International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 2995–3015

Analyzing and Predicting COVID-19 Effect using Semantic Web RDF Data

Rupal Gupta[1], Sanjay Kumar Malik[2]

[1]rupal.gupta07@gmail.com

[2]sdmalik@hotmail.com

[1]Research Scholar, USIC&T, New Delhi. [1]Assistant Professor, CCSIT, Teerthanker Mahaveer University, Moradabad.

[2]Professor, USIC&T, New Delhi.

Abstract:

In the present times, Corona virus (COVID-19) has affected the entire world which is of prime significance and needs immediate attention. The need is to analyze and predict about the huge COVID datasets scattered on web which may be in unstructured or semi-structured format. Semantic web has the potential to facilitate it by making this data machine understandable by its RDF technology. In this paper, in regard to this context, COVID-19 RDF data has been generated and analyzed. Further a comparative analysis has been made of different prediction models like linear regression, polynomial regression and Holts Models (Time Series forecasting). The prediction models proposed have considerable significance when it works with RDF dataset which is more flexible and extensible with context to semantic web. For the above, Semantic web technologies, SPARQL and RDF have been used along with python language for implementation. The prediction models have been compared and illustrated in context to COVID spread around the globe as well as in India.

Keywords: Linear Regression, Polynomial Regression, COVID-19 RDF data, Time Series Forecasting (Holts Model), Predictive model.

I. Introduction

COVID-19 is a pandemic which has affected the whole world and there are many countries where the effect of COVID-19 is dangerous and frightening. CORONA Virus is referred as (SARS-CoV-2) the severe acute respiratory syndrome corona virus 2 [1]. The information about COVID-19 infected, recovered and deceased cases around the world has been placed on several official websites of organizations like WHO (World Health Organization) and research group like John Hopkins University (Corona virus resource Centre) [2].

Semantic web is an intelligent web where a machine can understand and interpret the information stored on web and act accordingly [3]. In semantic web, RDF (Resource Description Framework) and OWL (Ontology web language) are two main sources where data and information can be stored. This paper analyzes the existing web data and integrates it with semantic web data. Further the analyses has been made on semantic web data with the

ISSN: 2233-7857 IJFGCN

assumption that in future these available data will resides in semantic web data formats which represents knowledge graphs and can be further utilized in linked open data (LOD).

In this paper, first an RDF data representation has been illustrated for COVID-19 data as well as trends of the COVID has been analyzed that can be summarized into four sections. In section1 the processing framework using a flowchart has been represented and illustrated. In section2 prediction models used for COVID RDF data has been explored. In section3 the analysis of COVID RDF data has been represented graphically in terms of active/recovered/deceased cases of the world as well as of the country India. Further in Section4 all the prediction models used on COVID RDF data have been illustrated and predictions made by the model have been compared with their accuracy measures.

II. Semantic Web Technology

Semantic web is a next generation web in which machine can act, understand and react intelligently. This emerging technology is performing smartly in various domains like social network, natural language processing, linked open data and many more. The data representation on web is smarter than the previous ways of storing it, due to the flexibility and rich taxonomy presentation among the data. RDF OWL and SPARQL are the three key technologies in the semantic web and have got a place in semantic web architecture.

2.1 RDF

RDF stands for resource description framework and is treated as a core method of data representation in Semantic web. RDF has also been treated as a standard for Semantic web by W3C [4]. RDF data provides the flexibility in data representation where data is represented using RDF graphs, which represents a basic data in form of triples. Each triple comprises of a subject, a predicate and an object (S, P, O). RDF data has variety of formats comprises of N-triples, RDF/XML, RDF-a, RDF3X, turtle and N3. These different formats gives in strength to semantic web with RDF data, also it provides more flexibility for working in this domain. RDFS (RDF-Schema) is the schema representation of relationship and rules added and constructed on RDF. Sesame and RDF- Bigdata are commonly used for RDF generation and Jena is a java framework which assists in development and deployment of RDF data with different tools and features.

2.2 OWL

Ontology Web Language is another representation of data in Semantic web, which has a rich taxonomy over RDF and RDF schema. These ontologies can be easily published over the web, where the base of all the rich taxonomy is over RDF / RDFS. It can be considered as a class, object, properties, inheritance and instance like concept for better data representations. For the ontology development a popular tool called Protégé is widely used.

ISSN: 2233-7857 IJFGCN

2.3 SPARQL

SPARQL (Standard protocol and an RDF Query Language) is a query language used to retrieve the data from RDF. It is a standard for querying RDF data and set as a standard by W3C. The SPARQL query is a SQL like query making SPARQL easy to understand. SPARQL query can also be viewed as a BGP (Basic Graph Pattern), through which a SPARQL query can be viewed as a Query Graph. SPARQL Query is a backbone in Semantic web framework because it provides flexibility of data integration from different sources of data like RDBMS and others.

III. Related Work

Gaglione [6] et.al proposed a mathematical model for forecasting the COVID-19 effect in terms of effect of the disease. The proposed model uses Bayesian sequential and adaptive dynamic estimated to predict and track the epidemiological curve with good accuracy on real dataset of regions of Italy and USA.

Khondoker [7] represents that Corona virus is a deadly and infectious virus that can easily be transmitted from one person to another. The best way to control the disease is to home quarantine or self quarantine. In this paper analysis of the real time data of five countries that is Russia, India, Brazil, Bangladesh and United Kingdom has been made with the help of the graphs. Fitting performance of daily detected symptomatic infectious and daily cumulative detected symptomatic infectious cases of all the five countries has been analyzed. Trust region reflective algorithm has been used to predict the corona virus future dynamics. A mathematical Covid-19 model has been proposed for susceptible individuals, exposed individuals, symptomatic infectious individuals, asymptomatic infectious individuals, quarantined, hospitalized, recovered cases and death cases.

Behnood et.al. [8] represent the factors which are responsible for spreading the corona virus like number of people living per unit area and climatic factors. Adaptive network-based fuzzy inference system and virus optimization algorithm has been used to determine the impact of the infectious virus. In warm and humid conditions spread of corona virus was low. Initialization, Replication, and Maintenance are the three main phases of virus optimization algorithm. The machine learning algorithm method ANFIS (Adaptive network-based fuzzy inference system) has been used to explore the climatic factors like temperature, precipitation, humidity and precipitation for the analysis.

Peng and Nahata [9] analyzed Non-linearity in structures, generalization, Hoeffding's inequality and overfitting in machine learning model. For the generalization of data for the future, appropriate pattern has recognized from the data. Some of the methods like deep neural networks and random forest also affect the generalization of data. Support vector regression has been used

ISSN: 2233-7857 IJFGCN

to find optimal solution between bias and variance. For classification and regression job Gaussian kernel and polynomial functions has been used.

