Authorship Confirmation

Please save a copy of this file, complete and upload as the "Confirmation of Authorship" file.

As corresponding author I, Pawan Singh , hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.

2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.

3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.

4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.

5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature: Pawan Singh Date: 04/21/2021

List any pre-prints: NA

Relevant Conference publication(s) (submitted, accepted, or published): NA

Justification for re-publication: NA

Research Highlights (Required)

- A novel time-series forecasting architecture that helps deal with sudden hike in the data.

- A pattern recognition architecture that helps to forecast unemployment rate in India in the post-pandemic timeframe.

- The current study proposes a novel architecture to deal with the rare unusual trends that are not seasonal in nature.

- Vanilla time-series forecasting frameworks remained ineligible to capture the rare sudden patterns.

- Study handles the issue of unavailability of adequate and proper data.

Prognosticating the effect on Unemployment rate in the post-pandemic India via Time-Series Forecasting and Least Squares Approximation

Ashutosh Agrahari[a], Pawan Singh[a,**], Ankur Veer[a], Anshuman Singh[a], Ankit Vidyarthi[b], Baseem Khan[c]

[a]Dept. of Computer Science and Engineering, Amity University, Lucknow, India.
[b]Department of CSE & IT, Jaypee Institute of Information Technology, Noida, India.
[c]Dept. of Electrical and Computer Engineering, Hawassa University, Hawassa P.O. Box 05, Ethiopia.

ABSTRACT

The current paper aims to analytically visualise the future outcomes that the post-pandemic India might have in store for its citizens. We use time series forecasting on various collected data and combined the statistics of economics-deciding parameters to forecast the trends that might be prevalent in the next year. Since, the data contains a single anomalous trend, even the Prophet model could not learn this property from the data since this trend is not seasonal in nature. The current study proposes a novel architecture to deal with these rare unusual trends by combining two models - one learning normal usual patterns and the other getting trained on usual as well as rare anomalous patterns. It could help in dealing with sudden hike patterns like due to COVID-19 in the data, and lead to better forecasting on future timeframes. We combined the results of two distinct time-forecasting models trained on two sets of data of varying timeline lengths, using parameters obtained from Least Squares Approximation (LSA). The LSA helps us find an approximate vector approximation so as to obtain a model performing closely to the actuals.

1. Introduction

COVID-19 pandemic has hit the world very harshly. India has now become a hotspot of this pandemic after the USA and Brazil, after surpassing Russia in the number of cases, and is soon to enter stage 3 considering fatality rate. The countries are helpless before this natural phenomenon and are forced to implement lockdowns to prevent its spread. As is evident from many other historical pandemics and epidemics in various parts of the world, there has always been a great deal of adverse effects on the economic conditions of the country or the world as a whole. Consequently, due to lockdowns, trading and working in the industries is stalled which in turn leads to loss of jobs of numerous employees. The employers are then not in the position to support themselves and thereby many companies and industries get shut down. Since industries run a country's economy, and with the industries shutting down or getting

in loss, the economy of the entire country gets disrupted. A disease outbreak, being at the scale of the world, affects the entire world. Since most countries are connected through trade and commerce, the pandemic can lead to utter loss of jobs and economy of nations.

The COVID-19 disease has been growing steadily at an exponential rate throughout the world. Since its inception, it has disrupted various sectors of human life including business, health, education and governance. As of July 20, 2020 the number of cases is 14,360,451, out of which 603,285 have lost their life while 8,071,937 have recovered (1). The economies have been adversely affected and their GDP (Gross Domestic Product) is following a negative slope in today's scenarios. In India, due to sequence of lockdowns, menial laborers have lost their daily-wage jobs, and are left with nothing to sustain in this pandemic-affected economy, and hence many have died not due to the disease but of scarcity of money to support one's family. This is a very arduous situation in the country, and its citizens must not leave hope. Government has been doing its best to control this

---

**Corresponding author: pawansingh51279@gmail.com

pandemic. It has taken many initiatives and cut down the cost of COVID-19 tests by half, but still the situation in the country is not under control and is surging high.

