

## **ABSTRACT**

The outbreak of the Novel Coronavirus or the COVID-19 in various parts of the world has affected the world as a whole and caused millions of deaths. This remains an ominous warning to public health and will be marked as one of the greatest pandemics in world history. This paper aims to provide a better understanding of how various Machine Learning models can be implemented in real-world situations. Apart from the analysis done on the world figures, this paper also analyzes the current trend or pattern of Covid-19 transmission in India. With the help of datasets from the Ministry of Health and Family Welfare of India, this study puts forward various trends and patterns experienced in different parts of the world. The data to be studied has been obtained for 154 days i.e. from January 22, 2020, till June 24, 2020. For future references, the data can be further analyzed, and more results can be obtained.

# TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No
	ABSTRACT	i
	LIST OF FIGURES	v
	LIST OF TABLES	vi
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Proposed Algorithms	1
	1.1.1 Polynomial Regression	1
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>5</b>
	2.1 LITERATURE SURVEY	5
<b>3</b>	<b>AIM AND SCOPE OF THE PRESENT INVESTIGATION</b>	<b>10</b>
	3.1 Purpose	10
	3.2 Scope	10
	3.3 Existing system	10
	3.4 Proposed system	11
	3.5 Software Environment	11
<b>4</b>	<b>EXPERIMENTAL OR MATERIALS AND METHODS; ALGORITHMS USED</b>	<b>18</b>
	4.1 Input Design	18
	4.2 Output Design	18
	4.3 Data Flow Diagram	19
	4.4 UML Diagrams	23
	4.5 Goals	24
	4.6 Modules	24
	4.6.1 Data collection	24
	4.6.2 Expected Spread And Forecasting Validation	25
	4.6.3 Model Comparison With The Seird Model	25
	4.6.4 Covid-19 Forecasting	25
		25

4.6.5	Accurate Forecasting	26
4.7	System Architecture	26
4.8	System Requirements	
<b>5</b>	<b>RESULTS AND DISCUSSION, PERFORMANCE ANALYSIS</b>	<b>28</b>
5.1	Test Plan	28
5.2	Verification	28
5.3	Validation	28
5.4	Basics of Software Testing	28
5.5	Black box Testing	28
5.6	White Box Testing	29
5.7	Types of Testing	29
5.7.1	Unit Testing	29
5.7.2	Integration Testing	29
5.7.3	Functional Testing	30
5.7.4	System Testing	30
5.7.5	Stress Testing	30
5.7.6	Performance Testing	30
5.7.7	Usability Testing	30
5.7.8	Acceptance Testing	31
5.7.9	Regression Testing	31
5.8	Requirement Analysis	31
5.9	Functional Requirements	31
5.10	Non-Functional Requirements	33
<b>6</b>	<b>SUMMARY AND CONCLUSION</b>	<b>35</b>
	<b>REFERENCES</b>	<b>36</b>
	<b>APPENDIX</b>	<b>37</b>
A.	SOURCE CODE	37-42
B.	SCREENSHOTS	42-44

## LIST OF FIGURES

<b>FIGURE NO.</b>	<b>FIGURE NAME</b>	<b>PAGE NO.</b>
1.1	Linear Model, Polynomial Model	3
1.2	Multiple Linear Regression Model	4
4.1	Bubble Chart	21
4.2	Work Flow Diagram	22
4.3	Prediction	23
4.4	System Architecture	26
6.1	No. Of Corona Virus Confirmed Cases	42
6.2	No. Of Corona Virus Confirmed Cases Per Country	43
6.3	Login Page	43
6.4	Welcome Page	44

## LIST OF TABLES

SL NO.	TABLE NAME	PAGE NO.
3.1	Methods	16

# CHAPTER-1

## INTRODUCTION

MACHINE learning (ML) has proved itself as a prominent field of study over the last decade by solving many very complex and sophisticated real-world problems. The application areas included almost all the real-world domains such as healthcare, autonomous vehicle (AV), business applications, natural language processing (NLP), intelligent robots, gaming, climate modeling, voice, and image processing. ML algorithms' learning is typically based on trial and error method quite opposite of conventional algorithms, which follows the programming instructions based on decision statements like if-else. One of the most significant areas of ML is forecasting, numerous standard ML algorithms have been used in this area to guide the future course of actions needed in many application areas including weather forecasting, disease forecasting, stock market forecasting as well as disease prognosis. Various regression and neural network models have wide applicability in predicting the conditions of patients in the future with a specific disease. There are lots of studies performed for the prediction of different diseases using machine learning techniques such as coronary artery disease, cardiovascular disease prediction, and breast cancer prediction. In particular, the study is focused on live forecasting of COVID-19 confirmed cases and study is also focused on the forecast of COVID-19 outbreak and early response. These prediction systems can be very helpful in decision making to handle the present scenario to guide early interventions to manage these diseases very effectively. The most challenging aspect of its spread is that a person can possess the virus for many days without showing symptoms. The causes of its spread and considering its danger, almost all the countries have declared either partial or strict lockdowns throughout the affected regions and cities. Medical researchers throughout the globe are currently involved to discover an appropriate vaccine and medications for the disease. Since there is no approved medication till now for killing the virus so the governments of all countries are focusing on the precautions which