Gabriele and M. Gianluca [10] represent the brief overview of evolution of Sars-Cov-2 in Italy. Disease Covid-19 caused by Sars-Cov-2 (Severe Acute Respiratory Syndrome Corona Virus 2). The main purpose of the paper is to find the missing peak. Different models were explored to study the evolution of it. Furthermore a new model similar to classical model was defined to find the downhill of Sars-2 and currently infected provided population is not undertaken into consideration. Graphical representations have been used to evaluate the infected rate, death rate, recovery rate and currently infected rate.

Sayantani and Roy [11] explore Sars-Cov-2, the virus which is responsible for causing Covid-19 and lead to the higher death rates. To understand the evolution of this virus a Long Short-Term Memory (LSTM) based model has been discussed in this paper. Test set, training set and validation set are split for Long Short-Term Memory. Prediction of covid-19 confirmed cases in different countries has been represented with the help of graphs. With the help of prediction model the impact of covid-19 can be easily analyzed by different countries and accordingly further decision can be taken.

Michel [12] focuses on how knowledge graph related to Corona can be available over web. The purpose is to increase the availability of information related to corona virus disease on the web. So that researchers can easily access and process the data set of Covid. Detailed documentation of dataset is available on github repository. Many tools were deployed to analyze the data set. Several resources were used to enrich the CORD-19 dataset like DBpedia Spotlight, Linked Open Data, Knowledge graph, Wikidata, RDF data and many more.

In the continuation, we have made an attempt to propose a prediction model which can be used to predict and analyze the COVID effect using semantic web RDF data, which is one of the most flexible dataset used over the web.

IV. Methodology Used

The Methodology used for applying predictions on COVID RDF data has been shown in figure1.

ISSN: 2233-7857 IJFGCN

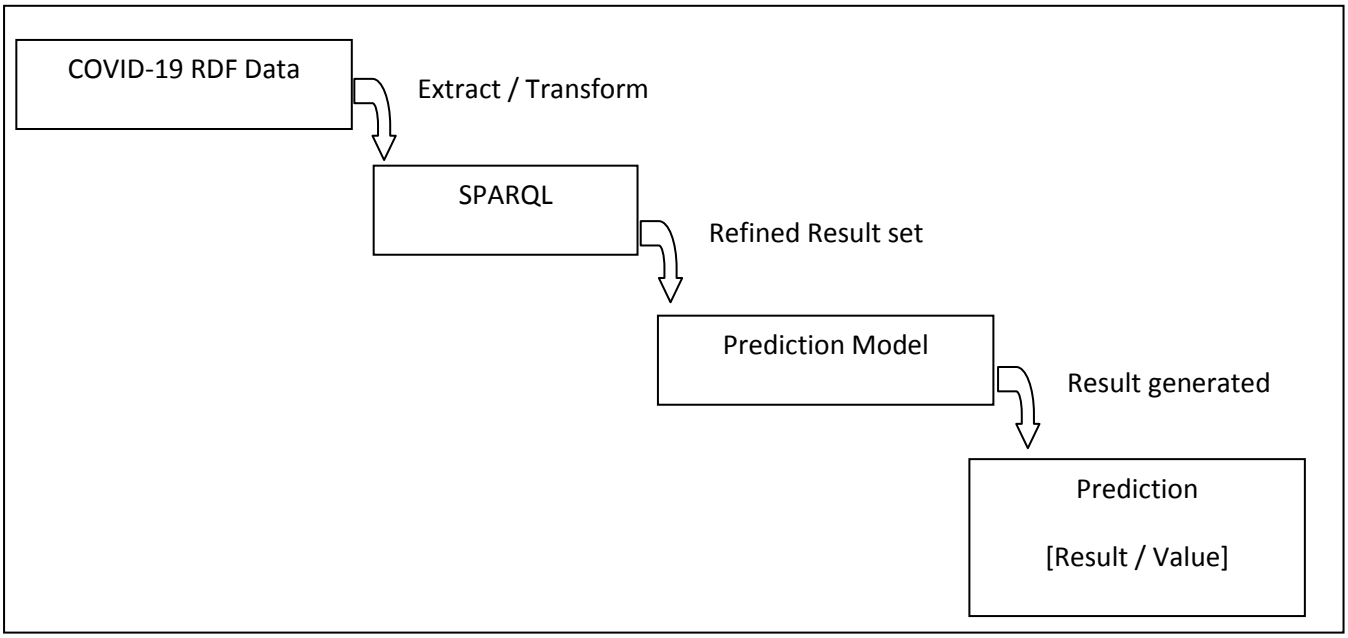

Figure 1: Flowchart of the workflow used in predicting values from COVID-19 RDF data

The processing workflow represented above using a flowchart has four basic steps and elaborated using Table1.

STEPS		Description	Key Points
Step-1	COVID RDF DATA	The Dataset can be built-in or can also be derived from other sources (Web Crawling)	• Domain knowledge of RDF triples to perform further steps ahead.
Step-2	SPARQL	In this Step the retrieval of RDF triples has been perform using SPARQL Query Language. The extraction and transformation of initial level can be made here to process the data into desirable format.	• Domain knowledge from step1 has been used to filter the data. • SPARQL Query has been used with RDFLIB python Library to retrieve data in to List (Python Object). • Further List data has been used to transform the data into data frame for further steps. • Define the column name to produce the refined dataset.
Step-3	Prediction Model	In this step the prediction models have been trained on the dataset with desired columns.	• Transformed the dataset in to desirable form to provide input for prediction models. • Define the training set and validation set for the prediction models.
Step-4	Prediction (Result)	In this step the prediction models has been used to predict the result on some desired input.	• The result can be predicted after the training of the models. • Further model can also be evaluated with their accuracy measures.

ISSN: 2233-7857 IJFGCN

Table 1: Description of Processing Flowchart shown in Figure 1.

In this paper the COVID-19 RDF dataset has been generated which further has been extracted and transformed using SPARQL. For the same rdflib library of python has been used along with SPARQL execution for the transformation and cleaning of the data, so that it can be further utilized by the predictive model to predict the future values. In this paper the predictive model like linear regression, Polynomial regression and time series forecasting has been used to predict the number of cases in the coming days. With the reference of the flow chart the predictive model has been applied using linear regression, Polynomial regression and time series forecasting method. These models has been used and compared with each other on different parameters and further compared by the actual outcome seen in future.