The unemployment rate in India had been swinging closely to around 2.7% in previous years but has been exaggerated to an astonishing 15% average during this pandemic, specially between March and May. The rules of social distancing and lockdown has added more to disruption rate of dwindling unemployment rate and consequently on the overall economy of the country. Apart from studying the effects of the pandemic on unemployment rate, inflation rate and economy growth rate, the paper also analyses the consequences of up-skilling at home, on academia and corporate world. We run time-series forecasting on the collected data, and present our findings through comprehensive analysis of the results.

The paper is organized as follows. Section 2 presents a discussion about the works that have been conducted and published in the past. Section 3 discusses the algorithms and time-series forecasting architectures used in the paper. Section 5 outlines the methodology that has been adopted to implement analyses presented in the paper, and also describes the workflow that is used in the implementation of the proposed architecture. Section 6 talks about the issue of data unavailability, its collection and curation. Section 7 presents the results of the algorithms employed to perform time-series forecasting on the curated dataset. Section 8 presents a comparative analysis of the proposed architecture with other available time-series forecasting frameworks and existing researches. Section 9 summarizes and concludes the paper. Finally, section 10 comments on the future scope of the current research.

2. Related Work

With the onset of the pandemic, a lot of researches have been conducted by researchers worldwide on the various aspects of the mankind's life from health and hygiene to social and environmental contexts. Sengupta et al. (2) estimated the date when the number of cases would reach the peak and when it will flatten out using methods shared from data analysis and machine learning paradigm. Gupta et al. (3) used time-series forecasting methods like ARIMA to predict the future trends in the rise of the COVID positive cases in India. Poddar and Yadav (4) concluded the relationship between the pandemic and the downfall of Indian Economy based on their null hypothesis. Paul (5) showed the effect of pandemic on Indian economy graphically which included representation of unemployment rate based on CMIE data and various other related parameters. Katris (6) analyzed different machine learning (neural networks, support vector regression) and time series (ARIMA, FARIMA) models to predict unemployment rates of several countries. Karlsson and Javed (7) used both univariate and multivariate time series models like SARIMA, SETAR and VAR, and various macroeconomic variablesto predict unemployment rate and presented the results with 95% forecasting confidence of SARIMA model. The paper also suggested that short term forecasting models give better result than long term. Vikas et.al. in (9) used decision tree approach on the basis of the level of pessimism to
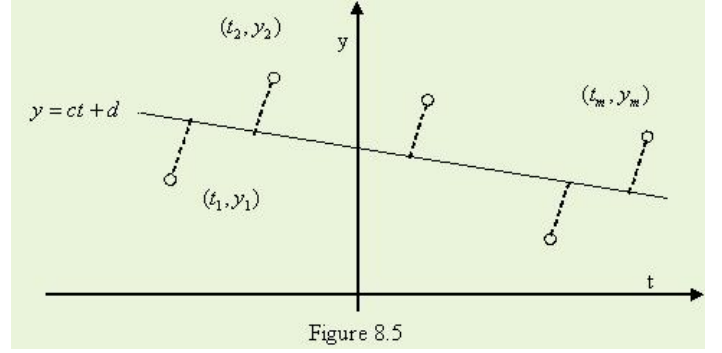

Fig. 1: LSA tries to minimize the deviations from the actual points while fitting the predicted line. Image Credits: www.math4all.in

predict the future unemployment rate in India. Khem et.al. in (10) showed how various economic parameters effect unemployment rate in India, and they showed that 48% of the effect is due to the GDP of the nation. They used correlation and regression analysis to come to their conclusion, and verified them against Okun's law and historical data. Dumičić in (11) experimented with various forecasting models to predict unemployment rate separately for both the genders. It is not pertaining to India, but it is relevant to our study since it used Linear Trend Equation in MINITAB software along with various levels of smoothening to arrive at results closer to the actual.

Most of the researches related to forecasting for COVID have been mainly to predict the number of cases, or predict the unemployment rate using vanilla models and algorithms. Our research focuses on the social sphere of the problems that arose as a consequence of the pandemic. We try to incorporate the sudden spike that happened as a consequence of the pandemic, which made making predictions over Indian economy difficult. In this paper, we handle the shortcomings of the unavailability of reliable data and also the sudden spike that disturbed the underlying mathematics of the models.

