COVID-19 PREDICTION IN INDIA USING POLYNOMIAL REGRESSION TECHNIQUE IN R

Ryanston Rodrigues
Department Of Statistics, V.P.M.'s B.N. Bandodkar College Of Science, Thane – 400601
ryanstonrodrigues2000@gmail.com

ABSTRACT

The Covid-19 pandemic, also known as the coronavirus pandemic, is an ongoing pandemic of the coronavirus disease 2019 (Covid-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2). The disease was identified in December 2019 in Wuhan, China. As of 16th October 2020, more than 73,72,394 cases have been confirmed with more than 1,12,219 deaths attributed to Covid-19. The rate at which the infection is spreading in India is alarming. This research aims to help in the study of Covid-19 cases in the country by predicting the number of daily infections, recoveries and deaths in the country with the help of polynomial regression technique in R software.

KEYWORDs: polynomial regression,overfitting,

INTRODUCTION

In statistics, polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as a nth degree polynomial in x. It is a special case of multiple linear regression with only one independent variable x,

The second order polynomial in one variable x is given as

$$y = B_0 + B_1x + B_2x^2 + \varepsilon$$

Polynomial regression fits a non-linear relationship between the value of x and the corresponding conditional mean value of y, denoted E(y|x). It is an effective curve fitting technique.

In polynomial regression, we regress a dependent variable in the powers of the independent variable.With the application of this technique, we aim to predict the daily increase in infections, daily recoveries and deaths in the country due to the ongoing Covid-19 pandemic. R, primarily being a statistical software provides the basis for this analysis and helps provide accurate results within a short time.

DATASET FOR EXPERIMENT

The data was collected from the website www.covid19india.org . The data shows the number of people belonging to each of the three groups (Cases Found, Recovered, Deceased) beginning from 1st March ,2020 to October 14th,2020.

Days	Cases Found	Recovered	Deceased
1	0	0	0
2	2	0	0
3	1	0	0
4	22	0	0
5	2	0	0
6	1	0	0
7	3	0	0
8	5	0	0
9	9	0	0

10	15	1	0
11	8	0	0
12	10	0	1
13	10	6	0
14	11	0	1
15	10	3	0
16	14	1	0
17	20	1	1
18	25	0	0
19	27	5	1
20	58	3	0
21	78	0	0
22	69	0	3
23	94	2	2
24	74	15	1
25	86	3	1
26	73	7	5
27	153	25	3
28	136	10	5
29	120	17	3
30	187	35	14
31	309	13	6
32	424	19	6
33	486	22	16
34	560	39	14
35	579	56	13
36	609	43	22
37	485	65	16
38	573	75	27
39	565	96	20
40	813	70	46
41	871	151	22
42	854	186	41
43	758	114	42
44	1243	112	27
45	1031	167	37
46	886	144	27
47	1061	258	26
48	922	273	38
49	1371	426	35
50	1580	388	38
51	1239	419	33
52	1537	703	53
53	1292	394	36

54	1667	642	40
55	1408	484	59
56	1835	442	44
57	1607	585	56
58	1568	580	58
59	1902	636	69
60	1705	690	71
61	1801	630	75
62	2396	962	77
63	2564	831	92
64	2952	911	140
65	3656	1082	103
66	2971	1295	128
67	3602	1161	91
68	3344	1475	104
69	3339	1111	97
70	3175	1414	115
71	4311	1669	112
72	3592	1579	81
73	3562	1905	120
74	3726	1963	137
75	3991	1594	97
76	3808	2234	104
77	4794	4012	120
78	5049	2538	152
79	4628	2482	121
80	6154	3032	146
81	5720	3113	134
82	6023	3131	148
83	6536	3280	142
84	6667	2576	142
85	7111	3285	156
86	6414	3012	150
87	5907	3585	173
88	7246	3434	188
89	7254	3171	176
90	8138	11735	269
91	8364	4303	205
92	8789	4928	222
93	7723	3882	201
94	8815	4531	222
95	9689	3789	259
96	9847	4390	274
97	9472	4770	286