4.1 COVID-19 Semantic web Data Set

In this paper the dataset used has been referred from John Hopkins University and Medicine (Corona virus resource Centre) [2] [5]. The COVID data has been generated using RDF data in triples format which can be seen as a triple comprises of a subject, a predicate and an Object. The data set which has been analyzed in the paper is shown using a snapshot below in figure 2. In figure 2 snapshot of RDF data generated has been shown a randomly selected record of India and Indonesia for the date 7th Jun, 2020.

```
<http://covid.org/covid/25479>
<http://covid.org/a#ObservationDate> "6/7/2020" .
<http://covid.org/covid/25479>
<http://covid.org/a#Country_Region> "India" .
<http://covid.org/covid/25479>
<http://covid.org/a#Last_Update> "6/8/2020 3:33" .
<http://covid.org/covid/25479> <http://covid.org/a#Confirmed>
"257486" .
<http://covid.org/covid/25479> <http://covid.org/a#Deaths>
"7207" .
<http://covid.org/covid/25479> <http://covid.org/a#Recovered>
"123848" .
<http://covid.org/covid/25480> <http://www.w3.org/1999/02/22-
rdf-syntax-ns#type> <http://covid.org/a#sample> .
<http://covid.org/covid/25480> <http://covid.org/a#SNo>
"93524" .
<http://covid.org/covid/25480>
<http://covid.org/a#ObservationDate> "6/7/2020" .
<http://covid.org/covid/25480>
<http://covid.org/a#Country_Region> "Indonesia" .
<http://covid.org/covid/25480>
<http://covid.org/a#Last_Update> "6/8/2020 3:33" .
<http://covid.org/covid/25480> <http://covid.org/a#Confirmed>
"31186" .
<http://covid.org/covid/25480> <http://covid.org/a#Deaths>
"1851" .
<http://covid.org/covid/25480> <http://covid.org/a#Recovered>
"10498" .
```

Figure 2:A Snapshot of COVID-19 RDF Data generated [sample record of India and Indonesia 7th JUNE, 2020.]

ISSN: 2233-7857 IJFGCN

The complete data of the above sample data snapshot, representing the COVID-19 RDF data in N-triple format and has been further utilized for analysis and predicting the number of cases in near future dates using the predictive model.