3. Background

3.1. Least Squares Approximation

LSA is a method to approximate the solution to a set of linear equations for whom a real solution does not exist. It works by minimizing the sum of squares of residuals obtained by each single equation. It is most commonly used in the domain of data fitting, wherein the residuals are the difference between observed and actual values, and the LSA tries to minimize the sum of these residuals.

Since, a line, representing solution, intersecting with the plane of set of equations, cannot be found, LSA tries to find a line that has minimum distance from the plane, thereby finding an approximate solution. It tries to find an approximate solution to the equation $Ax = b$, but since no solution exists for this so LSA finds a projection of the line on the plane $Ax - b = 0$ and minimizes the distance between line and plane. The final solution that LSA comes to is represented by Equation 1. In this equation, A^T is the transpose of the matrix A, $\hat{x}$ is the approximate solution, and b are actual values.

$$A^T A \widehat{x} = A^T b \quad (1)$$

3.2. Levenberg-Marquardt Algorithm

LMA is an optimization algorithm that combines the best of both the steepest gradient descent method and the Gauss-Newton method. It aims to find the minimum of linear or non-linear function having an array of parameters. The gradient descent method is useful when the initial point is far from minimum point of the function. The Gauss Newton algorithm is useful when initial point is close to the minimum point of the function. The combination of the two helps the algorithm become independent of initial point and hence Levenberg-Marquardt algorithm is derived. It can be expressed by Equation 2.

$$
\begin{aligned}
& x_{n+1} = x_n - (H + \lambda I)^{-1} \triangledown f(x_n) \\
& when, \\
& \lambda \to 0, \; Gauss - Newton's \; Method \\
& \lambda \to \infty, \; Gradient \; Descent \; Method
\end{aligned}
\quad (2)
$$

Here, $f(x)$ is a multi variable non-linear function, H is the Hessian matrix, I is the identity matrix, λ is a hyperparameter, and n is the number of iterations.

3.3. ARIMA

ARIMA or the Auto-Regressive Integrated Moving Average is a purely statistical model. It is mainly employed in the time-series models by statisticians and economists. There are two types of models under ARIMA, namely, seasonal and non-seasonal. The non-seasonal model takes into consideration three parameters - p, d and q. Here, p refers to the number of lags to be used to predict the series or it is the order of Auto Regressive term. q refers to the number of lagged errors or it is the order of the Moving Average term. Since, the very first operation involved in ARIMA model is making time series stationary, so, to perform that the most common way is to difference the previous value from the present value. So, here comes the third parameter d which refers to the minimum number of differencing required to make the time series stationary. The non-seasonal model doesn't account for any seasonal patterns, so to add that functionality, Seasonal ARIMA or the SARIMA (15) comes into picture. The mathematical form of the model involves two models - Auto Regressive model and Moving Average model. A pure Auto Regressive model can be mathematically expressed by Equation 3, where, Y_{t-1} is the lag 1 of time series, β_1 refers to the coefficient of lag 1, ϵ_1 is the error term, and, α refers to the intercept term.

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_1 \quad (3)$$

Similarly, Moving Average model can also be depicted mathematically by Equation 4, where error terms ϵ_i refer to errors from auto regressive model, ϕ_i are the coefficients of error terms and α is the intercept term.

$$Y_t = \alpha + \sum_{i=1}^{q} \phi_i \epsilon_{t-i} + \epsilon_t \quad (4)$$

Now, the ARIMA model consists of at least one differencing parameter to make time series stationary which combines Auto Regressive model and Moving Average model. The final mathematical form of the ARIMA model can be depicted by Equation 5.

