Additive Penalized based Quantile Regression (APQR) for Predictive Analytics for Pandemic Data

Gopi Suresh Arepalli[1], Rajasekhar Kommaraju[2], Shaik Sikindar[3], Dr Y Sowmya Reddy [4] , Dr Y Madhavi Reddy[5]

[1] Assistant Professor, Department of Computer Science and Engineering, Lakireddy Bali Reddy College of Engineering,My avaram, Andhra Pradesh, India. arepalli.0888@gmail.com

2 Sr. Assistant Professor, Department of Information Technology, Lakireddy Bali Reddy College of Engineering,Mylavaram, Andhra Pradesh, India. rajasekharkommaraju@gmail.com

3 Assistant Professor, Department of Computer Science and Engineering, Vignan's Foundations for Science, Technology & Research (Deemed to be UNIVERSITY), Guntur, Andhra Pradesh
shaik5651@gmail.com

4 Associate Professor, Department of Computer Science and Engineering, CVR College of Engineering, Ibrahimpatnam ,Telangana, India. yalla.soumya.reddy1980@gmail.com

5 Assistant Professor, Department of Mathematics, CVR College of Engineering, Ibrahimpatnam, Telangana, India. madhavireddy75@gmail.com

Article Info
Page Number: 784 – 796
Publication Issue:
Vol. 71 No. 3 (2022)

Abstract

In present, entire globe facing global challenges due to the pandemic situations raised by the known COVID-2019. Also, this can be impacted to most issues which can includes financial crises, medical emergency, education loss, chronic hunder, migration of people, etc. It can not only threat to the pre-existing people suffered with health issues also to the healthy people who are more conscious towards health. However, there is a necessity to apply a very good statistical analysis over such kind of issues. Moreover, many of the research works towards this problem entire globe. One of the popular techniques to apply statistical analysis to such kind of pandemic data is quantile regression. Need an extension version to the quantile regression to provide solution to the pandemic data analysis. Moreover, it is required to know prior the overall idea of the conditional distribution of a response variable. A penalized based quantile regression utilizes the minimization of the L1 norm to address heterogeneous pandemic data prediction. The study focused on minimization of L1 norm to effectively address such pandemic with additive penalized model for quantile regression known as APQR. The proposed model can help to find regression coefficients and control heterogeneity effectively. A numerical study using

Article History
Article Received: *12 January 2022*
Revised: *25 February 2022*
Accepted: *20 April 2022*
Publication: *09 June 2022*

simulated and real examples demonstrates the competitive performance of the proposed APQR would be helpful and recommended for pandemic data prediction than standard quantile methods.

Keywords: Covid-19, Machine Learning, Predictive Analysis, Regression Analysis, Quantile Regression, and Regularization.

---

1. Introduction

The entire globe across 211 countries is impacted with the issue of Corona virus disease 2019 (COVID-19) from January 2020, basically it is and an infection desease caused by the novel corona virus SARS-CoV-2. One of the dangerous task related suck kinds of pandemic COVID-19 transmission involving everyone at high risk of exposure [2-4]. It is reported that in January 2020, more than 7 million COVID-19 cases and 431541 deaths occurred around the world[1]. The disease first was found in the city of Wuhan, China in middle of the year 2019 accordingly. Generally, any people are suffering with the COVD-2019 had symptoms which includes fever cough, breath issue, pneumonia and kidney issue, health failure etc[6-7]. In this paper, focused on mainly India county and necessary data related to address issue of pandemic analysis collected data from the media and social media sites. The complete information related to the pandemic data over India is shown in Figure 1.