98	10408	5433	297
99	10882	5191	261
100	8536	5171	271
101	9981	5634	272
102	11156	6275	358
103	11135	6044	394
104	11306	7263	388
105	12039	8092	309
106	11405	7358	324
107	10032	10639	396
108	11086	7226	2004
109	13108	6890	341
110	13829	10741	343
111	14740	9029	364
112	15198	13974	308
113	15151	9075	426
114	13560	10879	312
115	15656	10462	468
116	16868	13089	424
117	18205	13983	401
118	18255	10246	381
119	20142	14229	414
120	19610	11631	384
121	18339	13497	417
122	18256	12565	506
123	19429	12064	438
124	21947	19999	378
125	22718	14417	444
126	24018	14746	611
127	23942	15829	421
128	22500	15315	473
129	23147	16836	479
130	25561	19508	492
131	25790	19408	479
132	27762	20289	520
133	27757	19981	543
134	29106	18198	500
135	28178	17683	541
136	29917	20977	587
137	35468	22867	680
138	32607	20646	614
139	34820	17476	676
140	37411	23583	543
141	40235	22730	675

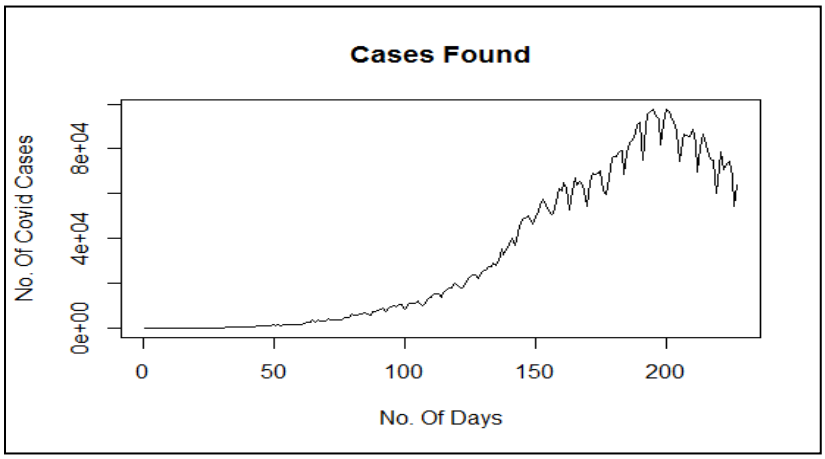
142	36806	24303	596
143	39170	27589	671
144	45601	31875	1130
145	48443	33326	755
146	48888	32514	763
147	50072	37125	703
148	48932	31512	704
149	46484	32354	642
150	49631	35683	774
151	52479	32886	775
152	54969	37425	784
153	57486	36554	764
154	55117	51368	854
155	52672	40355	760
156	50488	43070	806
157	51282	51220	849
158	56626	45583	919
159	62170	50141	899
160	61455	50387	936
161	65156	52135	875
162	62117	54474	1013
163	53016	47362	887
164	61252	56461	835
165	67066	57759	950
166	64141	54776	1006
167	65610	56920	989
168	63986	53116	952
169	58096	57404	952
170	54298	58172	880
171	65024	60455	1099
172	69196	59365	979
173	68518	61873	981
174	69029	62858	953
175	70067	59101	918
176	61749	56896	846
177	59696	66305	854
178	66873	64151	1066
179	75995	56191	1017
180	76827	59620	1066
181	76657	64475	1019
182	78479	64982	943
183	79461	60422	960
184	68766	64435	816
185	78168	62145	892

186	82860	67874	913
187	84156	67491	979
188	87115	69625	950
189	90600	73161	916
190	91723	69624	1008
191	75022	74123	1129
192	89852	74607	1107
193	95529	73057	1168
194	96760	70899	1213
195	97655	81456	1202
196	94414	77862	1111
197	93220	77748	1140
198	81913	79213	1054
199	91098	82849	1281
200	97860	82924	1140
201	96787	87788	1175
202	92973	95515	1221
203	92574	94389	1149
204	87395	92926	1135
205	74493	102070	1056
206	83362	89657	1085
207	86703	87459	1123
208	85921	81142	1144
209	85717	93331	1093
210	88759	92365	1124
211	82770	74691	1040
212	69669	85198	775
213	80500	86150	1178
214	86748	85274	1179
215	81784	78731	1099
216	79883	76339	1068
217	75479	81655	937
218	74770	76713	903
219	60130	75855	886
220	71869	81945	990
221	78809	83209	963
222	70797	78745	967
223	73305	82628	929
224	74418	89024	921
225	67789	71565	813
226	54261	78365	710
227	63717	74079	727