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{q} \phi_i \epsilon_{t-i} \quad (5)$$

3.4. Prophet

Prophet (17) is an open-source time-series forecasting model developed by Facebook. The parameters in the model can be easily adjusted for tuning while integrating the core business knowledge. Missing data, outliers can be handled easily and automatically. The three main components of Prophet are *trend*, *seasonality* and *holidays*. The mathematical model underneath Prophet is a decomposable additive model (Eqn. 6) to accommodate the in-dependency of importance and forecasting effects of the components of the model.

$$y_m = \beta_0 + \sum_{n=1}^{t} f_n(x_{mn}) + \epsilon_m \quad (6)$$

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (7)$$

The adapted version of the decomposable additive model can be expressed by the Equation 7. Here, $g(t)$ represents trend, $s(t)$ seasonality, $h(t)$ holidays and ϵ_t the error term associated with modelling. The main advantage is that unlike all other models which are natively made for univariate time-series, Prophet very easily accommodates multivariate dimension of time-series data paradigm. Prophet proved to be the most reliable algorithms out of all the three techniques we employed in our experiments. The Prophet architecture is made to train on the univariate data.

4. Engineering Applications of Time-series Forecasting

Time series forecasting has already proven its importance in various fields like weather forecasting, process and quantity control, stock market predictions etc. The new emerging applications of time series analysis can be found in energy consumption sectors. Chirag Deb et al (18) review, showed the time series forecasting used for buildings and campuses energy consumptions. The new building's energy consumption where the past data does not exist, computer simulation is used in such scenarios to predict future energy consumptions. While for existing buildings, where historical data is available time series forecasting is used to predict energy consumption and carbon emissions. Along with single time series data analysis, the energy data was also analyzed with other time series parameters like weather outside and inside conditions of environment. A hybrid model combining two or more forecasting techniques used, proved to be best for such time series forecasting.

The other important application of time series forecasting which has come out recently is forecasting of covid-19 hospital census. The planning of sufficient availability of covid-19

beds , personal protection equipment kits, intensive care units etc is important based on local infection incidence of covid-19. Hieu M Nguyen et al (19) used vector error correction model (VECM) to forecast COVID-19 Hospital Census: A Multivariate Time-Series Model Based on Local Infection Incidence. The VECM framework was used on local infection incidence and hospital census from May 15 to December 5, 2020 of North Carolina, United States. The 7 days ahead forecast was performance was measured by mean absolute percentage error (MAPE), with time series consolidation. The results achieved where then compared with regular ARIMA model and were proved far better than ARIMA.

The applications of time series forecasting is also seem to be in container logistics industry. Sonali Shankar et al (20) used long short term memory (LSTM) to predict container throughput. This multi-billion dollar industry produces a huge revenue if an accurate forecasting is done even at any minute strategical level in the industry. The systematic use of port data is directly proportional to economic development of the inland region lying behind port. The LSTM model made, was then also compared with the other time series forecasting methods like autoregressive integrated moving average (ARIMA), simple exponential smoothing, Holt–Winter's, error-trend-seasonality, trigonometric regressors (TBATS), neural network (NN) and ARIMA + NN. The evaluation was done on the dataset of Singapore port. The results indicated that LSTM outperformed these traditional methods.

5. Heuristic-based Model Training and Prediction

The prediction of future effects has been made using Facebook's time-series forecasting framework Prophet. The accumulated data was first cleaned and pre-processed as per the requirements. Due to the pandemic in 2020 and unavailability of required data, basic implementations of these models could not perform and give reliable results.

To overcome the predilection of the learning models on comparatively larger pre-pandemic historical data and the pandemic-induced sudden hike in unemployment rate, we propose a new computational model for time-series forecasting. We first held 10% of the data, and trained the models on rest of it. We also made two datasets - one with COVID-19 patterns and other without it. The COVID-19 patterns could be seen in Indian economics from the beginning of 2020. Since, we have the actual patterns, we combine the results of models on two datasets and minimize its difference with the actual values by finding the most approximate solution to the set of linear equations that we model. The validation-optimization pipeline could be well illustrated by Figure 2.