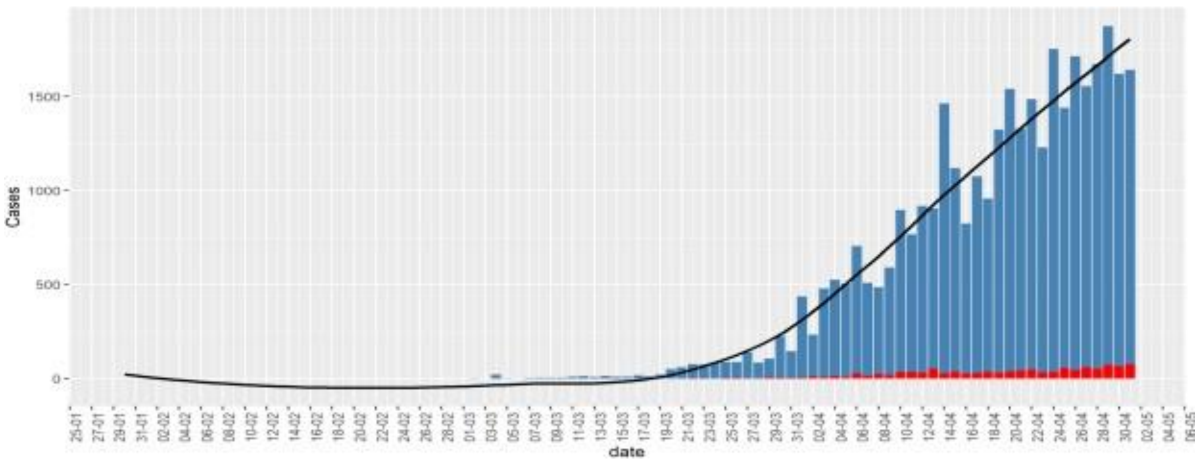

Figure 1: Pandemic data in India day wise

From the figure 1, it is shown the from 20 March 2020 onwards in India, observed that there is an exponential growth in the daily number of COVID-19 cases. Also it is observed that there few more initiatives also taken by the Indian government to control and prevention of the COVID-2019 situations [8-9]. The complete measures were listed in the Table 1.

Table 1: Total preventive measure initiated by the government of India.

Date	Measure
January 25th, 2020, to March 13th, 2020	Initiation of health screenings at both airports and border crossings
February 26th, 2020-March 20th ,2020	quarantine policies applied to the people coming over different countries
February 26th, 2020-March 13th ,2020	Applying Visa restrictions to covid effected counties
March 5th ,2020	Controlling public gathering
March 11th ,2020	Initiation of Border Closure and extensive checks
March 16th ,2020	Controlling of public gathering
March 18th ,2020	initiation of the restriction towards the Travel to infected countries
March 20th ,2020	Testing for COVID-19
March22nd ,2020	Suspend of Flights from effected countries
March22nd ,2020	Train Services Suspension
March24th ,2020	Suspension of Domestic Airplane Operations
March25th ,2020	Lockdown entire country about 21 days
March25th ,2020	Train Services Suspension
March30th ,2020	Improvement in facilities of quarantine/isolation
April 14th ,2020	Lockdown is extended up to 3 May 2020
May 1st , 2020	Lockdown is extended up to 17 May 2020

