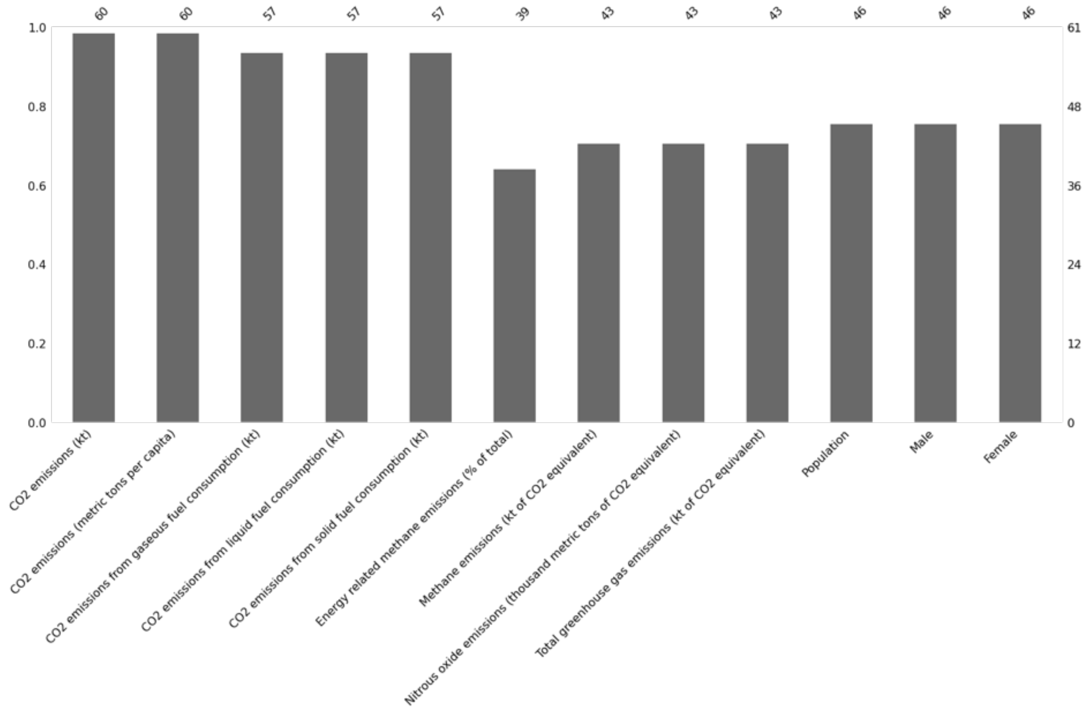

Figure 4.1: Visualization on Raw dataset

4.2 Project Work-Flow

The objective of this workflow is to use The Regression models of Machine Learning algorithms, as it allows predicting data by learning the relationship between the features of the given data. It is a statistical method that predicts continuous values and helps to understand the change of the values of dependent variables corresponding to an independent variable. Our models are Univariate Time Series Regression models for predicting emissions of different harmful air contaminants. Therefore, datasets need to be trained and tested accurately to plot a model for getting a high prediction accuracy score. In this regard, we have built several models such as the Linear Regression model that is a statistical approach for data analysis. Moreover, Artificial Neural Network(ANN) models such as Multilayer Perceptron Model and LSTM(RNN Architecture) models have been implemented to predict the best result considering the accuracy rate. In addition, we have used the performance measuring unit Mean Squared Error value to calculate RMSE values to compare the different model results to find out the best model. Before training the datasets, for statistical models, we have used the method of lag features observations (x_1, x_2, x_3 concatenated as X) as an input variable X and the output variable is y that shows the prediction result on the test dataset. On the other hand, ANN models used the 3 years of data as a sequence block in each epoch for predicting the output y. Henceforth, we have illustrated a comparative prediction-based analysis on harmful contaminants emissions. This process is a supervised learning approach where the goal is to approximate the mapping function efficiently and predict whether there is a new input data X. For the supervised problem, we need to train the given data on both the features and targets. Furthermore, we need to encounter the relationship between independent and dependent features for the regression analysis of our problem. The implementation format of different models in machine learning is almost even in most cases, rather the generalized steps to model the data are as follows:

- Data Pre-Processing
- Exploratory Data Analysis
- Feature Engineering and Selection
- Train and Test Split
- Training and Prediction

Years	int64
‘CO_2 Emissions (kt)’	float64
‘CO_2 Emissions (Metric Tons per Capita)’	float64
‘CO_2 Emissions from Gaseous Fuel Consumption (kt)’	float64
‘CO_2 Emissions from Liquid Fuel Consumption (kt)’	float64
‘CO_2 Emissions from Solid Fuel Consumption (kt)’	float64
‘Energy-Related Methane Emissions (percentage of total)’	float64
‘Methane Emissions (kt of CO_2 equivalent)’	float64
‘Nitrous Oxide Emissions (Thousand Metric Tons of CO_2 equivalent)’	float64
‘Total Greenhouse Gas Emissions (kt of CO_2 equivalent)’	float64
‘Population’	float64

Table 4.1: All the columns of our dataset

4.3 Data Pre-processing

The first and foremost task of our dataset was to impute the missing values for further analysis. Hence, there was the obstacle of insufficient data with huge amount of missing values. As most Machine Learning algorithm cannot work with the missing features, our main target was to prepare accurate and relevant data. We have followed several methods to impute NaN values, dropping the missing values were more convenient way for our work. Firstly, we tried to impute NaN values with special metrics (mean and median) for every column of our dataset. In that case, the accuracy rate was not good. Though our dataset is small, we had to drop all the rows that contain NaN values for better result. After the cleaning process, we had data available of 42 entries from the year 1971 to 2012. The final data was all about the total measurements of different air contaminants emissions such as Carbon Dioxide(CO_2), Methane(CH_4), Nitrous Oxide and Greenhouse Gas. Moreover, we divided each column as separate time series dataframes with single columns to perform Univariate time series data analysis for each type of emission measurement prediction. Besides, in order to check if there were any outliers in the dataset, box plot has been used for each of the columns. An example of an outlier detection has been shown in the Figure 4.2 for Greenhouse gas emissions. Figure 4.3 and Figure 4.4 is the representation of distribution of the features and outlier detection of total COtextsubscript2 and total Greenhouse gas emission. These plots helped our team to visualize the dataset properly which eventually lead to select the model.