4.2 SPARQL Query: SPARQL query has been used with python rdflib library on the COVID 19 RDF data and has been used for extraction and transformation. The following is the SPARQL query which has been applied on the dataset:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://covid.org/a#>
SELECT ?SNo ?ObservationDate ?Country_Region ?Last_Update ?Confirmed ?Deaths
?Recovered ?Province_State
WHERE
{
?x rdf:type ex:sample .
?x ex:SNo ?SNo .
?x ex:ObservationDate ?ObservationDate .
?x ex:Country_Region ?Country_Region .
?x ex:Last_Update ?Last_Update .
?x ex:Confirmed ?Confirmed .
?x ex:Deaths ?Deaths .
?x ex:Recovered ?Recovered  .
optional {?x ex:Province_State ?Province_State}
}
```

After the SPARQL query the extracted result has been transformed into a list (Python object) and further to the dataframe (Python object) in python. The first stage transformation has been made in SPARQL query to deal with the NULL values in the dataset. Further python utility of pandas has been used for transformation like domain, strings and the date formats on the COVID-19 RDF Data. Random sample of the transformed data set has been shown below with python snapshot below:

	SNo	ObservationDate	Country_Region	Last_Update	Confirmed	Deaths	Recovered	Province_State
0	17969	4/23/2020	Mainland China	4/24/2020 3:31	18.0	0.0	18.0	Qinghai
1	3954	3/7/2020	US	2020-03-07T04:43:02	1.0	0.0	0.0	Davis County, UT
2	6533	3/18/2020	Turkey	2020-03-18T02:33:06	98.0	1.0	0.0	None
3	13203	4/8/2020	US	4/8/2020 22:59	221.0	0.0	0.0	Wyoming
4	26002	5/18/2020	Czech Republic	5/19/2020 2:32	8586.0	297.0	5641.0	None

Figure 3: Snapshot of the data using Dataframe object in Python generated after processing COVID-19 RDF data using SPARQL

4.3 Prediction Models:

4.3.1 Linear Regression: Linear Regression is a technique of mapping various independent variables into one dependent variable which is continuous in nature. Dataset in case of linear regression has to be linear in nature as shown in figure 4. Result found using linear regression is more accurate and has very less error range.

ISSN: 2233-7857 IJFGCN

When number of independent variable is one, then it can be represented with the following equation (1):

$$Y=\alpha_0+\alpha_1.X_1 \quad (1)$$

Where α_i is the regression coefficient X_i is the independent variable and Y is the dependent variable whose value is derived from independent variable.

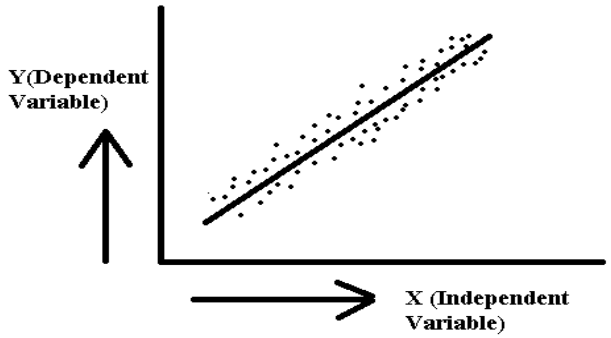

Figure 4: Representation of Linear regression using Independent and Dependent variable

Dots in the above figure represent values in the dataset. When independent variables are more than one in that case the multiple linear equation can be represented with the following equation (2) or equation (3).

$$Y=\alpha_0+\alpha_1X_1+\alpha_2X_2+\ldots\ldots+\alpha_nX_n \quad (2)$$

Alternatively,

$$Y=a_0+\sum_{n=1}^{m}(\alpha_nX_n) \quad (3)$$

Where α_i is the regression coefficient X_i is the independent variable and Y is the dependent variable

4.3.2 Polynomial Regression: If linear regression is applied to non-linear dataset then there will be huge loss of data. For example in the following graph given below in figure 5, the line goes through only 2 points and thus resulting in a large amount of data loss and inaccuracy of data.

ISSN: 2233-7857 IJFGCN

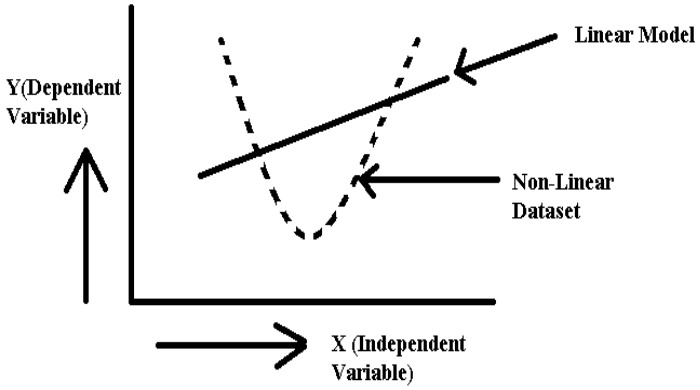

Figure 5: Representation of Linear and polynomial regression using Independent and Dependent variable