ANALYSIS AND INFERENCE

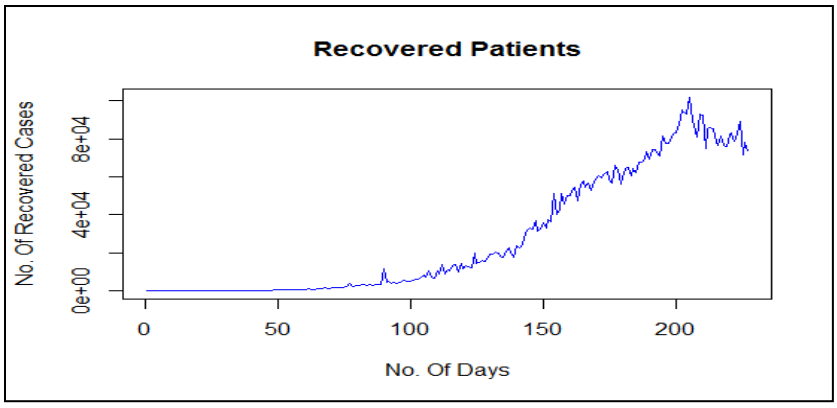
Following is the R code which is used for this experiment.

```
#Covid-19 Research Using Polynomial Regression in R
#Basic Visualizations on the data
#Importing the dataset
dataset=read.csv('Covid_Cases.csv')
plot(dataset$Days,dataset$Cases.Found,main = "Cases Found",
xlab = "No. Of Days",ylab = "No. Of CovidCases",type = "l",col = "black")
```



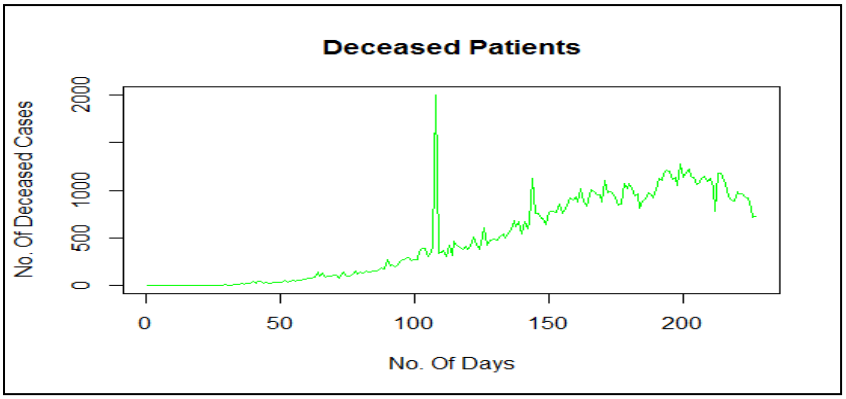
Line Graph of number of daily Covid-19 cases vs number of days.

```
plot(dataset$Days,dataset$Recovered,main = "Recovered Patients",
xlab = "No. Of Days",ylab = "No. Of RecoveredCases",type = "l",col = "blue")
```



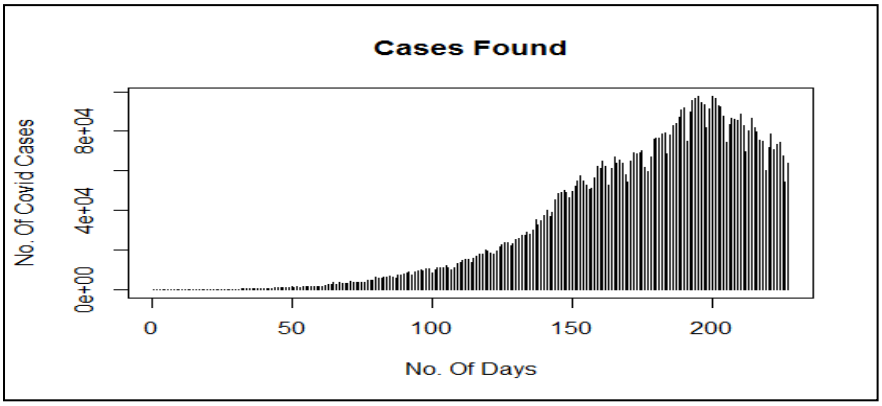
Line Graph of number of daily recovered cases vs number of days.

```
plot(dataset$Days,dataset$Deceased,main = "Deceased Patients",
xlab = "No. Of Days",ylab = "No. Of DeceasedCases",type = "l",col = "green")
```