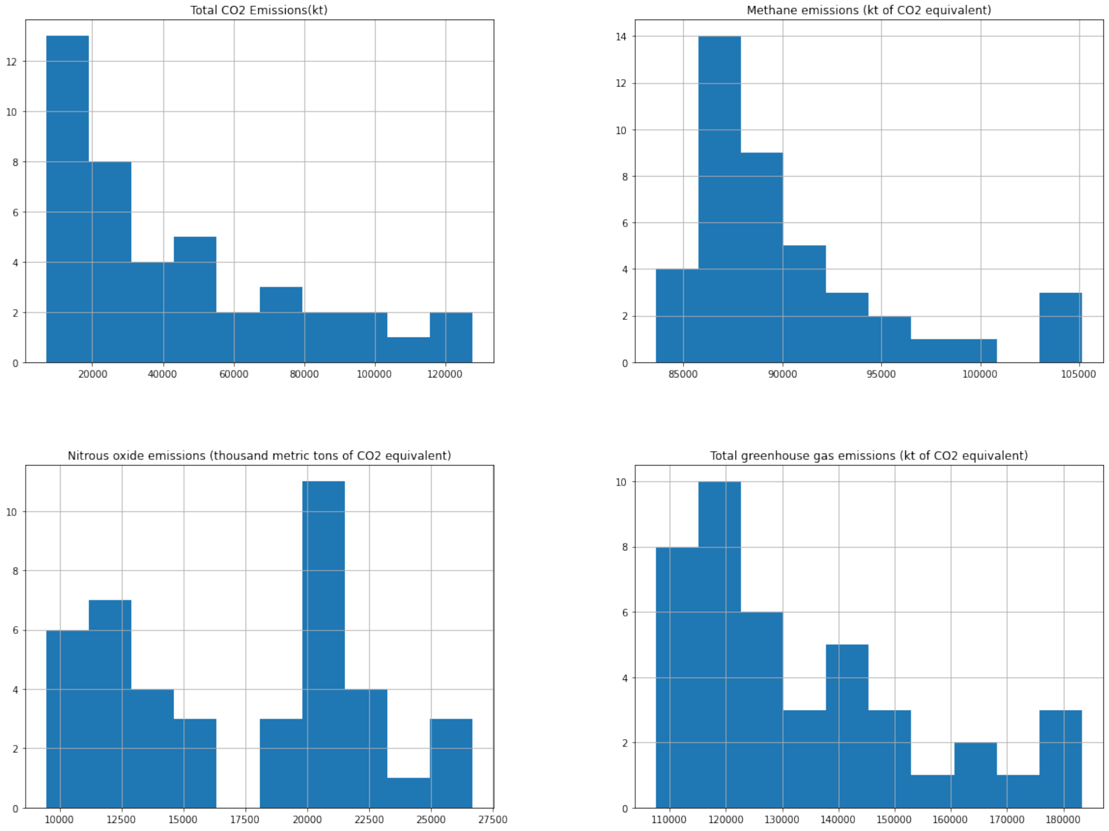

Figure 4.2: Histogram for each numerical attributes of the dataset

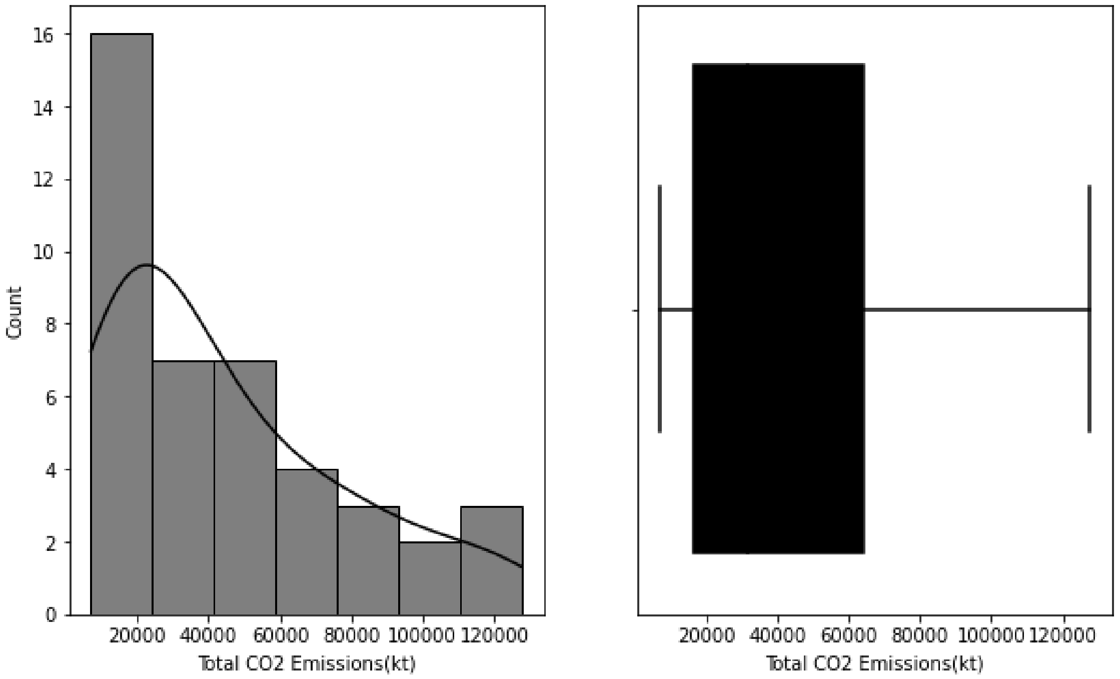

Figure 4.3: Distribution of the Features and Outlier Detection of Total CO_2 emissions

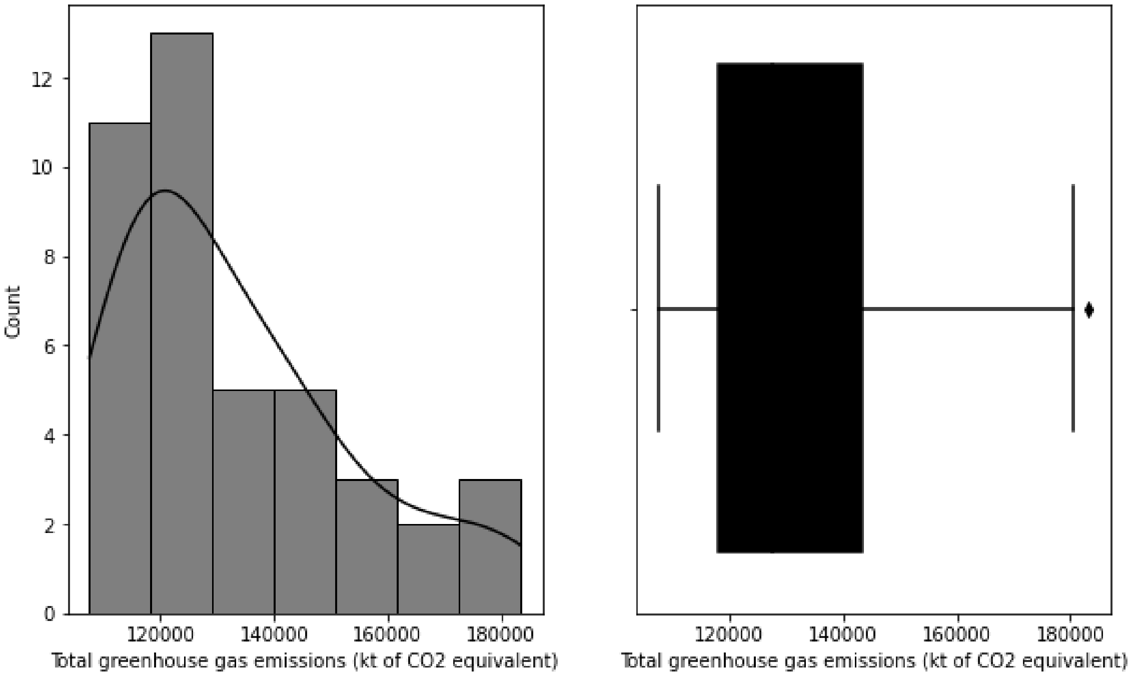

Figure 4.4: Distribution of the Features and Outlier detection of Total Greenhouse emissions

4.4 Feature Selection and Engineering

Feature Selection
Feature Selection reduces the number of Input Variables when developing a Predictive Model. In order to achieve better accuracy for any model, it is essential to identify related features and remove irrelevant or less important features from a set of data that do not contribute to the target variable. Nevertheless, selecting the most relevant features from the dataset depends on many factors. However, the features that have expected to exhibit extensive compatibility with the chosen target and the other less relevant or inaccurate features are removed, which is necessary for the model for understanding the new data.

Feature Engineering
Feature Engineering means selecting raw data and transforming the data into features to get the best possible results from the predictive models. The process is associated with improving the model accuracy on unseen new data. There are some techniques to observe the effect on model performances. For instance, outliers imputation for numerical variables as for some models outliers is a sensitive issue. Therefore, outliers are handled to overcome the problem. Although, in most mathematical operations, logarithm and square roots are used for variable transformation. Additionally, to some extent, some models may require performing feature scaling if the values of the dataset have a massive difference in numerical value ranges. On the other hand, one hot encoding concept or label encoder is used in categorical data transformation. Moreover, ensuring transparency for the machine learning models is a complicated process, as often there are different approaches needed for different types of data.

Feature selection and engineering go through several repetitions as it is an iterative process that is aimed to get an accurate result. Different models necessitate different types of feature engineering techniques, and different accuracy is achieved if before and after validation score is examined. Apart from the dataset's missing values, our data was clean enough, and to scale the numerical value ranges of the features, feature scaling was done with the MinMaxScaling method and there were no outliers to handle. We have generated the correlations between the features by illustrating a Heat-map in Figure 4.5 to visualize the importance of the variables. Through the feature scaling process, the final features were the total 'CO_2 Emissions (kt)', 'CH_4 Emissions (kt of CO_2 Equivalent)', 'N_2O (kt of CO_2 Equivalent)', 'Total Greenhouse Gas Emissions (kt of CO_2 Equivalent)'. Moreover, using each column we have created separate dataframes to perform univariate analysis.