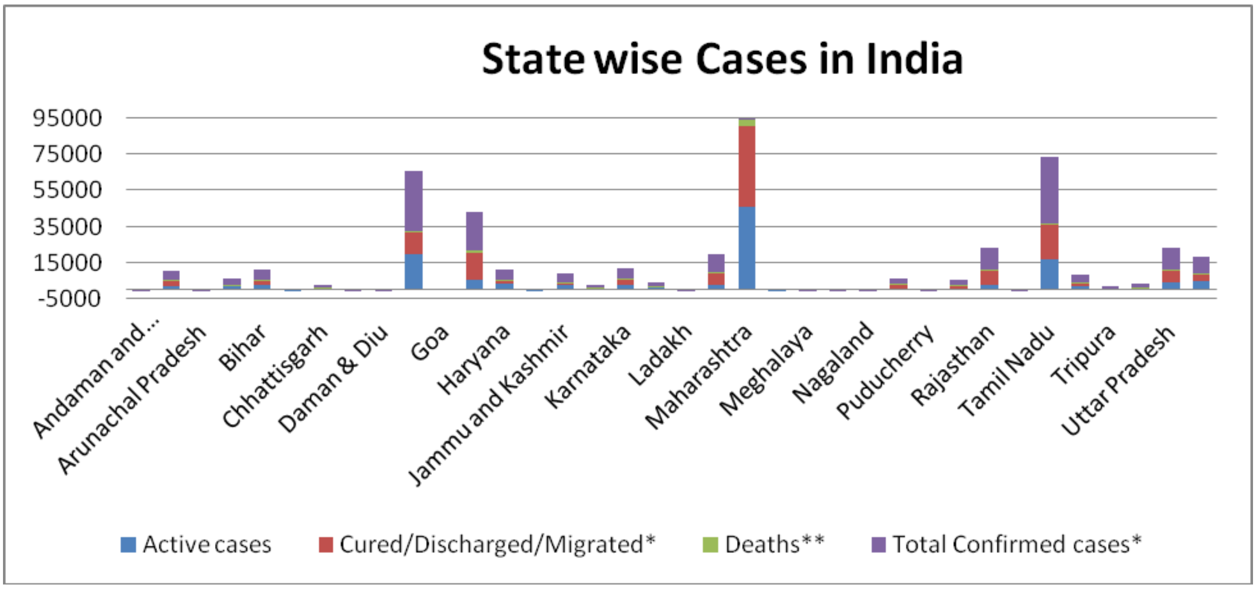

Figure 2. The total number of confirmed cases, Deaths, cured/discharged/migrated and actives on 3rd, June, 2020 in different states of India.

From the Figure 2, it was reported that most of the cases of COVID-19 recorded at state of Maharashtra, Tamil Nadu, Delhi, Kerela and Telangana. Also, it was seen that the highest deaths were seen in the On Maharashtra, Kerela, Haryana and Uttar Pradesh states accordingly. The complete information related to each state in terms of the total number of confirmed cases, Deaths, cured/discharged/migrated and actives is shown in Figure 2.

The quantile regression is different from the standard regression where it can used to do analysis among groups over the entire distribution of continuous response or specific quantile. It is also understanding the relationship among predictor and response variable outside of the conditional mean. To understand and analyze these kinds of pandemic data, quantile regression is one of the suitable method and it analyses better way among relations. However, to improve the performance of the pandemic analysis regularized version of quantile is the best solution and it produces better results. But the study in addition to regularization also added additive nature to cover the non-liner analysis of pandemic high dimensional covid-19 data. The study in which standard quantile makes utilized penalized technique which can utilize the minimization of the L1 norm to address heterogeneous kind of data analysis. Also, moreover, work also concentrated on the minimization of the L2 norm and it extends the analysis with understanding of complete conditional distribution of the response variable. In addition to penalization technique also included the additive in nature to the quantile regression and it also produce higher results in estimation of response variable. Also, the proposed method APQR is good for outliers in pandemic high dimensional data.

2. Related work:

The most useful method for the statistical analysis used for the prediction in linear regression. However, it is not good for providing proper relation among the response variable in the cumulative distribution. The quantile regression (QR), introduced in the year 1978, and is an extension to the standard linear regression model and it is good for analyzing the conditional distribution of a response variable at proper manner [17]. The method mainly focused on the regressor in distribution at every level of part and good for outliers. Moreover, it has been popular to wide range of application and produced good computation power. But this method first obeys the linear model drawbacks in addition also limit to few issues which can includes crossing of quantile, misspecification, and multicollinearity. To resolve such kind of mentioned issues penalization regression is adapted to the quantile regression and is helps to avoid overfitting, outlier and other issues raised by the linear regression. Number of penalized regression introduced by most of the researchers L1 penality, L2 penality , the elastic net penalty , and etc [11-15]. Introduce additive nature to the quantile regression with the concept of the bootstrapping and is developed and is helps to control smooth error density. [23], proposed quantile based predictive method for the COVID-2019. A to predict the fertility rate among cases raised in the county. Also, it is notice that data in this study is represented in the three-parameter beta (GB3) distribution. A study of adaptive lasso based regression method is applied to real data retrieved form the chile is developed [25]. The study which can identify the infection fatality rate of COVID-2019 based on the index model used to define risks. The advantage of this method is model free variable selection method. [26]

applied bootstrap method to quantile regression to predict Length-of-Stay hospitalized for patient with COVID-19 respectively

3. Proposed Framework

The study uses initially the quantile regression approach to design the predictive model for the COVID-2019 [16-20] in terms of the different types of cases and the idea is described in this section.

Consider the linear model to represent the COVID-2019 data:

$$Y_i = X_i^T \beta + e_i, i = 1, \quad , n \ (1)$$

From the equation, (Y^*) is represented as i^{th} observation, X_i is denoted as the i^{th} independent variable, and the error term is represented as e_i.