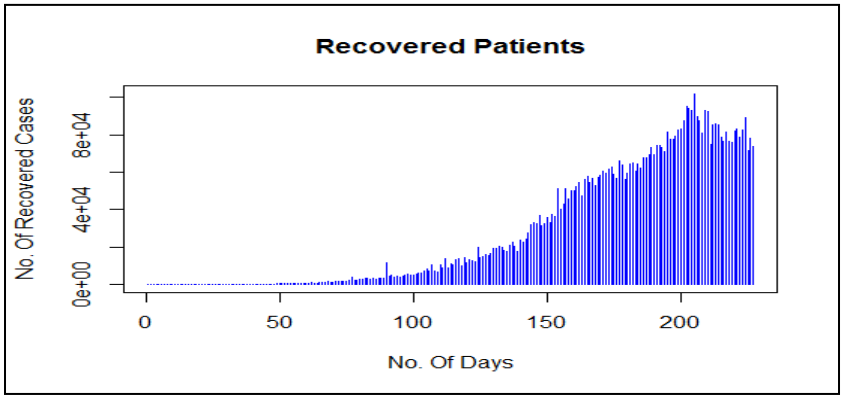
Line Graph of number of deceased cases vs number of days.

```
plot(dataset$Days,dataset$Cases.Found,main = "Cases Found",
xlab = "No. Of Days",ylab = "No. Of CovidCases",type = "h",col = "black")
```



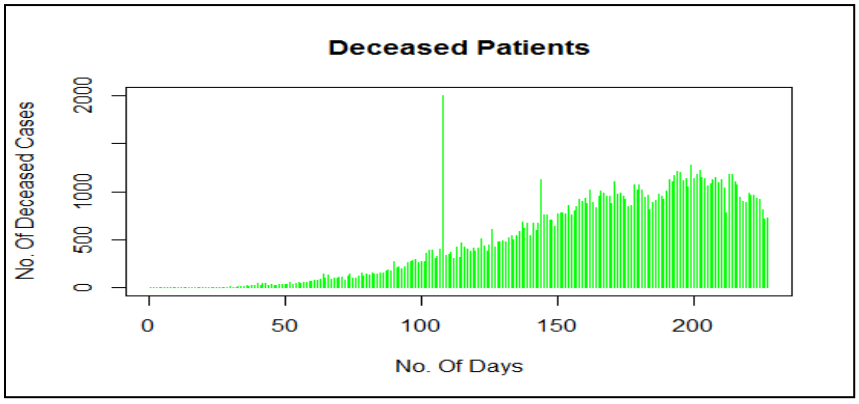
Bar Graph of number of daily cases vs number of days.

```
plot(dataset$Days,dataset$Recovered,main = "Recovered Patients",
xlab = "No. Of Days",ylab = "No. Of RecoveredCases",type = "h",col = "blue")
```



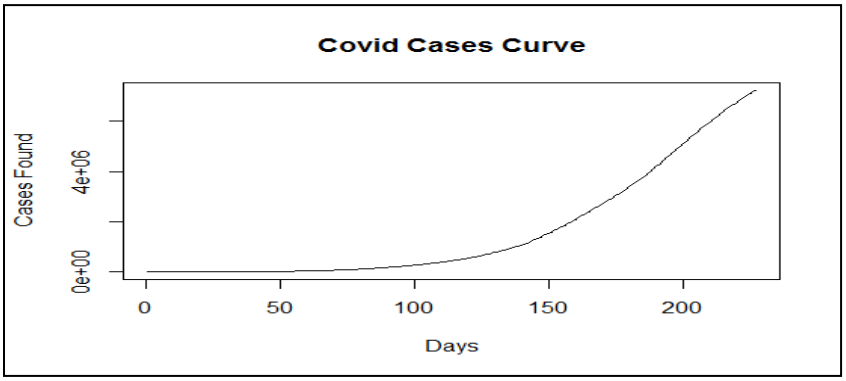
Bar Graph of number of recovered cases vs number of days.

```
plot(dataset$Days,dataset$Deceased,main = "Deceased Patients",
xlab = "No. Of Days",ylab = "No. Of DeceasedCases",type = "h",col = "green")
```