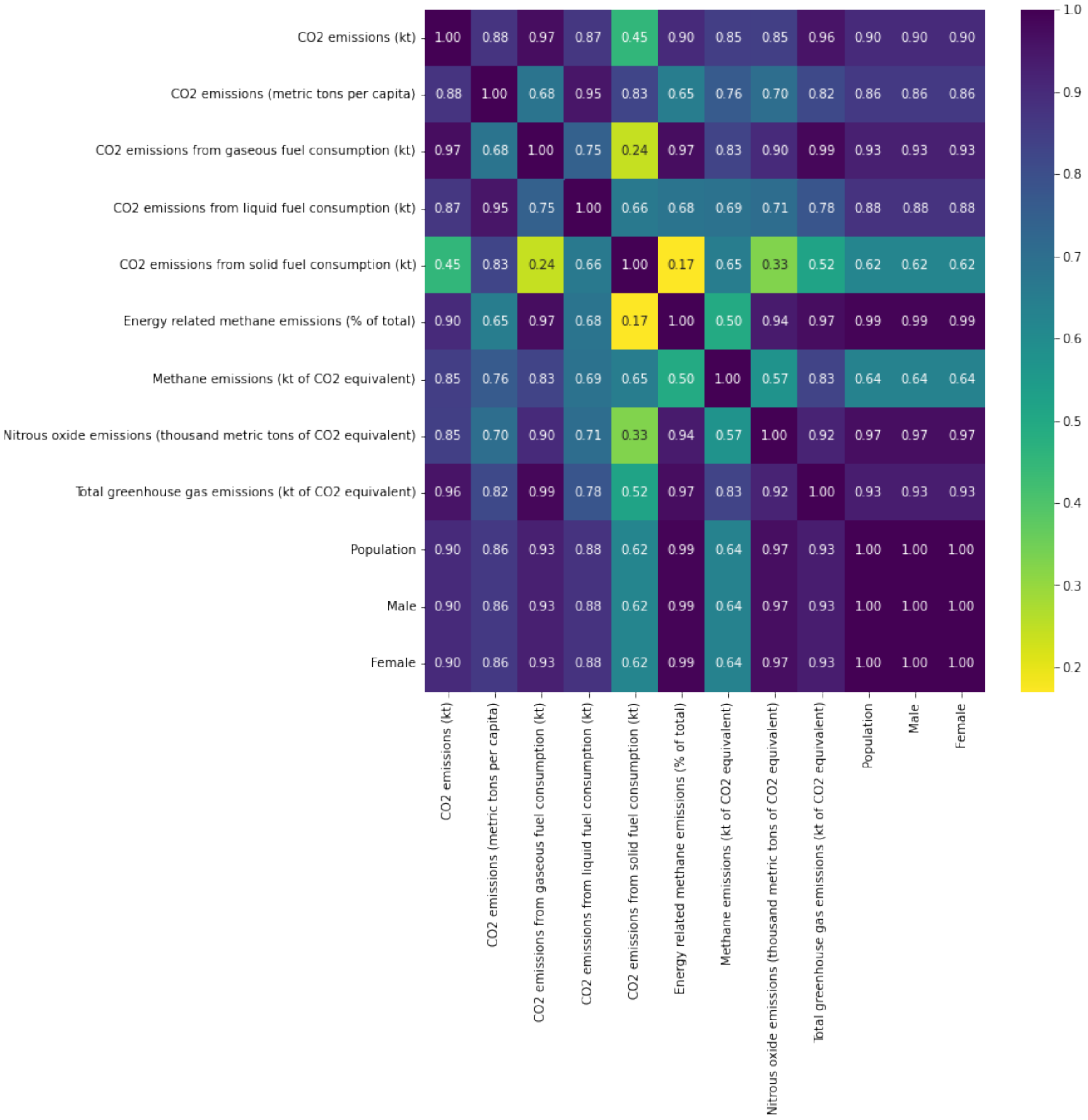

Figure 4.5: Correlation among Numeric Features using Heat-Map

For the Time Series Data Analysis, in the statistical model the Linear Regression and the ANN model MLP, data has been firstly transformed into a supervised dataset. Henceforth, the new input features are created considering the lag features observation technique. We employed the method of lag features observations (x_1, x_2, x_3 concatenated as X) as an input variable for statistical models, and the output variable is y, which provides the prediction outcome on the test dataset. Among the four different dataframes, if we consider the CO_2 emission analysis, then the new variables as lag features, x_1, x_2, x_3 correspond to the observation of the previous years' emission, the past two years' emission data and the past three years of emission data. The sequence of three observations is calculated using the shifting method that is concatenated for the prediction of next time steps. Moreover, for the LSTM model, an RNN architecture, employed the three years of data as a sequence block at each epoch to forecast the output y. The stacked LSTM model has been used for univariate time series data analysis. After completing all the feature engineering and selection, we have the following four separate time-series datasets for training, mentioned in the Table 4.2.

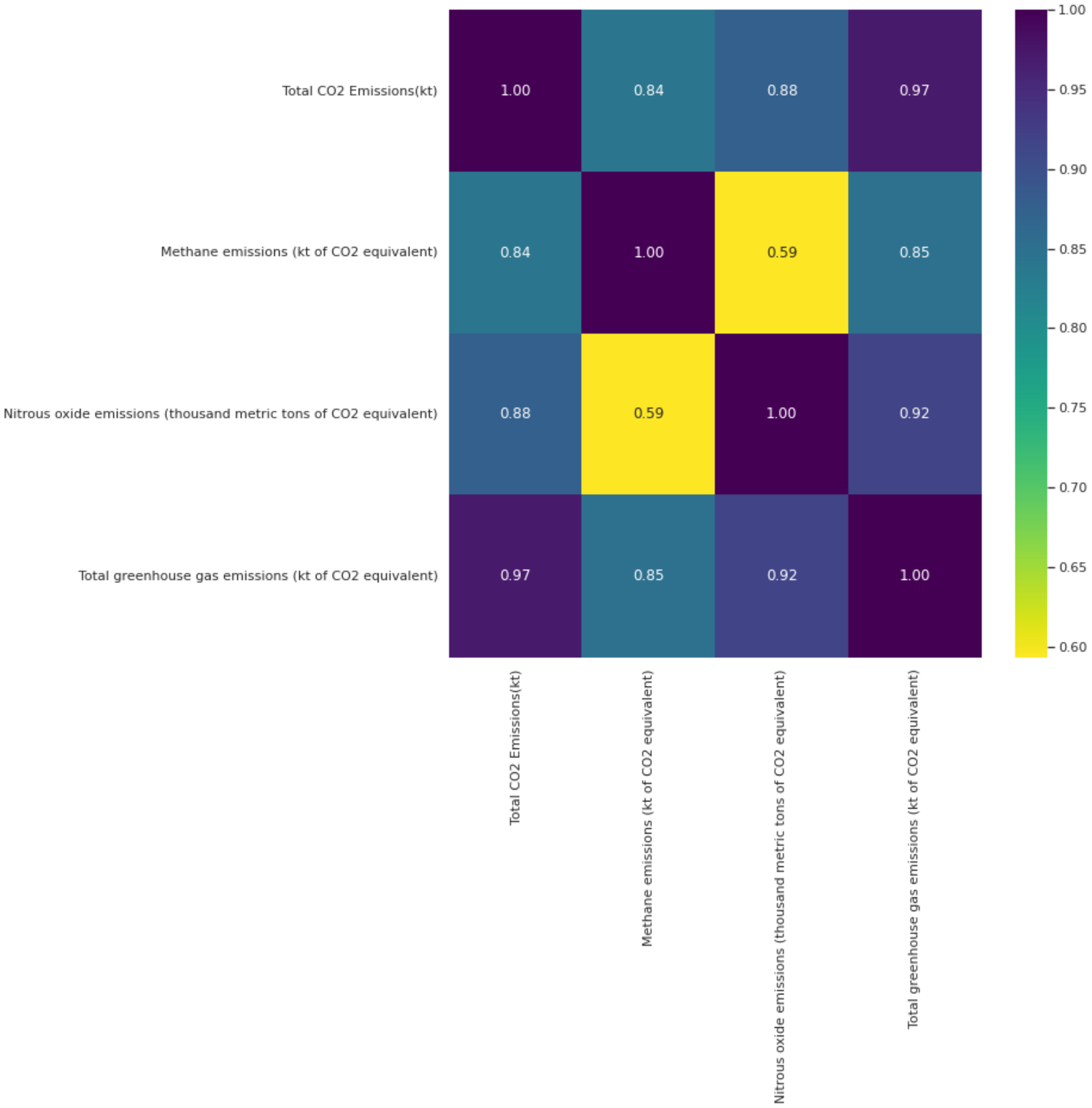

Figure 4.6: Correlation among the numerical features after Pre-processing

‘CO_2 Emissions (kt)’	float64
‘Methane Emissions (kt of CO_2 Equivalent)’	float64
‘Nitrous Oxide Emissions (thousand metric tons of CO_2 Equivalent)’	float64
‘Total Greenhouse Gas Emissions (kt of CO_2 equivalent)’	float64

Table 4.2: All the columns of our dataset after Pre-processing

4.5 Train-Test Split

Usually, data is split into two parts, Training Data, and Test Data. The model learns from the training dataset that contains the known output, and the learning helps to take part in the generalization of the other data later on. Therefore, the training dataset will be used for training our model, and for accuracy evaluation, we will use the test set that is unseen to the training data model. We have used the sklearn library to split each of the datasets into Train and Test, keeping the ratio, 65 percent in the training dataset and the 35 percent in the test dataset. Therefore, we have calculated the accuracy rate and RMSE rate on test datasets for prediction analysis.

Chapter 5

Model Implementation and Optimization

5.1 Work Flow Overview

In order to build the best possible model, it is necessary to choose the correct model space given the training data. As a result, our initial objectives are to find the accuracy result for the test data. Here is a quick rundown of our workflow 5.1. The following are the specifics:

- The raw dataset is cleaned and in the pre-processing part, the missing values have been dropped as imputation method, outliers were checked, and the feature scaling has been done using the MinMaxScalar method.
- Feature Selection and Engineering have been performed. For feature engineering, normalization has been performed for feature scaling and a correlation map has been used for the selection of important features and feature importance as reference. Feature engineering was conducted by outliers detecting using the box plotting visualization method.
- The Train Test split is done in the ratio 65:35, where the test set evaluates the final model.
- The model prediction results have been analyzed.