In addition to the linear regression , various other analysis also popular includes exponential, two degree, three degree and fourth degree polynomial for the statistical analysis. The description of these models is given below:

$$Y_i = aX_i^T + bX_i + e_i, i = 1, \ldots., n \ (2)$$

$$Y_i = aX_i^T + bX_i^T + cX_i + e_i \ (3)$$

$$Y = aX_i^T + bX_i^T + cX_i^T + dX_i + e_i (4)$$

where a, b, c , and d are called the parameters of regression analysis.

Let consider, a quantile level of interest $\tau\,(0,1)$ the conditional τ^{th} quantile of e_i given X_i is zero. The conditional quantile regression is as follows:

$$Q_\gamma(\tau / X) = X^T \beta(\tau) \ (5)$$

From (6) , $Q_\gamma(\tau / X)$ represented as the τ^{th} conditional quantile of the concern response Y for the related X. From the unknown functional, there estimated one parameter $\beta(\tau)$ .Applying the minimizing objective function the correspond point estimate $\beta(\tau)$ is determined respectively.

The study [29] suggested that to enhance statistical analysis more further better to adapt bootstrap method to the quantile regression (QR) analysis [21-23]. The bootstrap method, is one the popular method helps to do statistical analysis with few samples not depends on the error terms [27]. The standard deviation of the slope coefficient is determined using the bootstrap re-sampling method [28]. In that method, the standard error is defined using following procedure and is shown in Algorthm 1.

Algorithm1: Bootstrap method for resampling

Step1:Let consider a data (Y^*), from each sample (R) is estimated as $Y^*_{(1)}, Y^*_{(2)}, \ldots\ldots, Y^*_{(L)}$ by resampling method

Step2 : Next, calculate bootstrapcoefficient ($\mathbf{b}^*$)fromsamples

$\mathbf{b}^*_{(1)}, \mathbf{b}^*_{(2)}, \ldots\ldots., \mathbf{b}^*_{(L)}$

Step 3 :Finalerror is estimated from distribution of bootstrapslope

$$S_E(\mathbf{b}^*) = \left(\sum \frac{\sum_{k=1}((\mathbf{b}^*_k - \mathbf{b}^*(m))^2}{L-1} \right)^{1/2}, \qquad (6)$$

where $\mathbf{b}^*(m) = \frac{\sum_{k=1}^{L} \mathbf{b}^*_k}{L}$ is the mean of all bootstrpae slope coeffiencents

$\mathbf{b}^*$ = bootstrap OLSestimateof coefficient

L = No.of iterations / repetetions

$\mathbf{b}^*_k$ = tthe slope coefficient from the bootstrapped sample $Y^*_{(k)}$

Another robust method is called The Quantile Regression (QR) used for statistical analysis. In QR method, relationship among independent (X) and dependent (Y) variables are estimated based on the conditional mean function, $E(Y \mid X)$. The OLS minimizes the squared errors (S_{SE}) with the estimation of the conditional quantile function, $Q_{f\,(Y|X)}$ [22]. Moreover, such method can determine the slope coefficient will minimize error using the objective function and is defined as:

$$Q_f(\mathbf{b}) = \sum_{i=1}^{N} |Y_i - \mathbf{b}X|$$

(7)

Where , β is the slope coefficient, corresponding quantile values is denoted as q and number of observations is denoted as N and is described in Eq(7). is the number of observations. The method also minimizes the sum of absolute errors S_{AE}. Hence, with all above observations QR is also robust to the outlier [23].

By adding above two methods proposed method is determined combining both QR and bootstrap. The proposed method is derived advantages of both methods which can fit for large samples and robust to the error. Some of the related work related to this Bootstrap based QR is known as BQR is enhancement to QR and good for outlier.