As a solution to this problem, we prefer to use polynomial regression instead of linear regression so that higher accuracy can be maintained as well as non-linear data can be handled.

0 degree polynomial equation can be represented as:

$$Y=C \text{ (Constant)} \tag{4}$$

1 degree polynomial equation can be represented as:

$$Y=mX+C \tag{5}$$

2 degree polynomial equation can be represented as:

$$Y=ax^2+bx+c \tag{6}$$

N degree polynomial equation can be represented as:

$$Y=\alpha_0+\alpha_1X+\alpha_2X^2+\ldots\ldots+\alpha_nX^n \tag{7}$$

Polynomial Regression using linear model can be represented using the following equation summarizing all equations (4),(5),(6),(7):

$$Y=a_o+\sum_{n=1}^{m}(a_nX_n)+F_p$$

Linear Model

Polynomial function used for converting Linear model to Polynomial

4.3.3 Time Series Forecasting: In this paper Holt Winters time series forecasting algorithm has been used. This algorithm is used to extract and analyze the historical data and use it for accurately predicting futuristic values. In this method the more weight is assigned to more recent

ISSN: 2233-7857 IJFGCN

historical data which is well suited for the COVID-19 data analysis. It has been analyzed and illustrated that this model best suits for predicting number of infected candidate from COVID based on the trend of growth of the patients. Other forecasting model can also be analyzed as a future scope with comparison of other time series forecasting methods like Auto Regressive Integrated Moving Average (ARIMA).

V. Analysis and Prediction of Results

COVID-19 RDF data has been analyzed using the functionality of SPARQL with rdflib library. The data set has been further utilized to view the number of cases arise, recovered and decease around the globe, also for country India as well. This has been represented using python data visualization tools/packages like Seaborn and Matplotlib.

NOTE: The COVID dataset has been taken up to the date: 17th July, 2020.

5.1 Analysis of COVID Effect using RDF dataset:

The analysis has been represented using a snapshot in figure 6, which is representing the progress of COVID spread around the whole world with the weekly statistics with [no of cases, no of recovered cases and total death with red, green and blue colors respectively].

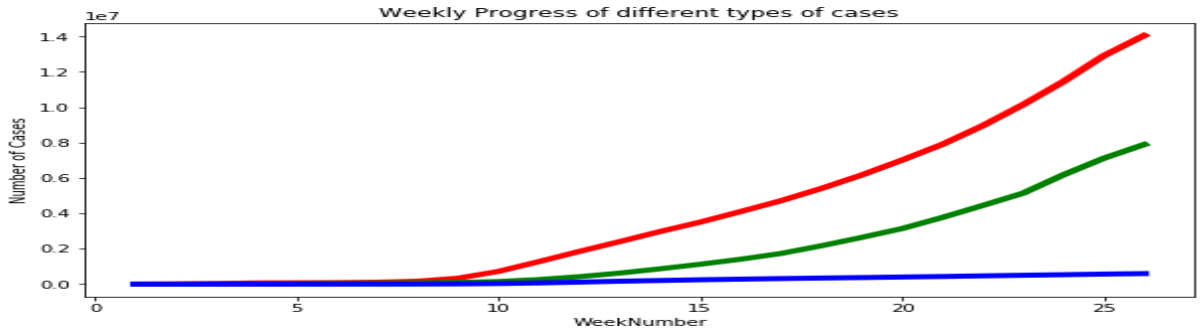

Figure 6 : Weekly progress of CORONA Spread around the World. [Growth of COVID-19 Weekly]

The similar report has been seen for India in figure 7 with same specifications.

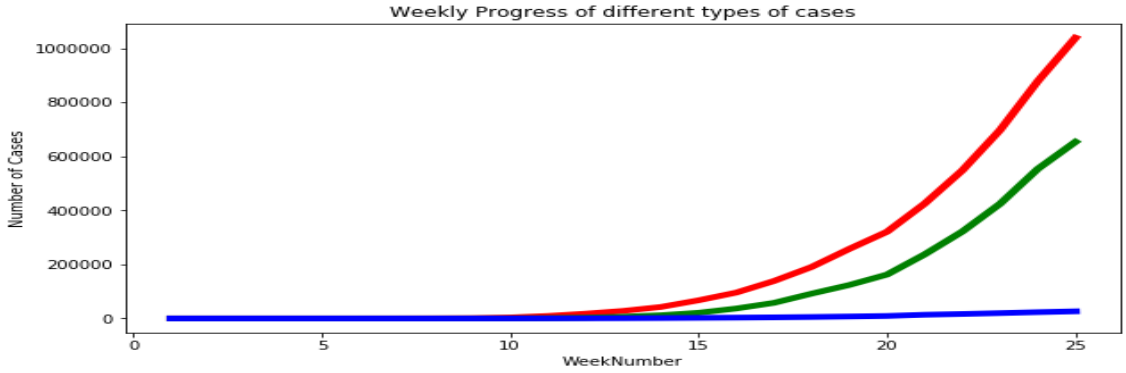

Figure 7 : Weekly progress of CORONA Spread around India. [Growth of COVID-19 Weekly]

As number of positives cases is increasing day by day as a current scenario, it has been represented using a snapshot in figure 8.

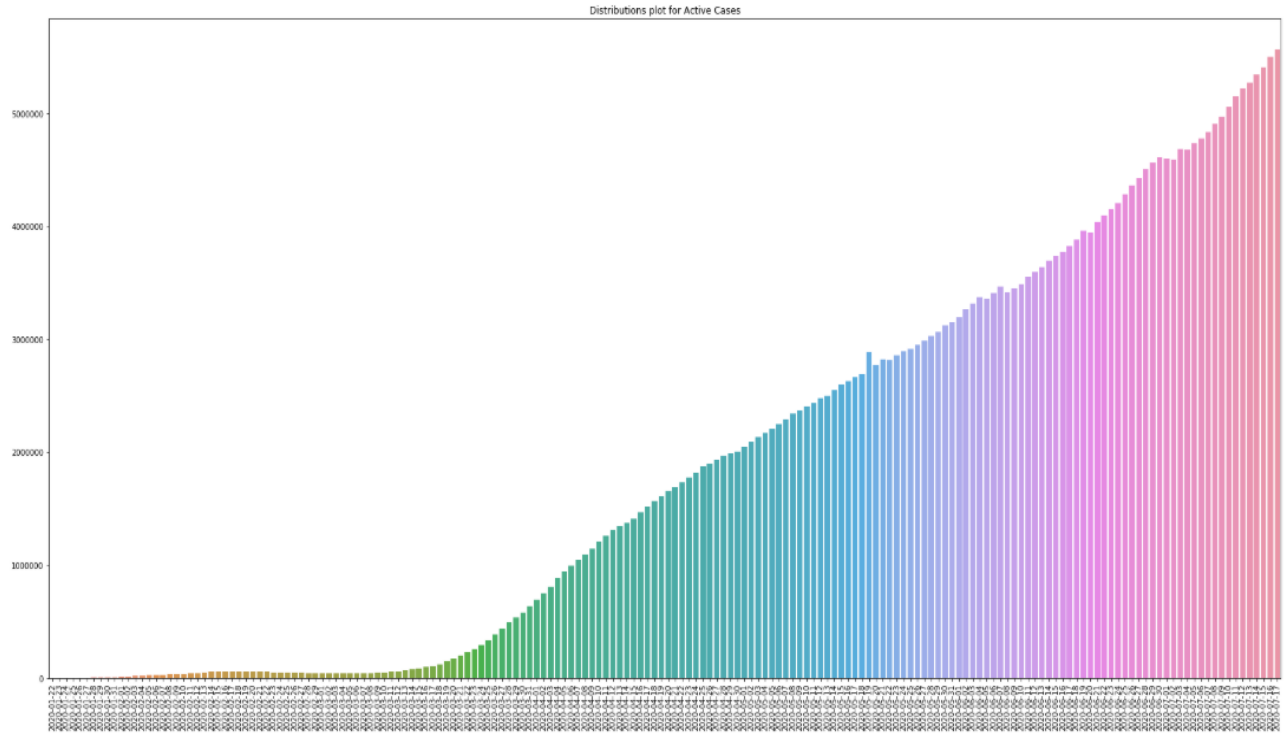

Figure 8: Distribution plot for No. of Active cases around the World.

From the above analysis it can be concluded that the CORONA impact will increase exponentially in near future. Figure 9 represent the weekly increase in active cases and death cases

ISSN: 2233-7857 IJFGCN

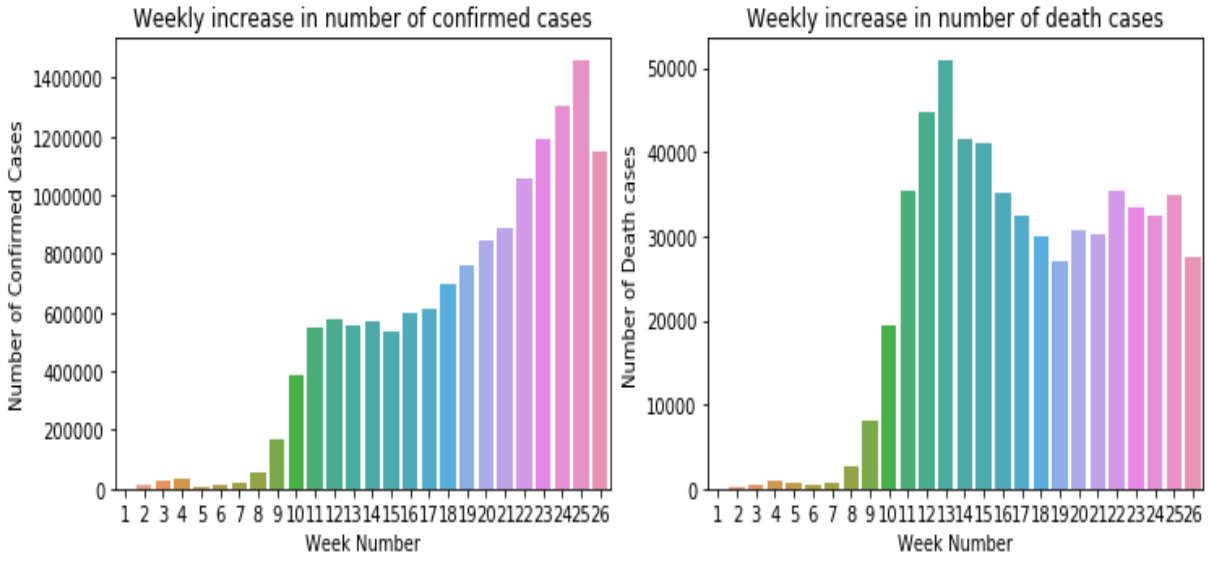

Figure 9: Weekly increase in No of confirmed and death cases around the globe.

5.2 Prediction Models representation on COVID-19 RDF dataset:

A linear and polynomial regression models has been trained with the training sets in figure10 and figure11 respectively. And it has been shown that polynomial regression model best fit the model representation.

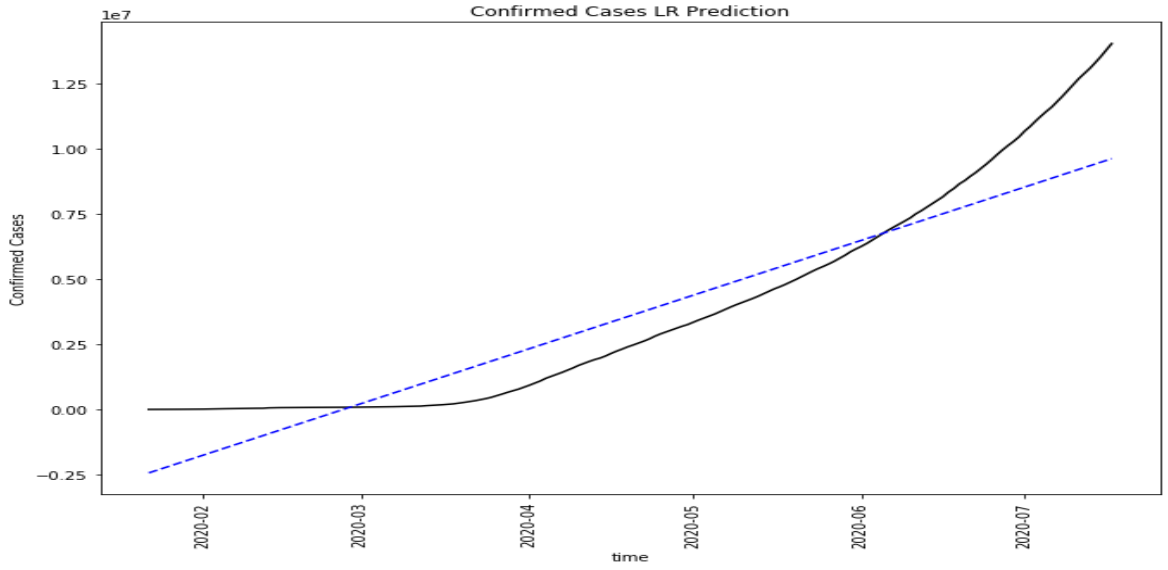