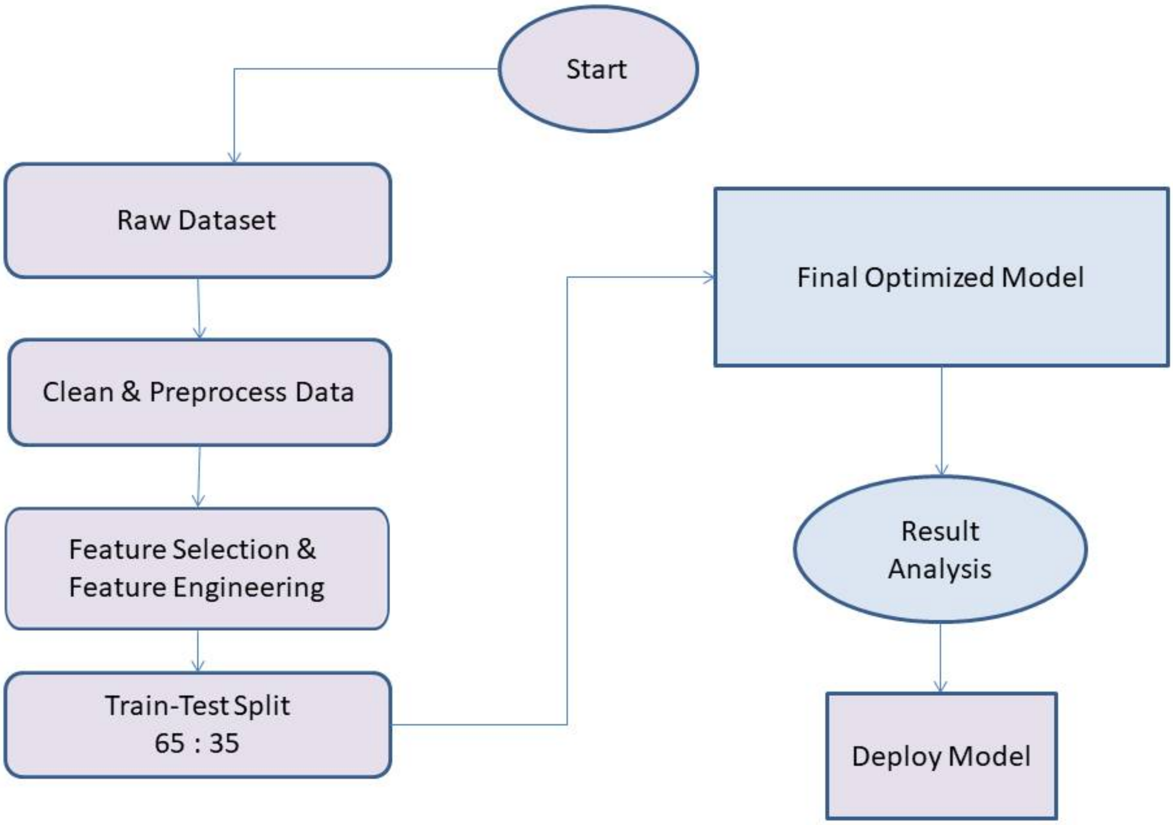

Figure 5.1: Model Workflow

5.2 Evaluating Machine Learning Models

For our supervised regression tasks, we have built, trained, and evaluated different Machine Learning models and analyzed the results which perform better. Three different machine learning models were trained and evaluated using the Scikit-Learn and TensorFlow library.

1. Linear Regression
2. Multi-layer Perceptron(MLP) Regression
3. Long Short -erm Memory, an RNN model for Regression.

5.3 Model Optimization

In machine learning, enhancing a model defines finding the best set of Hyperparameters for a specific problem. The difference between Model Parameters and model Hyperparameters are:

- Model Parameters are the ones that the model learns during training, helps in making predictions.

- Model Hyperparameters are best-considered settings for a Machine Learning algorithm that helps with the learning process that the data scientist tunes before training. For example, the number of Clusters in K-Means, the number of Trees in the Random Forest.

As our dataset is small, we have followed the parameter tuning process that tweaks the parameter values. Machine Learning Algorithms are driven by parameters, and these parameters majorly influence the outcome of the learning process. The main objective of Parameter Tuning is to find the Optimum Value for each Parameter to improve the model's accuracy. There should be a good understanding of these meanings and their impact on the model. Moreover, this process is repeated in well-performing models as well as the models with higher accuracy always do not perform better. For instance, in the MLP algorithm, various parameters such as max_iter, random_state, hidden_layer sizes, etc. have been tweaked several times. For an Intuitive Optimization of these parameter values that result in better and more accurate models.

5.4 Proposed LSTM Model (an RNN architecture)

Overview
The intuition behind using the LSTM model in our research is, it has been found that LSTM models perform better in the case of Time Series Prediction Analysis. This model is a variant of Artificial Recurrent Neural Network(RNN) that outputs a robust model for Predictive Time Series Regression Analysis. Figure 5.2 is a flow of Stacked LSTM Architecture.

Training and Architecture details
The Long Short-Term Memory (LSTM) network is the most common type of RNN architecture for taking input data and producing predictions, such as giving a class label or forecasting a numerical value, such as the next number in the sequence. RNN Architectures are most often used for the time-dependent data and in most of the cases that ended up with better results. Henceforth, we attempted to develop a Time Series Forecasting model with LSTM architecture. For our problem, the LSTM architecture is a Stacked LSTM with memory between batches with Time Steps to solve the Univariate Time Series Forecasting problem. In the case of Time Series Forecasting problems, the dataset is reframed to a supervised learning dataset before building any models. We have considered every three years of data to predict the fourth year's data. Then again considering the data from the second year to the fourth year to predict the fifth year's emission data. This is how the whole prediction process has been performed. Therefore, to calculate the value at the time step t, batches of a single vector for the timesteps (t-1), (t-2), (t-3) has been passed as input X_t at each time step, and that is three dimensional.

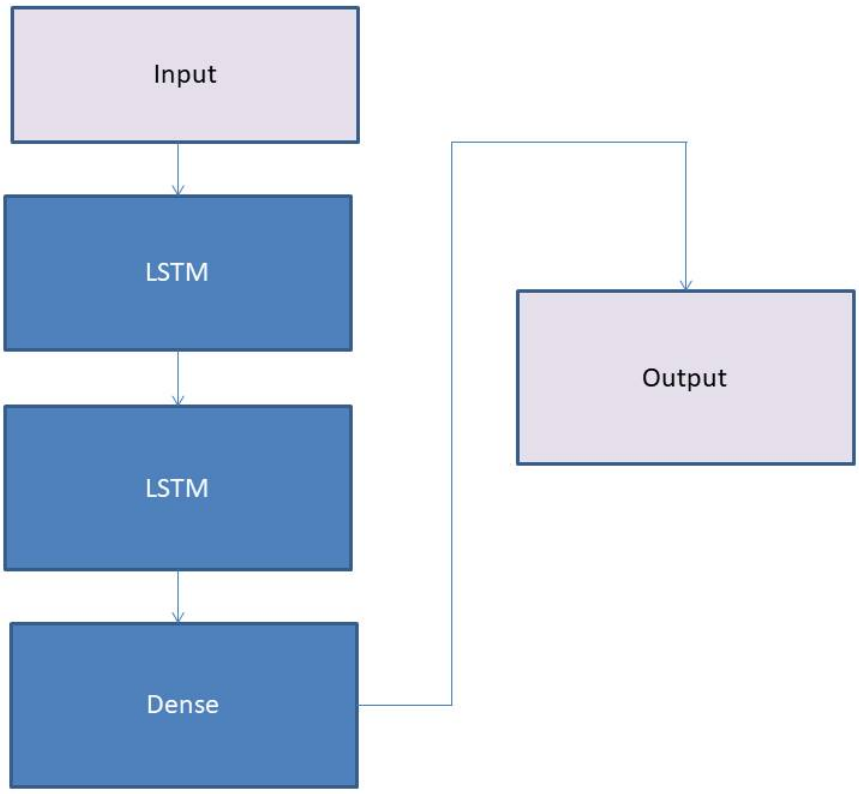

Figure 5.2: A Stacked LSTM Architecture

The data was non-stationary as there was no seasonality present in most cases for each data frame and an increasing trend that means the value is increasing with time from the visualization. Furthermore, in RNN architectures, it does not matter if the data is stationary or non-stationary because the LSTM network can learn from the data's complicated patterns. This is how the sequence of data is being passed as batches in the model for data prediction. 'Rectified Linear Unit (RELU)' Activation Function has been used here in Eq. 5.1.

$$\sigma ReLU(x) = max(x, 0) \tag{5.1}$$

The activation function has an advantage over Sigmoid activation and tanh activation function as it helps reduce the problem of vanishing gradients during back-propagation through the model [5]. To reduce the co-variate shift in the hidden unit values, we employed Batch Normalization [18] between the layers for regularization. Because it has a modest regularization effect, it also avoids over-fitting.
We have chosen Adam optimizer [16] as the model's optimizer since it is currently the most extensively used variation of Gradient Descent Algorithms in Deep Learning models. To fine-tune the model we have tweaked the parameters such as Increased The Number of Batches, The Number of Epochs. Moreover, used multiple layers of LSTM with different weights of neurons were to check the performance of the model considering the RMSE value calculation. For The Regression Problems, RMSE(Root Mean Squared Error) is one of the most used performance measures. This proposed

model has been implemented with Python and TensorFlow library and did not perform better than the other two models.

Chapter 6

Experimental Results Analysis

6.1 Comparative Analysis and Model Evaluation

After fitting the processed dataset into Google Colab, we have analyzed the accuracy and RMSE rate for the models Linear Regression (LR), Multi-layer Perceptron Regression (MLP) and Long Short-Term Memory(LSTM) algorithm. The following performance was observed from the best two models, Linear Regression and Multi-layer Perceptron(MLP) model given in Table 6.1.

Models (Accuracy Score)	CO_2	N_2O	CH_4	Greenhouse
LR	0.99053	0.89204	0.29344	0.93133
MLP	0.99041	0.86421	0.93518	0.92769

Table 6.1: Accuracy Scores of Best Performed models

For better visualization, the Bar Plot of these models accuracy are given in Figure 6.1, 6.2, 6.3, 6.4.