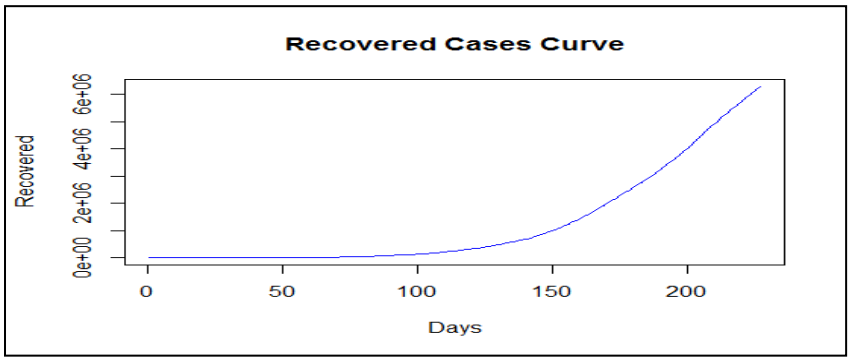
Bar Graph of number of deceased cases vs number of days.

```
#Plotting the Frequency Curve for each type of patient
plot(dataset$Days,cumsum(dataset$Cases.Found),type = "l",col = "black",main ="Covid Cases Curve",
xlab = "Days",ylab = "Cases Found")
```



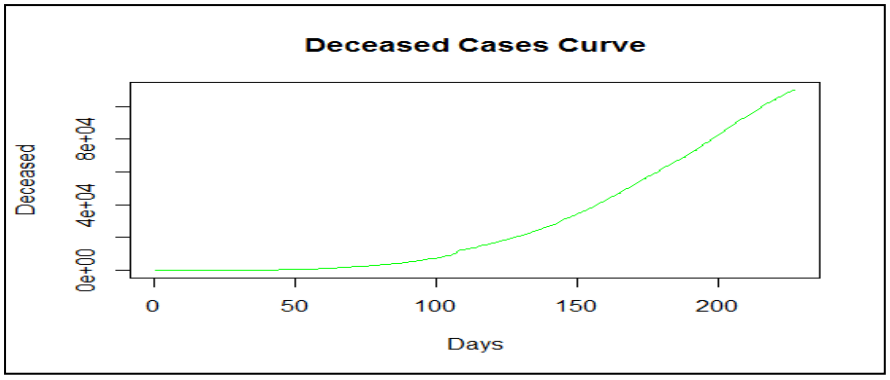
Line Graph of number of cumulative number of daily cases vs number of days.

```
plot(dataset$Days,cumsum(dataset$Recovered),type = "l",col = "blue",main ="Recovered Cases Curve",
xlab = "Days",ylab = "Recovered")
```



Line Graph of number of cumulative number of recovered cases vs number of days.

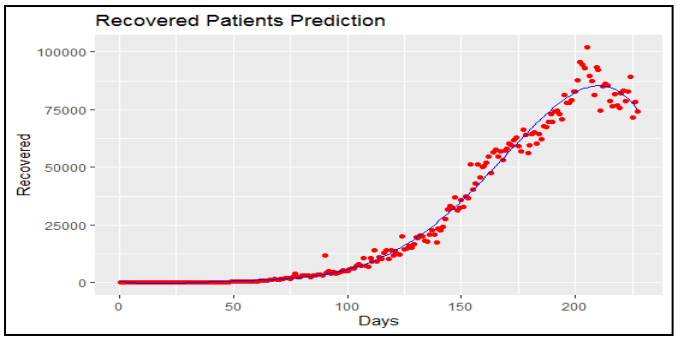
```
plot(dataset$Days,cumsum(dataset$Deceased),type = "l",col = "green",main ="Deceased Cases Curve",
xlab = "Days",ylab = "Deceased")
```