We aim to minimize the function $aP_1 + bP_2 - C$, where P_1 is the vectors of predictions by the forecasting model that ran on dataset with COVID-19 patterns; P_2 is the vectors of predictions by the forecasting model that ran on dataset without COVID-19 patterns; and C are the actual values for the held timeline. We then converted this equation to be represented using matrices, $\mathbf{AX} - \mathbf{C}$. Here, $\mathbf{A}$ is a $2 \times n$ matrix with each column containing P_1 and P_2 respectively, $\mathbf{X}$ is the vector of
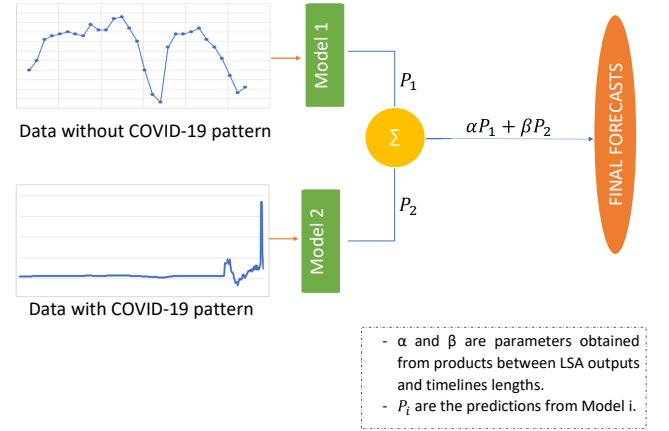

Fig. 2: Computational model for time-series forecasting to counteract the sudden COVID-19 pattern.

parameters a and b that need to be computed, and $\mathbf{C}$ is the vector of actual values. We then find an approximate solution to this matrix equation using Least Squares Approximation to obtain a vector approximation. Using this proposed technique, we could counteract the issue of sudden noise that got inserted into the data due to the pandemic. After getting optimal values of a and b, we retrain the models on full datasets, and then forecast for the next 13 months timeline by combining the results of two models using the model described in Figure 2. Though, the hike is total anomaly in the pattern but it cannot be ignored, so we smoothed down the predicted results of the two models using a heuristic based on the length of timelines for which the two models have been trained on.

Data: d_1 = time-series data 1; d_2 = time-series data 2
Result: Final optimum combined Predictions
m_1 = Prophet model trained on d_1
m_2 = Prophet model trained on d_2
P_1 = predictions with m_1
P_2 = predictions with m_2
C = actual observations
P = final predictions
$\widehat{x} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
$A = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$
Find $\widehat{x}$ such that $A^T A \widehat{x} = A^T C$
$P = aP_1 + bP_2$
Algorithm 1: Proposed Ensembling Procedure

6. Data Collection

There is an unavailability of proper data regarding unemployment rate in India prior to 2016. Adequate data started to be gauged from 2016 onward by CMIE (12), India. But this much

amount of data is not adequate and sufficient for training statistical and machine learning models - they need good amount of data to learn patterns. So, we also referred to World Bank and the International Labor Organization (ILO) for the data. After data aggregation and cleaning, the data before 2016 was just for the financial years and not for specific months, so we synthetically generated data for months using forward interpolation. The curated data are all numerical floats and sequentially distributed in time-series fashion over a period of December 1991 to July 2020. This data is broken into two chunks - one stalled until December 2019 and the other taken until July 2020. The latter data contains in itself the pattern introduced as a consequence of COVID-19.

Table 1: A snapshot of data.

Date	Unemployment Rate
1991-12-31	5.45
1992-12-31	5.50
1993-12-31	5.61

7. Results and Discussion

The experimentation and inferencing involves three steps. In the first step, we run the time-series forecasting framework on two types of datasets that we have created. Thereafter, we model the results in the form of matrix equations, and then in the last step, we combine the results and optimize it by finding suitable parameters *a* and *b* using Least Squares Approximation (LSA). The LSA utilises the Levenberg-Marquardt Algorithm (LMA)(13) underneath to find a vector approximation.

All the training and inferencing procedures were performed on Google's Colab Platform (14) with CPU runtime powered by a single core, two threaded Intel Xeon processor and 12.72 GB of RAM.