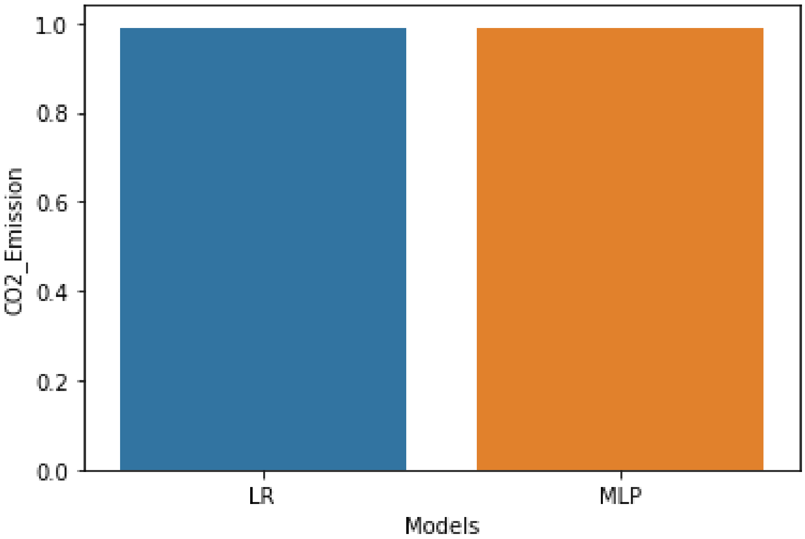

Figure 6.1: Comparison of LR and MLP accuracy for CO_2 Emission

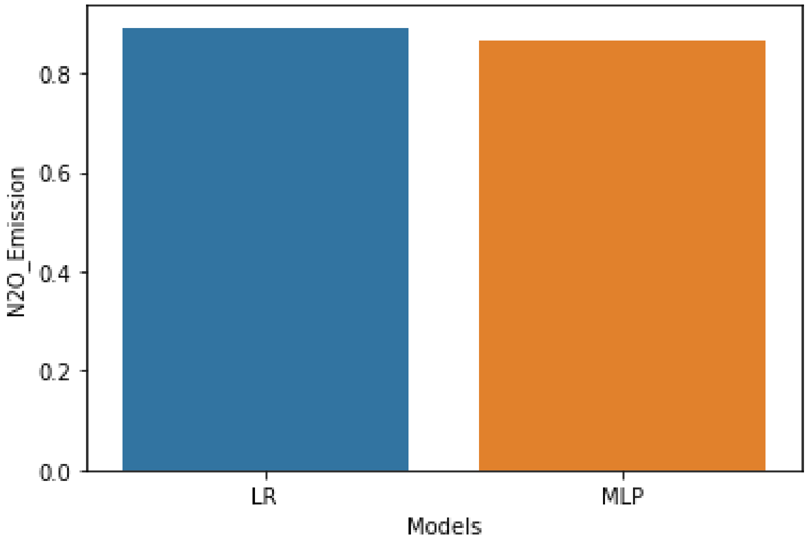

Figure 6.2: Comparison of LR and MLP accuracy for N_2O Emission

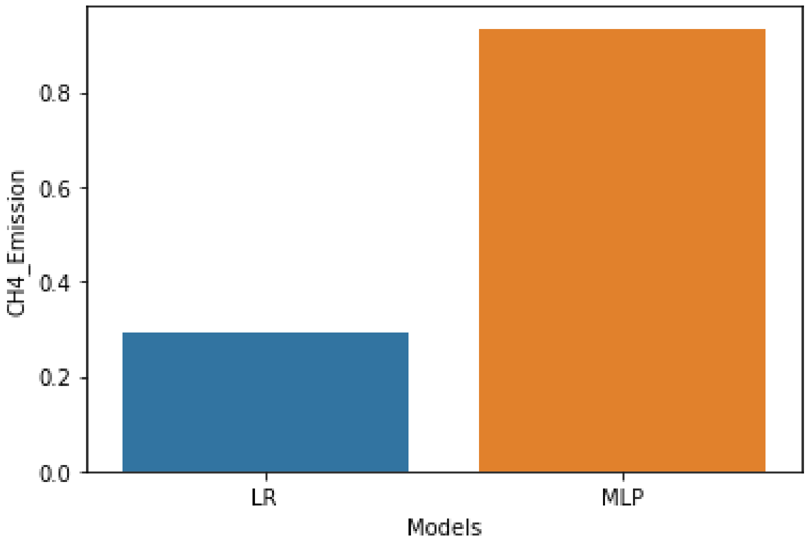

Figure 6.3: Comparison of LR and MLP accuracy for CH_4 Emission

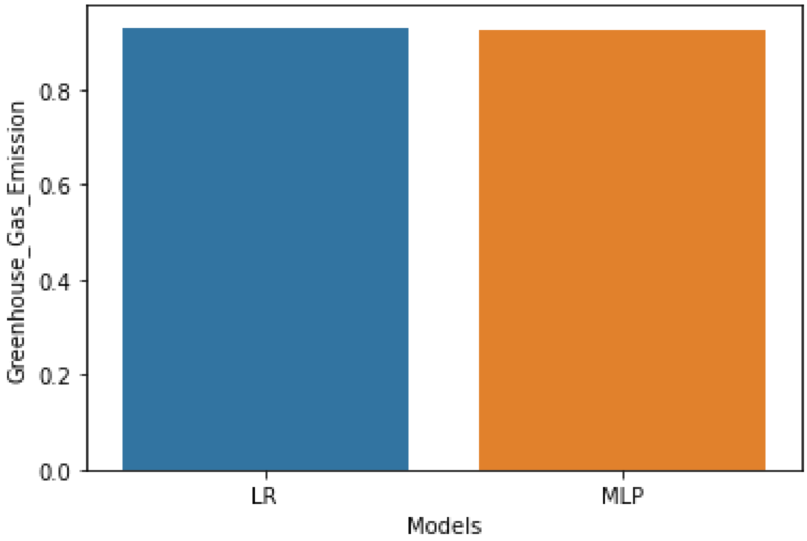

Figure 6.4: Comparison of LR and MLP accuracy for Greenhouse Gas Emission

The following performance was observed for all the models in terms of calculating Error Rate using performance measure RMSE value show in Table 6.2.

Models (RMSE Value)	CO_2	N_2O	CH_4	Greenhouse
LR	2401.9482	719.116	4718.217	4300.916
MLP	2410.167	806.487	1429.072	4413.547
LSTM	282474436	4135.353	10872.973	11001.619

Table 6.2: Comparison of Error Rates of different models

6.2 Model Evaluation

Best Models Evaluation

The models have been fit on the training dataset and evaluated depending on the data of test dataset. Among the three models, Linear Regression, Multi-layer Perceptron(MLP) Regression and LSTM model, Linear Regression and Multi-layer Perceptron(MLP) Regression outperformed on the test dataset. The real values and the models' predictions both had almost similar distribution. Figure 6.5, 6.6, 6.7, 6.8 are the visualization of prediction of LR and MLP.