Line Graph of number of cumulative number of deceased cases vs number of days.

```
#Covid 19 Research using Polynomial Regression (Recovered Patients)
#Importing the dataset
dataset=read.csv('Covid_Cases.csv')
dataset=dataset[-c(2,4)]
#Fitting Polynomial Regressor To The Dataset
dataset$Days2=dataset$Days^2
dataset$Days3=dataset$Days^3
dataset$Days4=dataset$Days^4
dataset$Days5=dataset$Days^5
dataset$Days6=dataset$Days^6
poly_reg=lm(formula=Recovered~.,data=dataset)

#Visualising the Polynomial Regressor
library(ggplot2)
ggplot()+
geom_point(aes(x=dataset$Days ,y=dataset$Recovered),
           colour='red')+
geom_line(aes(x=dataset$Days,y=predict(poly_reg,newdata = dataset)),
          colour='blue')+
ggtitle('Recovered Patients Prediction')+
xlab('Days')+
ylab('Recovered')
```



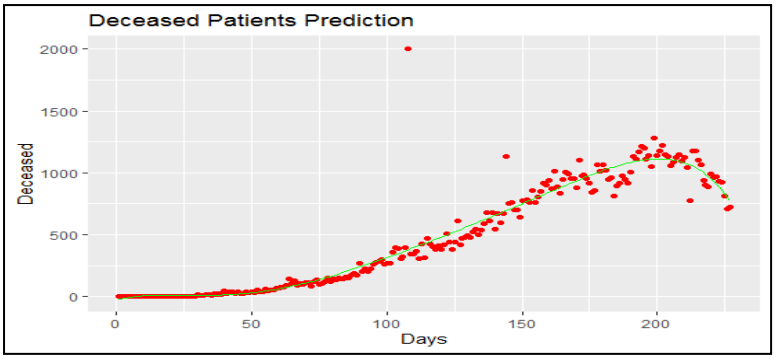
```
#Predicting new value using the Linear Regressor
y_pred1=predict(poly_reg,data.frame (Days=228,
                                     Days2=228^2,
                                     Days3=228^3,
                                     Days4=228^4,
                                     Days5=228^5,
                                     Days6=228^6))

#Predicting number of recovered patients over a span of 6 days
y = 228:233
for (i in y) {
  y_pred1=predict(poly_reg,data.frame(Days=y,
                                      Days2=y^2,
                                      Days3=y^3,
                                      Days4=y^4,
                                      Days5=y^5,
                                      Days6=y^6))
}
y_pred1

#Covid 19 Research using Polynomial Regression(Deceased Patients)
#Importing the dataset
dataset=read.csv('Covid_Cases.csv')
dataset=dataset[-c(2,3)]

#Fitting Polynomial Regressor To The Dataset
dataset$Days2=dataset$Days^2
dataset$Days3=dataset$Days^3
dataset$Days4=dataset$Days^4
dataset$Days5=dataset$Days^5
dataset$Days6=dataset$Days^6
poly_reg=lm(formula=Deceased~.,data=dataset)

#Visualising the Polynomial Regressor
library(ggplot2)
ggplot()+
geom_point(aes(x=dataset$Days ,y=dataset$Deceased),
           colour='red')+
geom_line(aes(x=dataset$Days,y=predict(poly_reg,newdata = dataset)),
          colour='green')+
ggtitle('Deceased Patients Prediction')+
xlab('Days')+
ylab('Deceased')
```