can stop the spread. Out of all precautions, "be informed" about all the aspects of COVID-19 is considered extremely important. To contribute to this aspect of information, numerous researchers are studying the different dimensions of the pandemic and produce the results to help humanity.

## 1.1 PROPOSED ALGORITHMS

### 1.1.1 Polynomial Regression

- Polynomial Regression is a regression algorithm that models the relationship between a dependent(y) and independent variable(x) as nth degree polynomial. The Polynomial Regression equation is given below:

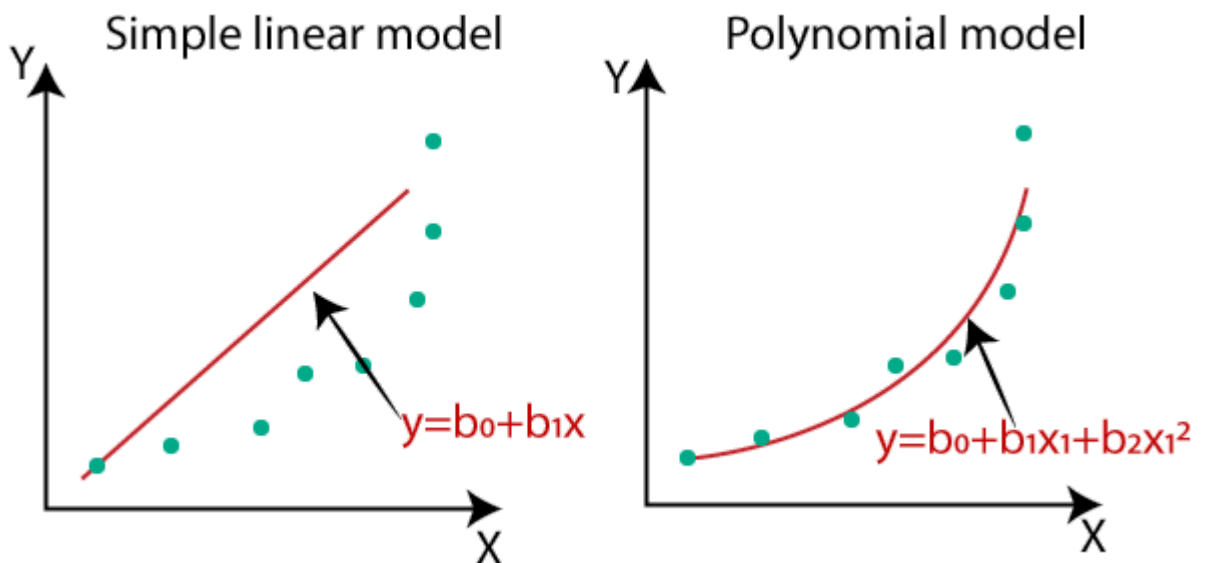
$$y = b_0 + b_1x_1 + b_2x_1^2 + b_3x_1^3 + \dots + b_nx_1^n$$

- It is also called the special case of Multiple Linear Regression in ML. Because we add some polynomial terms to the Multiple Linear regression equation to convert it into Polynomial Regression.
- It is a linear model with some modification in order to increase the accuracy.
- The dataset used in Polynomial regression for training is of non-linear nature.
- It makes use of a linear regression model to fit the complicated and non-linear functions and datasets.
- Hence, "In Polynomial regression, the original features are converted into Polynomial features of required degree (2,3,...,n) and then modeled using a linear model."

Need for Polynomial Regression:

The need of Polynomial Regression in ML can be understood in the below points:

- If we apply a linear model on a **linear dataset**, then it provides us a good result as we have seen in Simple Linear Regression, but if we apply the same model without any modification on a **non-linear dataset**, then it will produce a drastic output. Due to which loss function will increase, the error rate will be high, and accuracy will be decreased.
- So for such cases, **where data points are arranged in a non-linear fashion, we need the Polynomial Regression model**. We can understand it in a better way using the below comparison diagram of the linear dataset and non-linear dataset.