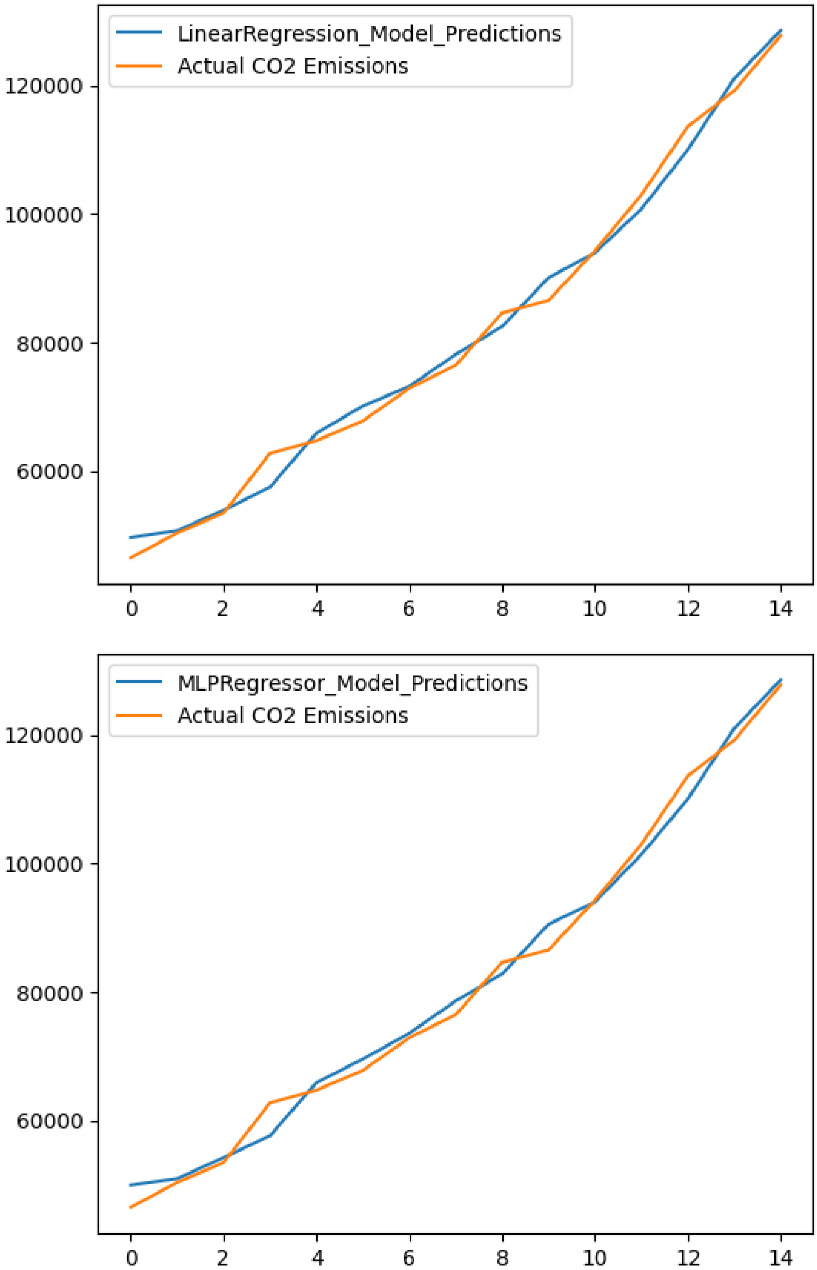

Figure 6.5: Prediction result of best models on CO_2 Emission

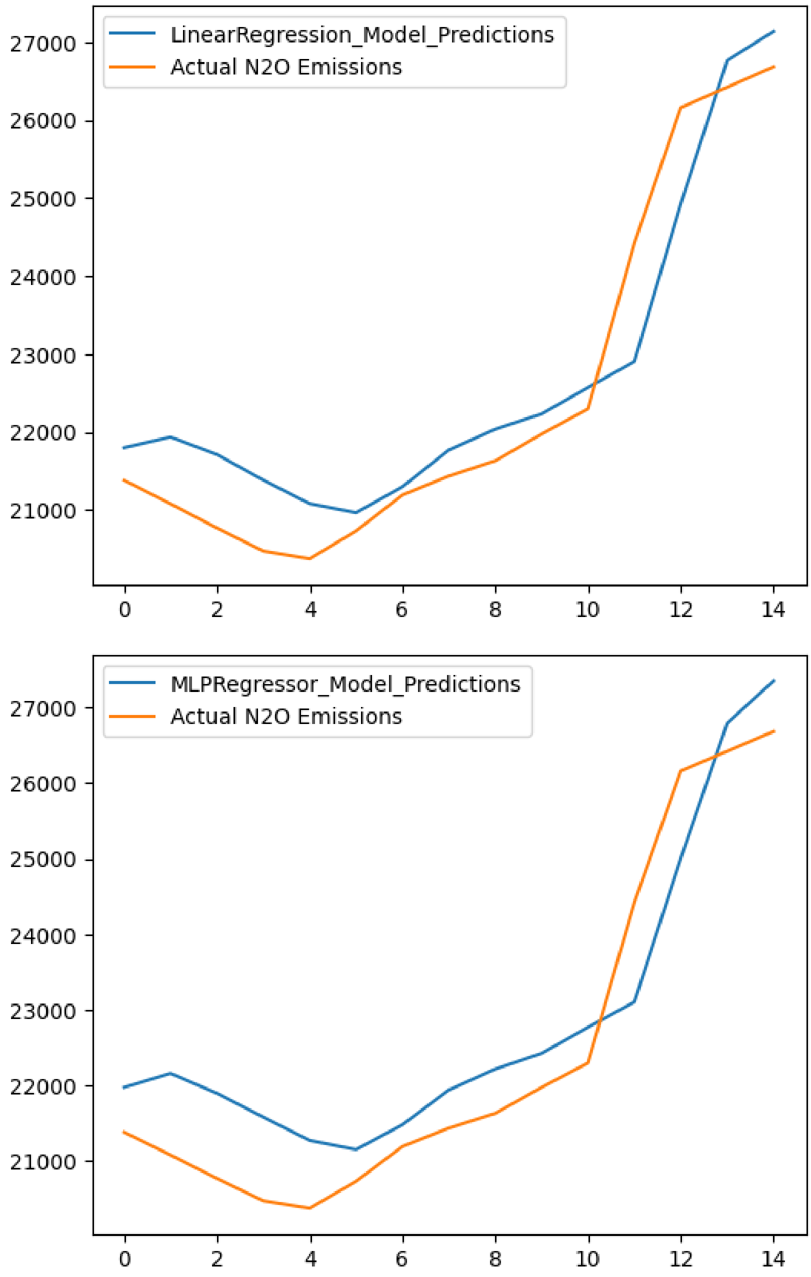

Figure 6.6: Prediction result of best models on N_2O Emission

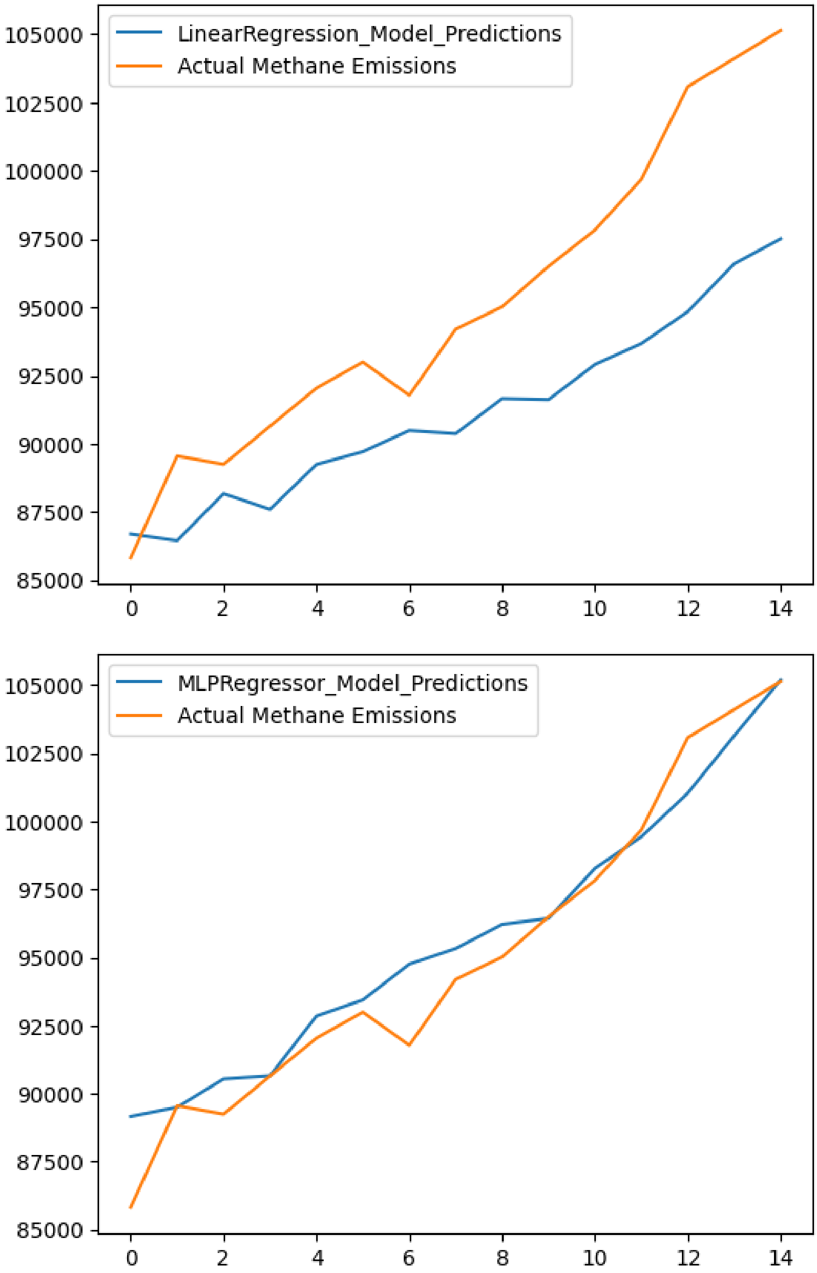

Figure 6.7: Prediction result of best models on CH_4 Emission

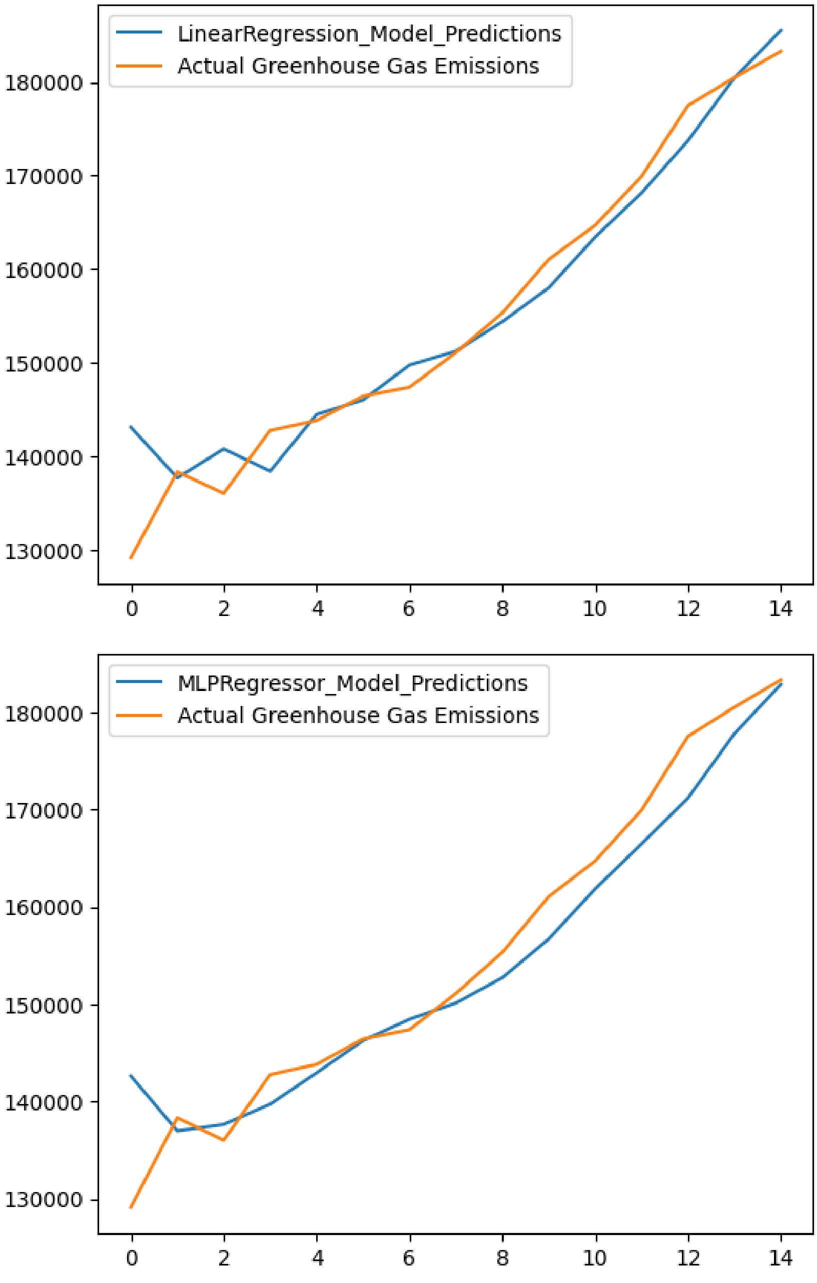

Figure 6.8: Prediction result of best models on Greenhouse Emission

6.3 LSTM Result Analysis

Experimental analysis has been done using LSTM model on each of the separate dataframes to predict the emission of different pollutants. It did not perform better in our small dataset. The model results have been evaluated using RMSE Performance measures.

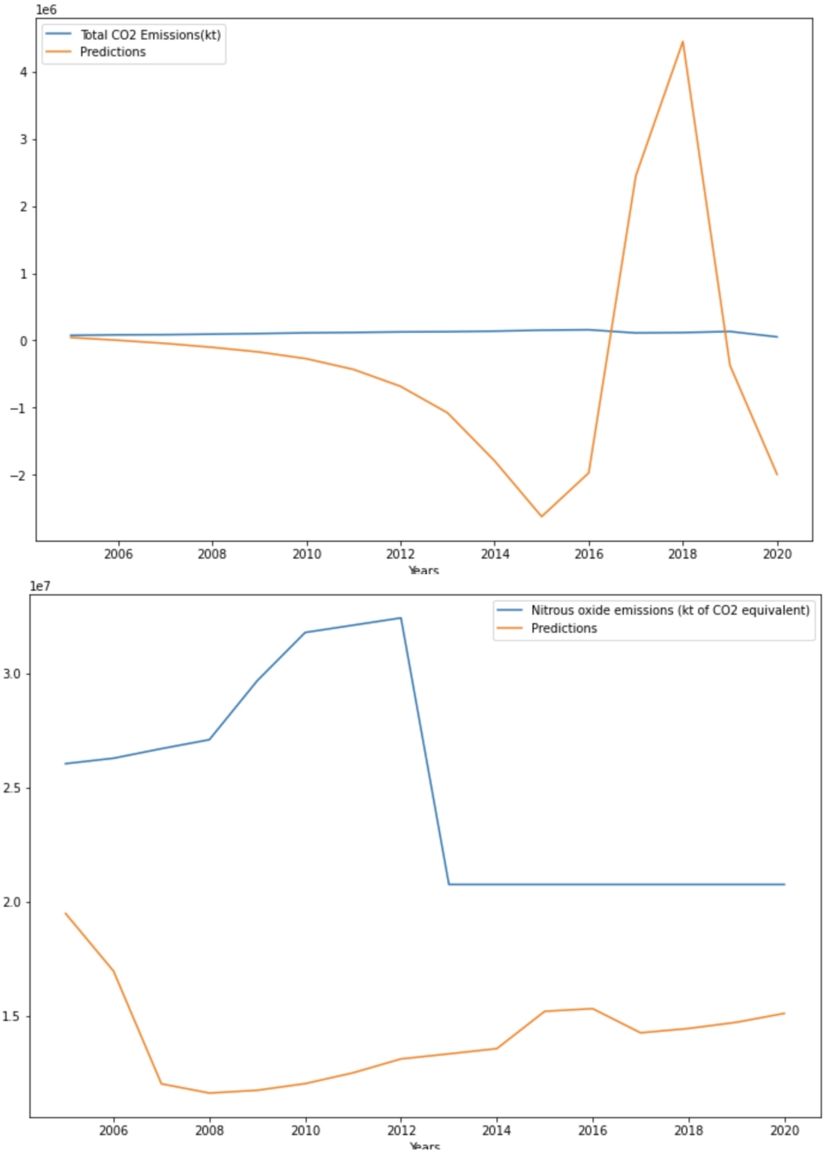

Figure 6.9: Prediction result on CO_2 and N_2O Emissions using LSTM model

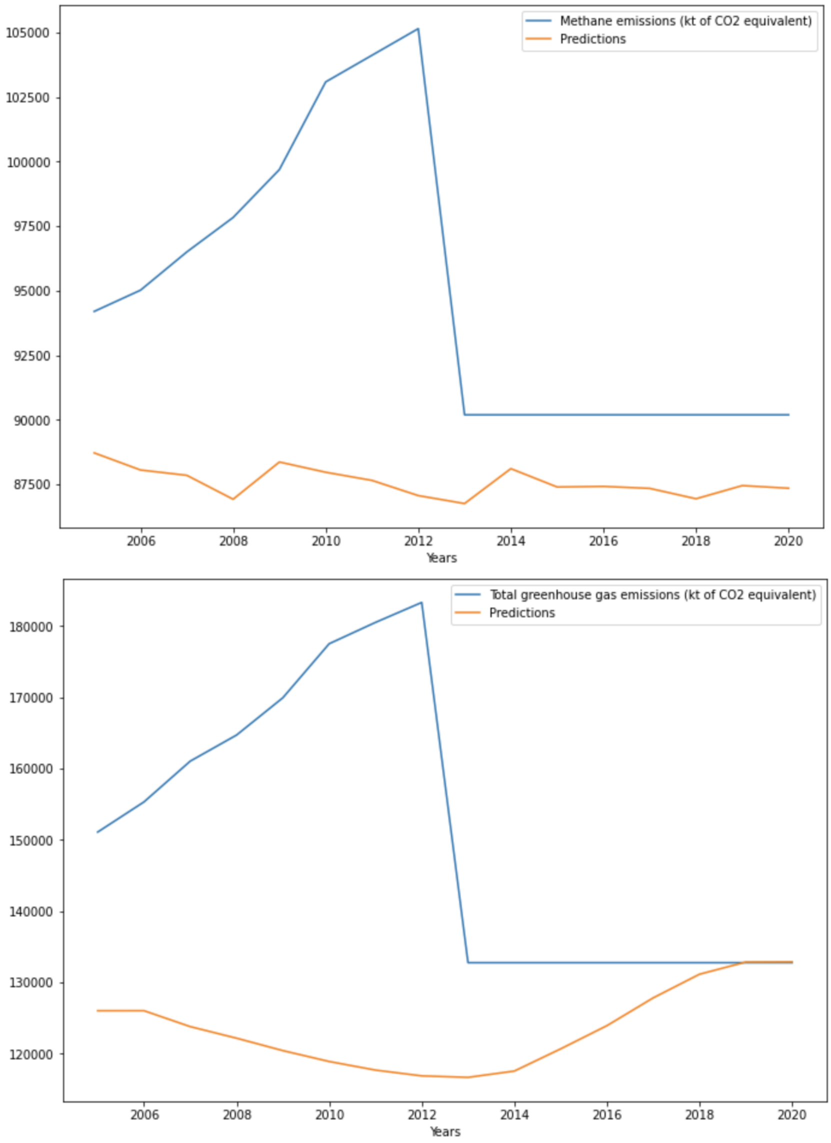

Figure 6.10: Prediction result on CH_4 and Greenhouse gas Emissions using LSTM model

Chapter 7

Conclusion and Future Work

To conclude, on Univariate Time Series Data, we achieved our goal of filtering out the best Supervised Regressor for Predictive Analysis. It is observed from our experiments that the MLP Regressor Model showed the best performance. Moreover, for our tiny dataset, LSTM did not perform better like usual times. To find out the best model, we considered the accuracy level of the implemented models.
Bangladesh is a highly populated country amid a rapid industrialization process. Furthermore, as an agricultural country, the significance of environmental preservation is unavoidable. We discovered how the harmful compounds in the air enlarge in a brief period using Machine Learning models. This finding can guide furthermore work to rescue the air from getting worst. Furthermore, this model can be applied for monthly-basis or even in weekly-basis datasets to generate predictions for smaller time series. Henceforth, this statistical approach can analyze valuable works that need future predictions to sustain or make decisions. Therefore, the implementation of Machine Learning to estimate harmful air contaminant emissions should be a significant concern for future studies. The following works include predictive analysis on stock market exchange prediction, economic growth rate with time, or any univariate time series data analysis.