The Lasso penalized linear regressions method does not have the selection consistency. To overcome this limitation, Zou (2006) proposed the adaptive Lasso penalized linear regressions method which enjoys the selection consistency. The adaptive Lasso penalized methods can shrink the estimates of the parameters of unimportant covariates to 0 as the sample size $n \to \infty$ thus, removing all the unimportant covariates automatically and yielding sparse estimation results.

$$Q_f(\mathbf{b}^*_L) = \mathbf{b}^*_L = \arg\min \left\{ \frac{1}{2n} \left\| Y_{*,L} - X^T \mathbf{b}_K \right\|_2^2 + \mu_L \sum_{j=1}^{p} w_{Lj} \left| \mathbf{b}_{K,j} \right| \right\} \qquad (8)$$

The tuning parameters μ_L with $L = 1,....,d$ are chosen by the cross-validation method. By solving the minimization problem eq () , we obtain a sparse estimate of β .

4. Experimental Results and Discussion’

4.1. Data set:

Now, the study, consider the COVID-19 data collected from the different web sources and compared performance using APQR model. The method is compared with linear, standard quantile regression respectively. To analyze and provide statistical and predictive analysis, total five data sets were collected from the different states of India. The states includes Telangana, Kerala, Delhi , Andhra Pradesh, and Tamil Nādu. For all these five data sets applied statistical and predictive analysis in the further sections detailly. Each data set consider the common variables and complete details about variables is shown in Table 2.

Table 2. Data and its Variables

Notation	Variable Names (As of 14th April 2020)
N_C	Number of positive cases
N_D	Number of deaths
H_{Gdp}	Health expenditure by government (% of GDP)
H_{Gexp}	Current expenditure by Government Health expenditure
P_{65p}	Above age 65 population
B_{SL}	Least services of sanitization facility

From the Table 3, the individual values of all the 5 data sets shown detail. It is shown that few of the variables are measured in terms of the metrics and remaining are per cent. According to the study of data from the Table, it is retaining that Kerala is the highest effected place compared to other states. However, Telangana is identified as lower risk state for the COVID-2019.

Table 3. Sample States consider into the study

Variable	Telangana	Kerala	Delhi	Andhra Pradesh	Tamilnadu
N_C	113	637581.6	14645.5	7322.7	10902.7
N_D	0	91940.5	912.9	456.5	679.6
H_{Gdp}	0.179	14.36	3.62	1.81	2.69
H_{Gexp}	4.586	202.54	50.97	25.49	37.95
P_{65p}	0.931	37.89	9.54	4.77	7.10
B_{SL}	9.455	305.64	76.92	38.46	57.26

Study also done statistical analysis over the five data sets defined by the individual data sets. It is shown that analysis in the Table 4 & 5 ,provided the basic statistical analysis which includes max, min, average values of each individual elements in the data set.

Table 4. Descriptive Statistics of Variables

Variable	**Minimum**	**Maximum**	**Mean**	**Standard deviation**
N_C	113	637581.59	134113.06	281498.54
N_D	0.00	91940.45	18797.88	40889.33
H_{Gdp}	0.18	14.36	4.53	5.64
H_{Gexp}	4.59	202.54	64.31	79.14
P_{65p}	0.93	37.89	12.04	14.79
B_{SL}	9.46	305.64	97.55	118.96

Table 5. Statistical analysis: Slope Equation Test

Variables	**Dependent variable:** N_C ***Vs*** ***Independent variables (***H_{Gdp} , P_{65p}, B_{SL})					
	$\tau_{(0.25,0.50,0.75,0.90)}$		$\tau_{(0.25,0.50,0.75)}$		$\tau_{(0.75,0.90)}$	
	F-statistic	***p-value***	***F-statistic***	***p-value***	***F-statistic***	***p-value***
H_{Gdp}	0.84	0.46	0.01	0.92	0.13	0.87
P_{65p}	0.95	0.41	0.03	0.86	1.23	0.29
B_{SL}	0.53	0.64	0.17	0.67	0.33	0.71

Observe from the linear regression, only one of the coefficients H_{Gdp} is only significant. However, by applying the quantile regression, except the $H_{Gdp\ all}$ the remaining are significant. Hence, the result is shown that lower quantile of cases occurred in the states which are providing excellent expenditure towards public health. For example, consider the Kerala district which can put higher expenditure and it helps to reduce the mortality rate or health risk.