Figure 10: Predictive Model Mapping with the existing dataset using Linear Regression.

ISSN: 2233-7857 IJFGCN

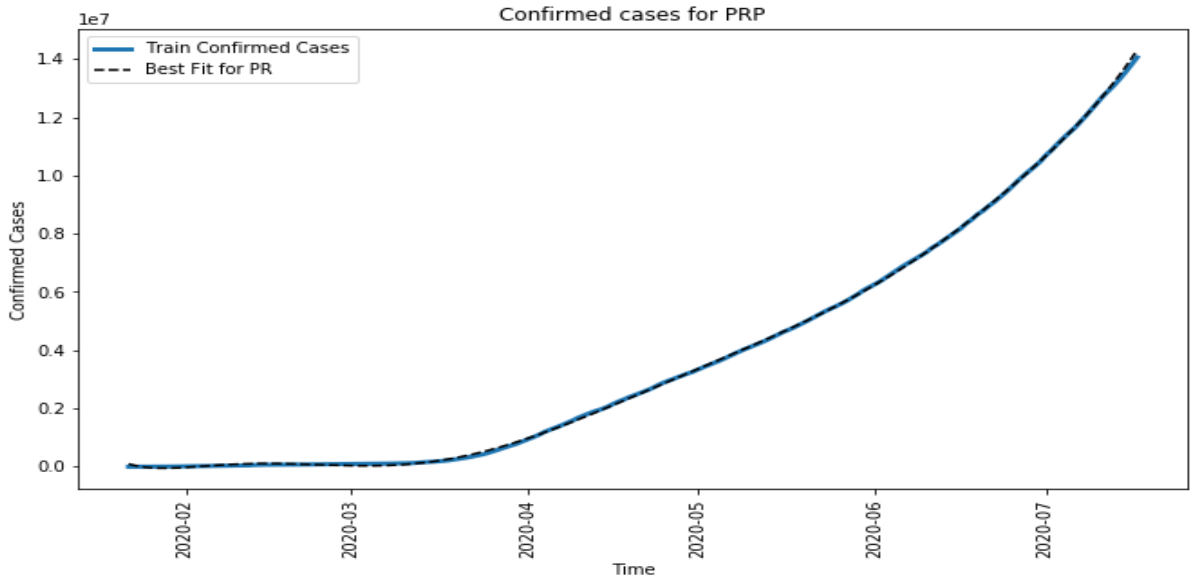

Figure 11: Predictive Model Mapping with the existing dataset using Polynomial Regression.

The next analysis has been made using the time series model Holts Linear model for prediction. The Figure 12 present the predictive model on training set and the validation set, which has been generated for future dates. The validation set has been shown using a yellow color line.

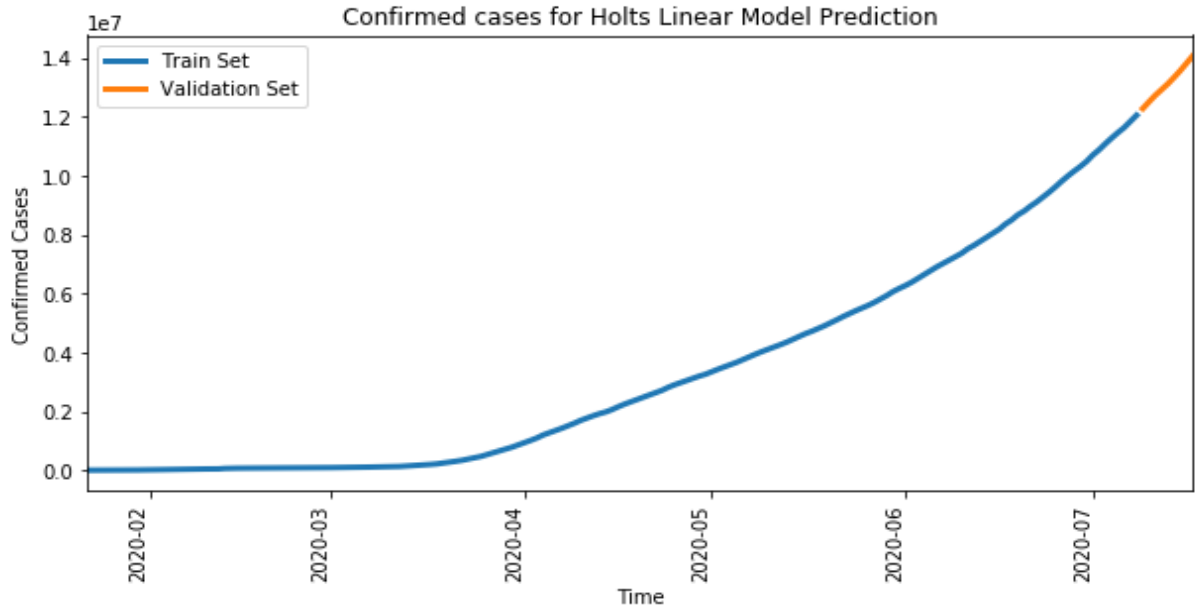

Figure 12: Time Series forecasting with Holts predictive Model around the world

ISSN: 2233-7857 IJFGCN

5.2.1 Results and Comparison of Prediction Models:

In this section the result produced by the different prediction models has been represented and been compared with the actual output in Table2.

DF OFFSET	Dates	LRP	PRP	Holts Model	Actual Confirmed Cases
0	2020-07-18	9702201	14611199	13984903	14055299
1	2020-07-19	9770404	14934078	14180852	14292922
2	2020-07-20	9838608	15274619	14376801	14507491
3	2020-07-21	9906811	15634620	14572750	14707960
4	2020-07-22	9975015	16016032	14768699	14947428
5	2020-07-23	10043218	16420974	14964648	15229740
6	2020-07-24	10111422	16851742	15160597	15511157
7	2020-07-25	10179626	17310819	15356546	15792390
8	2020-07-26	10247829	17800883	15552495	16046986
9	2020-07-27	10316033	18324825	15748443	16252541
10	2020-07-28	10384236	18885754	15944392	16487669

ISSN: 2233-7857 IJFGCN

11	2020-07-29	10452440	19487013	16140341	16691527
12	2020-07-30	10520643	20132193	16336290	17029155
13	2020-07-31	10588847	20825141	16532239	17309805
14	2020-08-01	10657050	21569978	16728188	17850479
15	2020-08-02	10725254	22371112	16924137	18079516
16	2020-08-03	10793458	23233253	17120086	18282208

Table 2: Predictive Model predicting No. of cases on future dates around the world.

The Figure 13 present the predictive model on training set and the validation set, which has been generated for future dates for India. The validation set has been shown using a yellow color line.

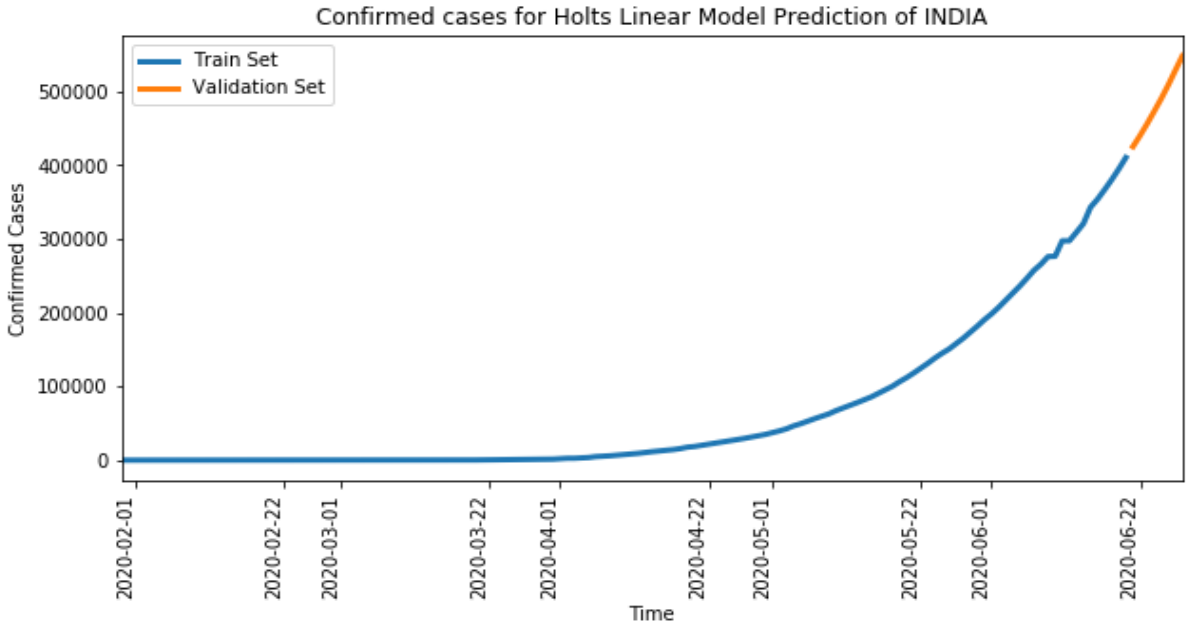

Figure 13: Time Series forecasting with Holts predictive Model for India

ISSN: 2233-7857 IJFGCN

All the three Predictive models have been compared and shown below in Table3 with the predictive values for India. Further the mean square errors of the models have been calculated and also shown below.

DF OFFSET	Dates	LRP (INDIA)	PRP (INDIA)	Holts Model (INDIA)	Actual Confirmed Case (INDIA)
0	2020-07-18	427431	1092209	1008624	1039084
1	2020-07-19	430805	1131598	1032947	1077781
2	2020-07-20	434178	1172558	1057270	1118206
3	2020-07-21	437552	1215165	1081593	1155338
4	2020-07-22	440926	1259498	1105916	1193078
5	2020-07-23	444299	1305639	1130239	1238798
6	2020-07-24	447673	1353675	1154563	1288108
7	2020-07-25	451047	1403696	1178886	1337024
8	2020-07-26	454420	1455798	1203209	1385635
9	2020-07-27	457794	1510078	1227532	1435616
10	2020-07-28	461168	1566641	1251855	1480073

ISSN: 2233-7857 IJFGCN

11	2020-07-29	464541	1625595	1276178	1483156
12	2020-07-30	467915	1687053	1300501	1581963
13	2020-07-31	471289	1751132	1324825	1634746
14	2020-08-01	474*662	1817955	1349148	1750723
15	2020-08-02	478036	1887652	1373471	1803695
16	2020-08-03	481410	1960356	1397794	1855745

Table 3: Predictive Model predicting No. of cases on future dates around the India.