Bibliography

[1] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain. psychological review," *Psychological Review*, vol. 65, pp. 386–408, 1958. DOI: 10.1037/h0042519.

[2] E. M. Azoff, "Neural network time series forecasting of financial markets," *John Wiley Sons, Inc.605 Third Ave. New York, NYUnited States*, Jun. 1994.

[3] J. D. Hamilton, "Time series analysis," *Princeton university press Princeton*, vol. 02, 1994, ISSN: 0-691-04289-6.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *ResearchGate*, vol. 9, pp. 1735–1780, Dec. 1997. DOI: 10.1162/neco.1997.9.8.1735.

[5] P. F. S. Hochreiter Y. Bengio and J. Schmidhuber, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," 2001.

[6] J. S. Felix A. Gers Douglas Eck, "Applying lstm to time series predictable through time-window approaches," *In: Tagliaferri R., Marinaro M. (eds) Neural Nets WIRN Vietri-01. Perspectives in Neural Computing. Springer, London.*, pp. 193–200, 2002, ISSN: 978-1-4471-0219-9. DOI: 10.1007/978-1-4471-0219-9_20. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4471-0219-9_20#citeas.

[7] W. SC, "Artificial neural network. in: Interdisciplinary computing in java programming," *The Springer International Series in Engineering and Computer Science*, vol. 743, 2003. DOI: 10.1007/978-1-4615-0377-4_5.

[8] S. Y. Edward S Rubin Margaret R Taylor and D. A. Hounshell, "Learning curves for environmental technology and their importance for climate policy analysis," *Energy*, vol. 29, pp. 1551–1559, Aug. 2004, ISSN: 0360-5442. DOI: 10.1016/j.energy.2004.03.092. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544204001392.

[9] G. Agirre-Basurko Ibarra-Berastegi and I. Madariaga, "Regression and multilayer perceptron-based models to forecast hourly o_3 and no_2 levels in the bilbao area," *Environmental Modelling Software*, vol. 21, pp. 430–446, 2006. DOI: 10.1016/j.envsoft.2004.07.008.

[10] S. Dasgupta, "Who suffers from indoor air pollution? evidence from bangladesh," *Health Policy and Planning*, vol. 21, pp. 444–458, Oct. 2006. DOI: 10.1093/heapol/czl027.

[11] K. P. Shiblee M. and C. B., "Time series prediction with multilayer perceptron (mlp): A new generalized error based approach," *Springer, Berlin, Heidelberg*, 2009, ISSN: 978-3-642-03040-6. DOI: 10.1007/978-3-642-03040-6_5. [Online]. Available: https://doi.org/10.1007/978-3-642-03040-6_5.

[12] T. K. I. Arel Derek C. Rose, "Deep machine learning - a new frontier in artificial intelligence research [research frontier].," *IEEE Computational Intelligence Magazine*, vol. 05, pp. 13–18, 2010. DOI: 10.1109/MCI.2010.938364.

[13] S. M. Reza Ebrahimpour Hossein Nikoo and M. S. G. Mohammad Reza Yousefi, "Mixture of mlp-experts for trend forecasting of time series: A case study of the tehran stock exchange," *International Journal of Forecasting*, vol. 27, pp. 804–816, Sep. 2011, ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2010.02.015.. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207010000920.

[14] K. S. Muntaseer Billah Ibn Azkar Satoru Chatani, "Simulation of urban and regional air pollution in bangladesh," *JOURNAL OF GEOPHYSICAL RESEARCH*, vol. 117, Apr. 2012. DOI: 10.1029/2011JD016509.

[15] D. Bereitschaft, "Urban form, air pollution, and co2 emissions in large u.s. metropolitan areas," *The Professional Geographer*, vol. 65, pp. 612–635, Nov. 2013. DOI: 10.1080/00330124.2013.799991.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014.

[17] T. D. Ross Girshick Jeff Donahue and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, pp. 580–587, Oct. 2014.

[18] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015.

[19] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015, ISSN: 0893-6080. DOI: 10.1016/j.neunet.2014.09.003.

[20] "Artificial neural network tutorial - javatpoint," 2016. [Online]. Available: https://www.javatpoint.com/artificial-neural-network.

[21] M. A. Choubin B. Khalighi-Sigaroodi S. and K. Ö, "Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals," *Hydrological Sciences Journal*, vol. 61, pp. 1001–1009, 2016. DOI: 10.1080/02626667.2014.966721.

[22] B. Z. Hossein Hosseini Sreeram Kannan and R. Poovendran, "Learning temporal dependence from time-series data with latent variables," *IEEE International Conference on Data Science and Advanced Analytics*, Oct. 2016. DOI: 10.1109/DSAA.2016.34.

[23] R. R. Kavitha S Varuna S, "A comparative analysis on linear regression and support vector regression," *Online International Conference on Green Engineering and Technologies (IC-GET)*, 2016. DOI: 10.1109/get.2016.7916627.

[24] H. B. Ke Hu Ashfaqur Rahman and V. Sivaraman, "Hazeest: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors," *IEEE Sensors Journal*, vol. 17, pp. 3517–3525, May 2017. DOI: 10.1109/JSEN.2017.2690975. [Online]. Available: https://ieeexplore.ieee.org/document/7892954..

[25] I. N.-C. Laura Varela-Candamio and M. T. García-Álvarez, "The importance of environmental education in the determinants of green behavior: A meta-analysis approach," *Journal of Cleaner Production*, vol. 170, pp. 1565–1578, Nov. 2017. DOI: 10.1016/j.jclepro.2017.09.214. [Online]. Available: https://doi.org/10.1016/j.jclepro.2017.09.214.

[26] C. K. Chang Y. and W. G., "Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Applied Soft Computing*, 2018. DOI: 10.1016/j.asoc.2018.09.029.

[27] W. X. Fangyi Li Xilin Xiao and K. L. Dawei Ma Zhuo Song, "Estimating air pollution transfer by interprovincial electricity transmissions: The case study of the yangtze river delta region of china," *Journal of Cleaner Production*, vol. 183, pp. 56–66, May 2018. DOI: 10.1016/j.jclepro.2018.01.190. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0959652618302129.

[28] M. Herndon, "Air pollution, not greenhouse gases: The principal cause of global warming," *Journal of Geography, Environment and Earth Science International*, Nov. 2018, ISSN: 2454-7352. DOI: 10.9734/JGEESI/2018/44290.

[29] A. A. Z. Ghaemi and M. Farnaghi, "Lasvm-based big data learning system for dynamic prediction of air pollution in tehran," *Environ Monit*, Apr. 2018. DOI: 10.1007/s10661-018-6659-6. [Online]. Available: https://link.springer.com/article/10.1007/s10661-018-6659-6.#citeas.

[30] X. Z. Wenquan Xu Hui Peng and X. P. Feng Zhou Xiaoying Tian, "A hybrid modelling method for time series forecasting based on a linear regression model and deep learning," *Appl Intell*, vol. 49, pp. 3002–3015, Feb. 2019. DOI: 10.1007/s10489-019-01426-3. [Online]. Available: https://link.springer.com/article/10.1007/s10489-019-01426-3#citeas.

[31] "Air quality data," *U.S. Embassy in Bangladesh*, Sep. 2020. [Online]. Available: https://bd.usembassy.gov/embassy/air-quality-data.

[32] M. HOQUE, "Air pollution should be treated as national crisis: Environment minister," *United News of Bangladesh (UNB)*, Feb. 2020. [Online]. Available: https://unb.com.bd/category/Special/air-pollution-should-be-treated-as-national-crisis-environment-minister/44061.

[33] "Importance of forest regions for maintaining ecological balance," *QS Study*, 2020. [Online]. Available: http://www.qsstudy.com/biology/importance-of-forest-regions-for-maintaining-ecological-balance.

[34] S. Nathanson and P. Scarf, "Air pollution," *International Journal of Forecasting*, vol. Encyclopedia Britannica, Oct. 2020. [Online]. Available: https://www.britannica.com/science/air-pollution.

[35] "Population by country," *Worldometers.info.*, 2020. [Online]. Available: https://www.worldometers.info/world-population/population-by-country.

[36] "World air quality index (aqi) ranking," *AirVisual*, Sep. 2020. [Online]. Available: https://www.iqair.com/us/world-air-quality-ranking.

[37] N. Donges, "A guide to rnn: Understanding recurrent neural networks and lstm networks," Jun. 2021. [Online]. Available: https://builtin.com/data-science/recurrent-neural-networks-and-lstm.

[38] M. Usman and Y.-S. Ho, “Covid-19 and the emerging research trends in environmental studies: A bibliometric evaluation,” *Environmental Science and Pollution Research*, vol. 28, pp. 16 913–16 924, Feb. 2021, ISSN: 1614-7499. DOI: 10.1007/s11356-021-13098-z. [Online]. Available: https://doi.org/10.1007/s11356-021-13098-z.

[39] “What is perceptron: A beginners guide for perceptron,” 2021. [Online]. Available: https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron.

[40] “Forest biodiversity statistics,” *Arannayk.org.*, [Online]. Available: https://www.arannayk.org/index.php?option=com_content&view=article&id=17&Itemid=132.