4.2. Performance measures

The measure of difference between the original and the predicted values over n number of samples and is defined as:

$$MSE = \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n-(k-1)} \tag{9}$$

The MAPE, is one of the popular statistic benchmark tool to find the best accuracy in terms of the error rate and is defined as relative difference among the true and predictive values and is defined as:

$$MSE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{O_i - P_i}{O_i}\right| \quad (10)$$

How model is liked to explain the total variance in the data is described by the two statistical tools known as R^2 and adjusted R^{*2} and their ranges always between 0 to 1, i.e., $0 \le R^2 \le 1$ and are defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - Mean_i)^2} \text{ and } R^{*2} = R^2 - \frac{(1-R^2)k}{n-(k+1)} \quad (11)$$

4.3. Predictive Analytics:

The study also, focused on the predictive analytics over covid-2019 data about the different states of the India. Major special focus is done on the state Tamil Nadu.

Table 6. Comparison of proposed with linear and quantile regression model

States	Linear regression (LR)			Qunatile Regression (QR)			APQR		
Telangana	0.7098	0.7879	0.5871	0.8621	0.8602	0.4256	0.9995	0.9995	0.2422
Kerala	0.6007	0.5969	0.787	0.9615	0.9611	0.6724	0.9995	0.9995	0.9099
Delhi	0.751	0.7474	0.6213	0.9191	0.9179	0.5327	0.9989	0.9988	0.6465
Tamilnadu	0.973	0.9727	0.3035	0.6939	0.6903	0.2763	0.9959	0.9957	0.5506
Andhra Pradesh	0.7422	0.7403	0.4205	0.5945	0.5915	0.3852	0.9747	0.9737	0.2781

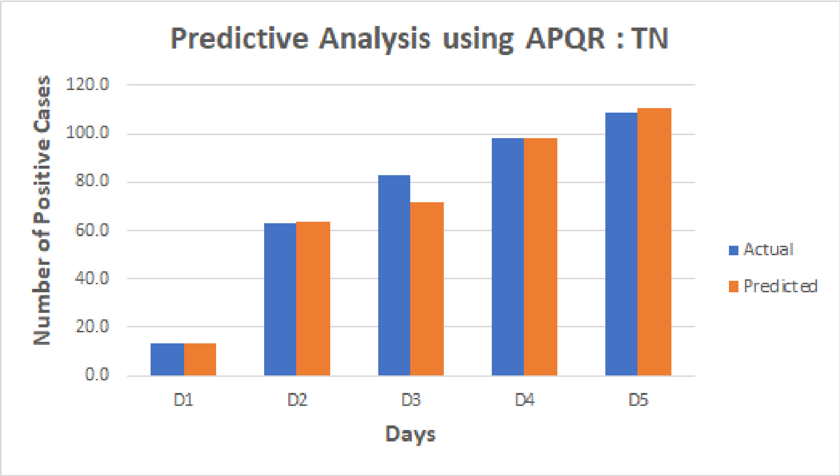

Fig 3. Comparison of positive cases actual results with proposed AQAR for training the COVID-19 data in TN State

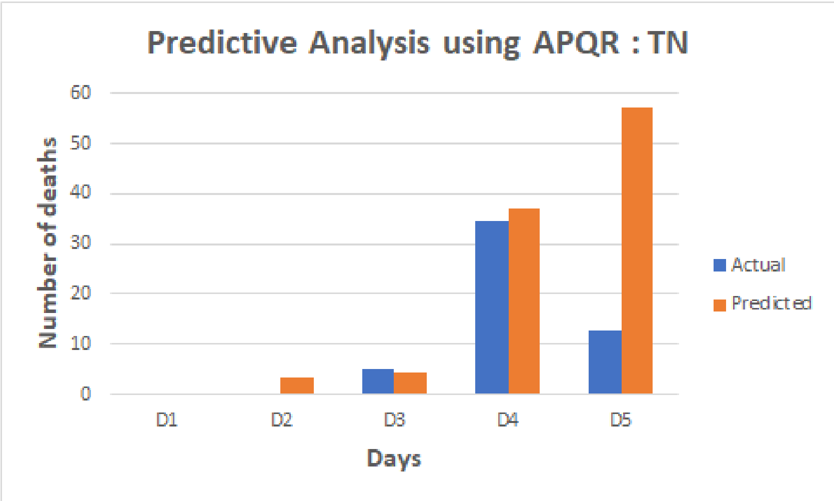

Fig 4. Comparison of number of deaths actual results with proposed AQAR for training the COVID-19 data in TN State

From the figure 3 to 5, and Table 6 It is shown that proposed method is performing better in terms of predictive analysis compared to the linear and quantile regression . Also, it is retained that predictive values computed from the different days is more closed to the actual values.