The root mean squared error for all the models has been analyzed separately on World data and on India's data. And the result has been shown in figure 14, which is showing that Polynomial regression performs better in both the cases. And Holt Winters model works well on world data as its RMSE value is less. It has also been analyzed that on different instances of the datasets, Holts model also works well even better than Polynomial model but only on few instances. On an average it can be concluded that polynomial regression works well on an average with different instances.

ISSN: 2233-7857 IJFGCN

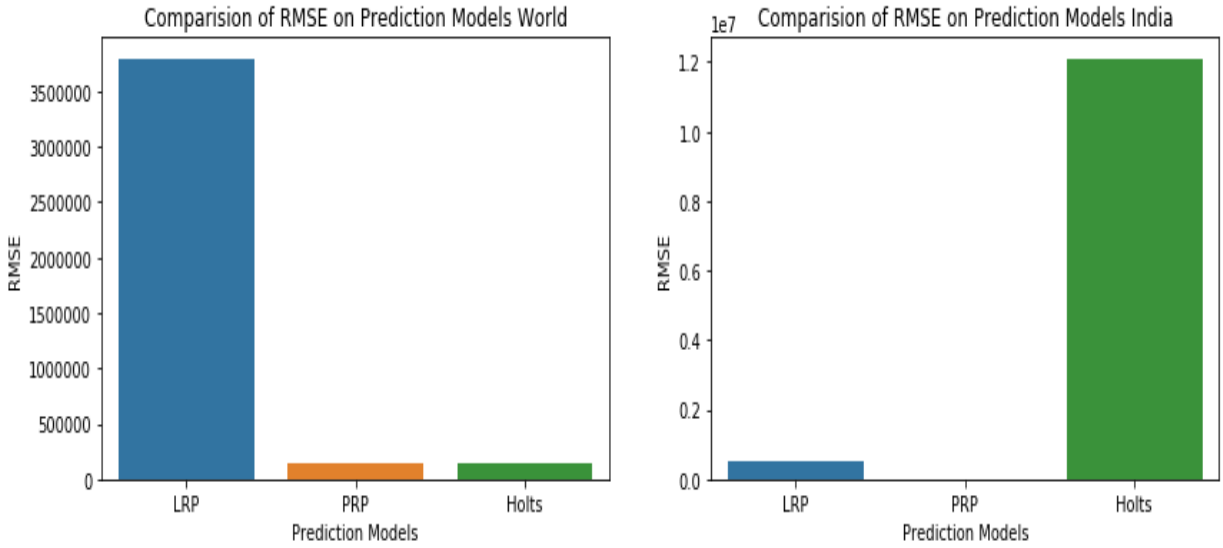

Figure 14: Comparison of RMSE on the data set representing World data followed by comparison of RMSE on India dataset

The accuracy has also been checked by getting the model trained on September, 2020 COVID RDF data. The r2 score model has been used to find the accuracy of the models on the testing dataset. The result of the accuracy measure has been shown in figure15, which also states that polynomial model is having more accuracy than the other models which have been illustrated in the paper.

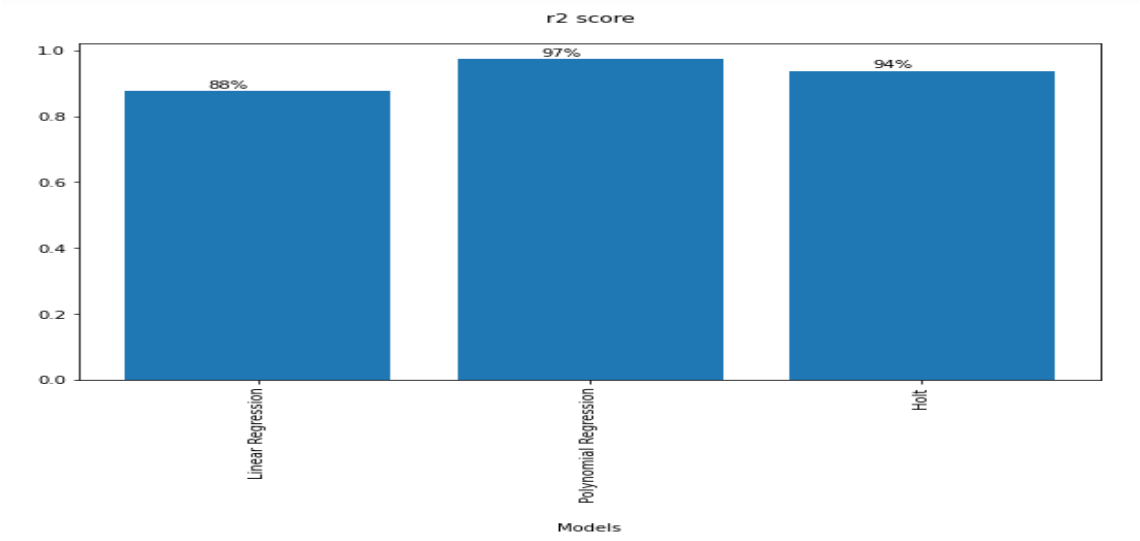

Figure 15: Comparison of accuracy on world COVID RDF data for Sept,2020 using r2 score.

ISSN: 2233-7857 IJFGCN

VI. Significance of work

The analysis work undergone in the paper is helpful for the domain of semantic web by adding COVID dataset as a knowledge graph over web. The significance of the work is that it has been based on RDF data used in semantic web for interlinking data over web. This paper deals with the real world data on a global issue which is still ongoing and many researchers can use the generated Covid RDF dataset as a resource available to analyze different models further. The prediction models proposed in this paper on such modern issues has significance specially when it works on RDF dataset which is more flexible and extensible dataset used in Semantic web.

VII. Conclusion and Future Scope

COVID is a pandemic for the whole world and many countries are still facing a disaster like situation at this time. As all information is needed globally to analyze what is going around the world. This paper analyzes that if the data is there residing in the RDF format of semantic web, it can be analyzed and interpreted easily. Further how to build a prediction model from the data. Three models of predictions has been analyzed and evaluated with root mean squared error values. As per the analysis and prediction model used for COVID RDF data it is concluded that its growth around the world in exponential in nature. Further it has been analyzed that polynomial regression model is the most effective model used for predicting future values on an average from all the models discussed in the paper. Also it has been analyzed with Holt Winters model that it produces better result with a condition that the growth of COVID over the time must be same. Both polynomial and Holt Winter model produces the result which are very close predictions to the actual upcoming results for such type of instances. Further as a future scope many other predictions model like ARIMA (Auto Regressive Integrated Moving Average) and others for time series forecasting can be applied and evaluated to analyze for more effective prediction results. Also the existing model can also be analyzed how it will work after a specific time period by adding new results in COVID-19 RDF datasets.

REFERENCES

[1] World Health Organization (WHO), “Novel Coronavirus – China,” Jan. 2020. [Online]. Available: http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/