**Fig:1.1: Linear Model, Polynomial model**

- In the above image, we have taken a dataset which is arranged non-linearly. So if we try to cover it with a linear model, then we can clearly see that it hardly covers any data point. On the other hand, a curve is suitable to cover most of the data points, which is of the Polynomial model.
- Hence, if the datasets are arranged in a non-linear fashion, then we should use the Polynomial Regression model instead of Simple Linear Regression.



**Note:** A Polynomial Regression algorithm is also called Polynomial Linear Regression because it does not depend on the variables, instead, it depends on the coefficients, which are arranged in a linear fashion.

Equation of the Polynomial Regression Model:

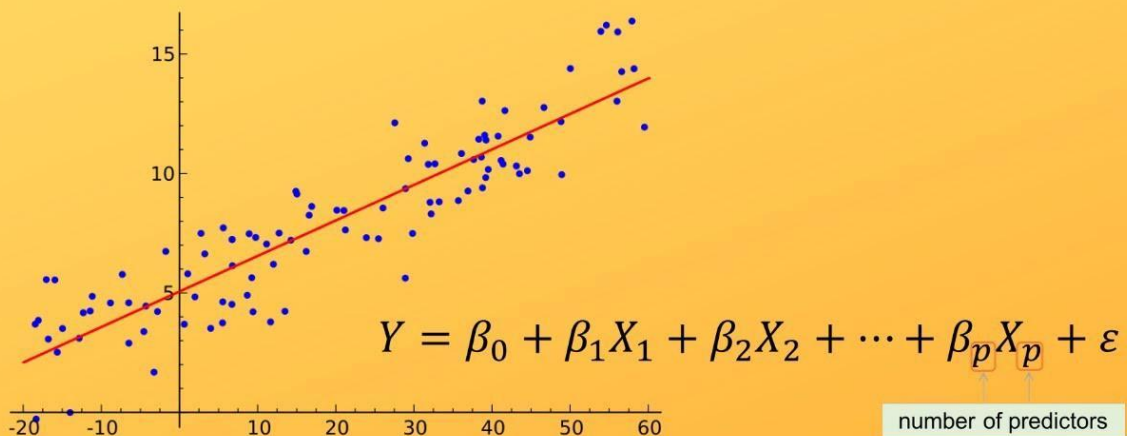
**Simple Linear Regression eq:**  $y = b_0 + b_1x$  .....(a)

**Multiple Linear Regression eq:**  $y = b_0 + b_1x + b_2x_2 + b_3x_3 + \dots + b_nx_n$  .....(b)

**Polynomial Regression eq:**  $y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n$  .....(c)

When we compare the above three equations, we can clearly see that all three equations are Polynomial equations but differ by the degree of variables. The Simple and Multiple Linear equations are also Polynomial equations with a single degree, and the Polynomial regression equation is Linear equation with the nth degree. So, if we add a degree to our linear equations, then it will be converted into Polynomial Linear equations.

## Multiple Linear Regression



***Fig:1.2: Multiple Linear Regression Model***

## **CHAPTER-2**

### **LITERATURE REVIEW**

#### **2.1 LITERATURE SURVEY**

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

## **[1] COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms**

The catastrophic outbreak of Severe Acute Respiratory Syndrome - Coronavirus (SARS-CoV-2) also known as COVID-2019 has brought the worldwide threat to the living society. The whole world is putting incredible efforts to fight against the spread of this deadly disease in terms of infrastructure, finance, data sources, protective gears, life-risk treatments and several other resources. The artificial intelligence researchers are focusing their expertise knowledge to develop mathematical models for analyzing this epidemic situation using nationwide shared data. To contribute towards the well-being of living society, this article proposes to utilize the machine learning and deep learning models with the aim for understanding its everyday exponential behaviour along with the prediction of future reachability of the COVID-2019 across the nations by utilizing the real-time information from the Johns Hopkins dashboard.

## **[2] COVID-19 in India: State-wise Analysis and Prediction**

The highly infectious coronavirus disease (COVID-19) was first detected in Wuhan, China in December 2019 and subsequently spread to 212 countries and territories around the world, infecting millions of people. In India, a large country of about 1.3 billion people, the disease was first detected on January 30, 2020, in a student returning from Wuhan. The total number of confirmed infections in India as of May 3, 2020, is more than 37,000 and is currently growing fast.

## **[3] COVID-19 Outbreak Prediction with Machine Learning**

Several outbreak prediction models for COVID-19 are being used by officials around the world to make informed-decisions and enforce relevant control measures. Among