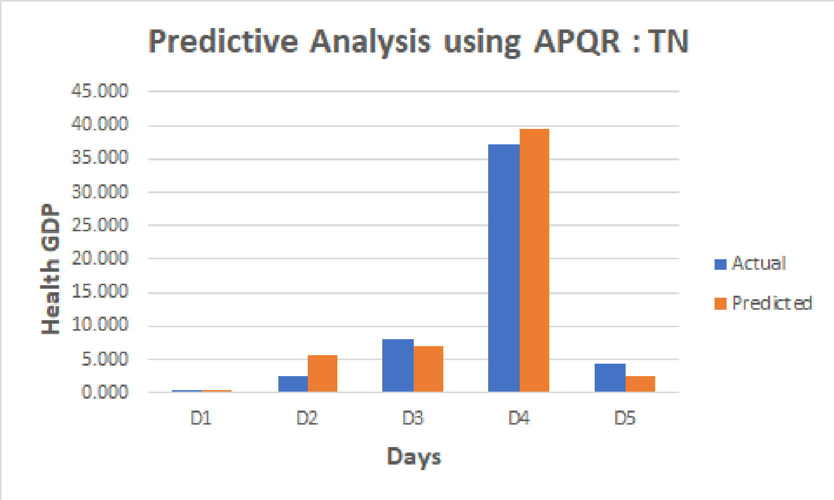

Fig5. Comparison of health expenditure by the government actual results with proposed AQAR for training the COVID-19 data in TN State

The overall analysis is done in terms of the three variable which includes Number of cases, Number of deaths and Health expenditure done by government to control and manage Covid-2019. From the all the three variable the proposed retain the close value. Therefore, the proposed method is very useful for future prediction of the COVID-2019 outbreak to the next coming days from the current date.

5 Conclusion

One of the popular techniques to apply statistical analysis to such kind of pandemic data is quantile regression. Need an extension version to the quantile regression to provide solution to the pandemic data analysis. Moreover, it is required to know prior the overall idea of the conditional distribution of a response variable. A penalized based quantile regression utilizes the minimization of the L1 norm to address heterogeneous pandemic data prediction. The study focused on minimization of L1 norm to effectively address such pandemic with additive penalized model for quantile regression known as APQR. The proposed model can help to find regression coefficients and control heterogeneity effectively. A numerical study using simulated and real examples demonstrates the competitive performance of the proposed APQR would be helpful and recommended for pandemic data prediction than standard quantile methods.

References

1. Wu Z, McGoogan JM (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention.
2. World Health Organization (2020) Coronavirus disease 2019(COVID19): situation report, p 67.
3. Pirouz B, Shaffiee Haghshenas S, Shaffiee Haghshenas S, Piro P(2020) Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of

COVID-19(new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. Sustainability12(6):2427

4. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z (2020)Review of the clinical characteristics of coronavirus disease 2019(COVID-19). J Gen Int Med 35:1545–1549. https://doi.org/10.1007/s11606-020-05762-w

5. Pirouz B, Shaffiee Haghshenas S, Shaffiee Haghshenas S, Piro P(2020) Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19(new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. Sustainability12(6):2427.

6. Benvenuto D, Marta G, Lazzaro V, Silvia A, Massimo C (2020) Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in brief 105340. MedRxiv. https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1.

7. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in IEEE Access, vol. 5, pp.8869-8879, 2017.

8. Yadav RS. Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India. In: Int. j. inf. tecnol. Springer; 2020. https://doi.org/10.1007/s41870-020-00484-y.

9. https://www.kaggle.com/imdevskp/corona-virus-report. [Accessed 4 June 2020].

10. Kalipe G, Gautham V, Behera RK (2018) Predicting malarial outbreak using machine learning and deep learning approach: are view and analysis. In: 2018 International conference on information technology (ICIT). IEEE, pp 33–38.