[2] JHU, “Coronavirus information and resources for JHU (John Hopkins University)-Baltimore, Maryland”, Mar.2020 [Online]. Available: https://coronavirus.jhu.edu/map.html/

[3] T.B. Lee, J. Hendler, O. Lassila, “The Semantic Web”, Scientific American 194.5, 2001 PP: 19-37.

[4] RDF Working Group, “RDF,” W3C Working Group, February 2014.[online] Available: https://www.w3.org/RDF/

ISSN: 2233-7857 IJFGCN

[5] "COVID-19 data repository by the center for systems science and engineering (CSSE) at Johns Hopkins University." [Online]. Available: https://github.com/CSSEGISandData/COVID-19

[6] D. Gaglione, P. Brace et.al (2020) Adaptive Bayesian Learning and Forecasting of Epidemic Evolution – Data Analysis of the COVID-19 Outbreak, IEEE Acess.

[7] N.N. Khondoker, "Forecasting COVID-19 Pandamic: A data driven analysis",Chaos,Solitons and Fractals- Nonlinear Science and Nonequilibirium and Complex Phenomena Journal, Elsevier (2020) https://doi.org/10.1016/j.chaos.2020.110051.

[8] A Behnood, E.M. Golafshani, S. M. Hosseini, "Determinants of the infection rate of the COVID-19 in the U.S using ANFIS and Virus optimization algorithm (VOA)", Chaos,Solitons and Fractals- Nonlinear Science and Nonequilibirium and Complex Phenomena Journal, Elsevier (2020).

[9] Y. Peng , M.H. Nagata , "An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data", Chaos,Solitons and Fractals - Nonlinear Science and Nonequilibirium and Complex Phenomena Journal, Elsevier (2020).

[10] M. Gabriele and M. Gianluca, "Modelling of the Sars-Cov-2 in Italy and universal forecasting of the epidemic in the world" , Chaos,Solitons and Fractals - Nonlinear Science, and Nonequilibrium and Complex Phenomena, Journal Elsevier (2020).

[11] B. Sayantani and C.H. Roy, "Going by the numbers : Learning and modeling COVID-19 disease dynamics", Chaos ,Solitons and Fractals - Nonlinear Science, and Nonequilibrium and Complex Phenomena, Journal Elsevier (2020).

[12] F. Michel et al. (2020) Covid-on-the-Web: Knowledge Graph and Services to Advance COVID-19 Research. In: Pan J.Z. et al. (eds) The Semantic Web – ISWC 2020. ISWC 2020. Lecture Notes in Computer Science, vol 12507. Springer, Cham. https://doi.org/10.1007/978-3-030-62466-8_19.

Authors Profile

Mr. Rupal Gupta is a Research Scholar at USIC&T, Guru Gobind Singh Indraprastha University, Delhi, India. His area of interests are Semantic Web, SPARQL Query Processing, Big Data and Data Mining. He received his Master's, MCA from UPTU, Lucknow and M.Tech.(IT) from USIT, GGSIPU, New Delhi . He has published papers in IEEE conferences. He is currently working as an Assistant Professor at Teerthanker Mahaveer University, Moradabad and having 14 years of teaching experience.

Dr. Sanjay Kumar Malik completed his Ph.D. in the area of Semantic Web from USIC&T, GGSIP University, Delhi. He is currently working as Professor in University School of Information, Communication and Technology, GGSIP University. He has more than 20 years of industry and academic experience in India and abroad (Dubai and USA). His areas of research interest are Semantic Web and Web Technologies. He has several research papers in reputed international conferences (India/abroad) and Journals. He has been session chair in several international IEEE/Springer conferences and honored with third best researcher award, 2011 by GGSIP University for his contributions in research.

ISSN: 2233-7857 IJFGCN