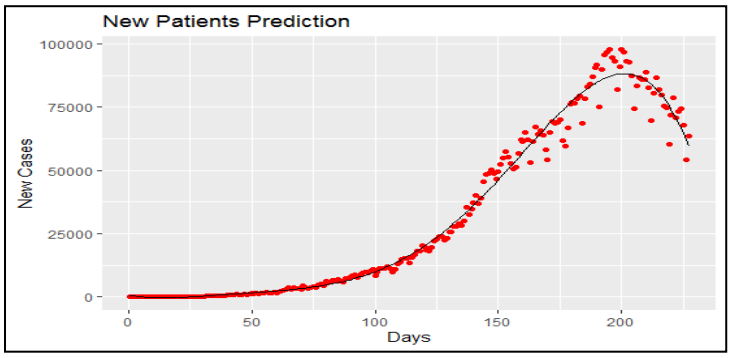



```
#Predicting new value using the Linear Regressor
y_pred2=predict(poly_reg,data.frame(Days=228,
                                    Days2=228^2,
                                    Days3=228^3,
                                    Days4=228^4,
                                    Days5=228^5,
                                    Days6=228^6))
#Predicting number of deceased patients over a span of 6 days
y = 228:233
for (i in y) {
  y_pred2=predict(poly_reg,data.frame(Days=y,
                                      Days2=y^2,
                                      Days3=y^3,
                                      Days4=y^4,
                                      Days5=y^5,
                                      Days6=y^6))
}
y_pred2

#Covid 19 Research using Polynomial Regression(New Cases)
#Importing the dataset
dataset=read.csv('Covid_Cases.csv')
dataset=dataset[-c(3,4)]

#Fitting Polynomial Regressor To The Dataset
dataset$Days2=dataset$Days^2
dataset$Days3=dataset$Days^3
dataset$Days4=dataset$Days^4
dataset$Days5=dataset$Days^5
dataset$Days6=dataset$Days^6
poly_reg=lm(formula=Cases.Found~.,data=dataset)
```

```
#Visualising the Polynomial Regressor
library(ggplot2)
ggplot()+
geom_point(aes(x=dataset$Days ,y=dataset$Cases.Found),
           colour='red')+
geom_line(aes(x=dataset$Days,y=predict(poly_reg,newdata = dataset)),
          colour='black')+
ggtitle('New Patients Prediction')+
xlab('Days')+
ylab('New Cases')
```




```
#Predicting new value using the Linear Regressor
y_pred3=predict(poly_reg,data.frame(Days=228,
                                    Days2=228^2,
                                    Days3=228^3,
                                    Days4=228^4,
                                    Days5=228^5,
                                    Days6=228^6))

#Predicting number of new patients over a span of 6 days
y = 228:233
for (i in y) {
  y_pred3=predict(poly_reg,data.frame(Days=y,
                                      Days2=y^2,
                                      Days3=y^3,
                                      Days4=y^4,
                                      Days5=y^5,
                                      Days6=y^6))
}
y_pred3
```

The trend in Covid-19 patients can be evidently seen from the above graphs. In the beginning, there was an increasing trend in the number of cases. Recently, however the cases have begun to

show a decline. The predict() function is used with given parameters to estimate the number of people in each case. To avoid overfitting, the experiment is carried out at 6 levels.
The for loop in the above code can be used to find out the number of people over the course of 6 days.
The predicted number of people obtained by the code above for the following days (first 6 days) can be tabulated as follows:

Days	Cases Found	Recovered	Deceased
228	57104	73552	743
229	54380	72027	709
230	51513	70391	673
231	48499	68641	634
232	45335	66773	594
233	42016	64785	551

As more data is entered into the dataset, the dataset can be expanded and accordingly, the above code (with minor modifications) can be used to make predictions.

CONCLUSION

Polynomial Regression provides a basis in the process of estimation of Covid-19 patients. Graphs obtained provide a graphical representation for the trend in the cases. Care should be taken while deciding the number of levels in the experiment so as to avoid overfitting or underfitting. A study like this in today's time will prove to be valuablein the fight against the novel coronavirus.

ACKNOWLEDGEMENT

The author would love to thank the Statistics Department, V.P.M.'s B.N. Bandodkar College Of Science for their constant support. The author also gives his heartfelt thanks to Ms. Divya Nairand Ms. Sumaiya Shaikh for critically reading the paper and for their valuable suggestions and corrections.

REFERENCES

Douglas C. Montgomery, Elizabeth A. Beck, G. Geoffrey Vining(2012), Introduction to Linear Regression Analysis, Wiley, Fifth Edition,

Eva Ostertagová(2012), Modelling using polynomial regression, Procedia Engineering 48, 500-506. India COVID-19 tracker. 2020. https://www.covid19india.org/

World Health Organization (WHO), Coronavirus disease (COVID-19) outbreak situation. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.