11. K. Lavanya, L. S. S. Reddy and B. Eswara Reddy,” Modelling of Missing Data Imputation using Additive LASSO Regression Model in Microsoft Azure”, Journal of Engineering and Applied Sciences,2018,Vol 13,Special Issue 8,pp:6324-6334.

12. Little RJ, Rubin DB. Statistical analysis with missing data. 2nd ed. New York: Wiley; 2002.

13. K. Lavanya, L. S. S. Reddy, and B. Eswara Reddy,” A Study of High-Dimensional Data

14. Imputation Using Additive LASSO Regression Model”, Computational Intelligence in Data Mining,2019 Advances in Intelligent Systems and Computing 711. (SCOPUS)

15. Lavanya K., Reddy, L., & Reddy, B. E. (2019). Distributed Based Serial Regression Multiple Imputation for High Dimensional Multivariate Data in Multicore Environment of Cloud. International Journal of Ambient Computing and Intelligence (IJACI), 10(2), 63-79.doi:10.4018/IJACI.2019040105. (SCOPUS) (Web of Science) (ESCI)

16. Rubin DB. Multiple imputations for nonresponse in surveys. New York: Wiley; 1987.

17. Koenker, R. and Bassett, G. Jr., 1978, “Regression Quantiles,” Econometrica, 46(1),pp. 33–50.

18. Santri, D., 2016, “Statistical Downscaling Modeling with Quantile Regression Using Lasso to Estimate Extreme Rainfall,” Proceedings of the 7th SEAMS UGM International Conference on Mathematics and Its Applications, http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.4940862.

19. Soleh, A.M., Wigena, A.H., Djuraidah, A., and Saefudin, A., 2015, "Statistical Downscaling to Predict Monthly Rainfall Using Linear Regression with L1 Regularization (LASSO)," Applied Mathematical Sciences, 9(108), pp. 5361–5369.
20. K. Lavanya, K. Harika, D. Monica, K. Sreshta," Additive Tuning Lasso (AT-Lasso): A Proposed Smoothing Regularization technique for Shopping Sale Price Prediction", International Journal of Advanced Science Technology, Vol 29,No 05,pp 878-886,2020. (SCOPUS)
21. Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society Series B, 58(1), pp. 267–288.
22. Zou, H. and Hastie, T., 2005, "Regularization and Variable Selection via the ElasticNet," Journal of the Royal Statistical Society Series B, 67(2), pp. 301–320.
23. K. Lavanya, G.V.Suresh," An Additive Sparse Logistic Regularization Method for Cancer Classification in Microarray Data", The International Arab Journal of Information Technology, Vol. 18, No. 2, March 2021.https://doi.org/10.34028/iajit/18/10, ISSN: 1683-3198E-ISSN: 2309-4524, Impact Factor is 0.654
24. Farid, S., Naeem, M. A., Paltrinieri, A., & Nepal, R. (2022). Impact of COVID-19 on the quantile connectedness between energy, metals and agriculture commodities. Energy economics, 109, 105962. https://doi.org/10.1016/j.eneco.2022.105962.
25. Li, Zitong & Sillanpää, Mikko. (2012). Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik. 125. 419-35. 10.1007/s00122-012-1892-9.
26. Kazemi M, Nazari S, Motamed N, Arsang-Jang S, Fallah R. Prediction of Hospitalization Length. Quantile Regression Predicts Hospitalization Length and its Related Factors better than Available Methods. Ann Ig. 2021 Mar-Apr;33(2):177-188. doi: 10.7416/ai.2021.2423. PMID: 33570089.
27. Plonsky L, Egbert J, Laflair GT. Bootstrapping in applied linguistics: Assessing its potential using shared data. Applied Linguistics. 2015; 36(5): 591–610.
28. Larisa Nikitina, Rohayati Paidi, Fumitaka Furuoka., Using bootstrapped quantile regression analysis for small sample research in applied linguistics: Some methodological considerations, PLOS ONE, January 2019, 10.1371/journal.pone.0210668.
29. Nikitina L, Furuoka F. Expanding the methodological arsenal of applied linguistics with a robust statistical procedure. *Applied Linguistics*. 2018; 39(3): 422–428. 10.1093/applin/amx026