7.1. ARIMA and its flavors

The architecture that we had designed did not work well with ARIMA. ARIMA could not capture the sudden hike pattern present in the data that got introduced due to COVID. Seeing the inefficiency of ARIMA to pick up the trend in the data, we decided not to move forward with the ARIMA model. Though, SARIMAX could somehow learn the pattern, but still it was not capturing the non-stationarity of the data. Also, the Augmented Dickey-Fuller test (16) showed weak evidence against null hypothesis of the data being non-stationary, indicating that time series has a unit root and it is non-stationary. So, a better model was required to capture this non-stationarity in the data, and Prophet was an obvious choice for such type of data.

Time series are stationary if they do not contain trend or seasonality in the data. Stationary time series is usually easier to model. We have considered that if 'p-value' is less than equals 0.05 then the data is stationary else it is non-stationary. In our experimentation, data proved to be non-stationary naively, and it stayed the same after first step of differencing as well. It is

only after second iteration of differencing on the data, that it got to be stationary.

Table 2: Augmented Dickey Fuller Test results.

	ADF Statistic	p-value
Naive	1.614075	0.997899
After 1st diff.	-0.4406437	0.903088
After 2nd diff.	-4.0788587	0.001049459

7.2. SVR

Support Vector Regression or SVR did try to learn some trend from the data and was able to "nicely" fit on the existing data, but it was a completely different case with unseen data. The model seemed to not learn the intrinsic sudden trend from the dataset which was a consequence of COVID. SVR gave a static value for the entire unseen timeframe.

7.3. VAR

Vector Auto Regression (VAR) is a multivariate time series method. The model built through VAR represents, two or more time dependent variables that are linear combination of their respective past values as well as the past values of other variables present in the set. Since our dataset is univariate having only one time dependent variable that is the unemployment rate of each month, considering this case VAR model cannot be built on this dataset. Therefore, the predictions through VAR model were not taken into consideration.

7.4. LSTM

Long-Short Term Memory or LSTM performed better than SVR but still was not able to closely learn and fit the dataset. It learned the pattern behavior but was deviating in good numbers from the actual observations, resulting in a large mean squared error.

7.5. Prophet

Prophet proved to be good model in the current research. The model not only learned the pattern in the data, but also was able to pick up the anomalous trend that got introduced into the data as a consequence of COVID. It gave a mean squared error very close to the proposed model. Unsurprisingly, due to the high promising results of Prophet, it was chosen to build the ensemble model, acting as the prime algorithm in the proposed model.

7.6. Proposed Model

As discussed in aforementioned sections, we train a model on two chunks of data, so as to accommodate the sudden hike in the data, which occurred as a consequence of COVID-19, heuristically. Our dataset contains records up till July 2020. Model 1 is trained on dataset from 1991 to 2019, and model 2 is trained on dataset from 1991 to July 2020. The second model takes into account the case of COVID-19 unusual trend in the data. The first model's data is available yearly, so as discussed earlier, we ran forward linear interpolation to generate monthly data.After

Table 3: Parameters used in our heuristic for Prophet model.

Parameter	var_1	var_2
Parameter 1	-0.1113	1.1055
Parameter 2	-0.0547	0.5616
Parameter 3	-0.0565	0.5438

Fig. 3: Visual representation of table 4. Data after July 2021 is not available.

this, we ran Prophet with seasonality mode as *multiplicative* on both the datasets, and generated predictions for the time-frame ranging from December 2019 to November 2020. Thereafter, we combine the two predictions in three ways to obtain the best forecast pipeline, which not only takes into account the hike in the data but also smoothens out after the effect. The parameter set 1 outputted through the LSA proved to be the best hyperparameters for ensembling procedure.

$$
p = \frac{D1.length}{(D1.length+D2.length)}
$$
$$
q = 1 - p
$$
$$
where,
$$
$$
D1 \ is \ dataset \ 1 \ and \ D2 \ is \ dataset \ 2
$$
(8)

Three parameters were created out of the variables outputted from the least squares approximation algorithm. These parameters as enlisted under Table 3, are generated by generating three different product combinations of the parameters a and b with *data ratio* parameters p and q as is shown by Equation 9. The data ratios are the ratio of the timeline length of the two sets of data on which two different models are being trained.

$$
a, b = p, q
$$
$$
a_1, b_1 = a \times p, \ b \times q
$$
$$
a_2, b_2 = a \times q, \ b \times p
$$
(9)

Thereafter, predictions from two models are combined as $var_1 \times P1 + var_2 \times P2$, and three different sets of results are obtained. These results are generated as per our heuristic, and we choose the pair of parameters that show the least deviation from the actual data. First we checked the three different models on existing data to see how much they are affected by the sudden hike in the data. From Table 5, it is clearly indicated that model with parameter set 1 undoubtedly gave the least MSE as compared to proposed model with other two sets of parameters.

After checking the pattern followed by the models, we then use our model architecture heuristically combined with three parameters to predict on the future timeframe, that is from December 2020 to December 2021. The results are elucidated in Table 4. Model with parameter 1 uses vanilla parameters obtained from the Least Squares Approximation algorithm which does seem to give results close to prevailing in the future timeframe. The other two heuristic's parameters seem to give somewhat constant results. After comparing these with the current ongoing data, we conclude that the parameter 1 combination gives the best result.
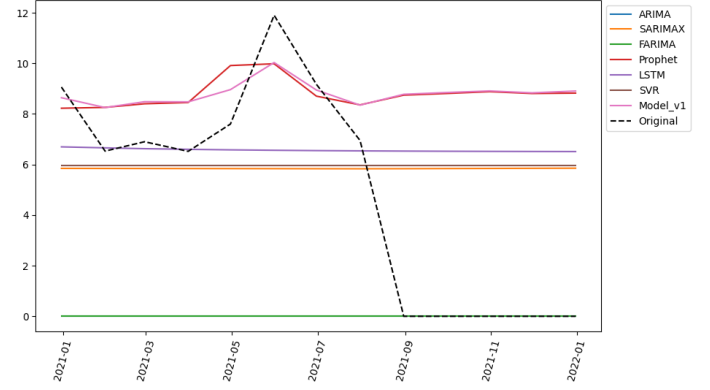
8. Comparative Analysis

8.1. Advantages over Other Vanilla Models

As shown in section 7, the vanilla models (except Prophet) could not correctly learn the pattern in the data. They were either over-hyped due to the sudden COVID-19 hike or they were just not being effected by this anomalous trend in the data. Likewise, ARIMA, FARIMA, LSTM and SVR could not learn the pattern , while SARIMA did try to learn and adjust to the pattern in the data, it still was unable to address the non-stationarity property that is present in the data, it just predicted strictly increasing linear model that was not good, since in future normal scenarios are bound to return and the predictions should not shoot up in those timeframes, rather they should become akin to the trend that was followed prior to the anomalous COVID-19 sudden pattern. As can be illustrated by Table 4, the predictions are hiked in case where COVID-19 pattern is present, and strictly decreases when this anomalous pattern is absent.

It can be observed that vanilla models could not learn the non-stationarity of the data properly. Even Prophet could also not learn this property because it was built to gauge seasonality and here in our data the unusual trend just comes for once, so it is not able to pick it up efficiently. Nevertheless, our proposed architecture is able to counteract this issue, and is successfully able to respond to the sudden COVID-19 pattern and also normalize itself thereafter in post-pandemic timeframe. As can be clearly observed from the plots in Figure 3 and Table 4, only Prophet and the proposed model's version 1 are able to perform best and give the best results. Comparing both, the proposed model wins, as can be inferred from the Mean Squared Error (MSE) of the models from the Table 5.

8.2. Comparison with Existing Researches

With the onset of COVID-19, many researches erupted in the academia, but they were mainly concerned with predicting the number of cases or in the pharmacological domain in finding new drugs for its treatment. Very few researches have been carried out to forecast the financial parameters of the nation. Unemployment rate is a deciding parameter in a nation's economy since it indirectly affects many spheres of the nation. With lockdowns being implemented, many people are at their homes out of which many lost their jobs, and this effect is apparent from

10. Future Scope

The current paper discussed a novel time-series forecasting approach that helps deal with two inherent issues that existed with the unemployment rate forecasting related problem domain, which are, unavailability of adequate data and sudden non-seasonal trend due to the COVID wave. The paper stands successful in addressing these two issues. By ensembling and optimizing the predictions produced by two models trained on separate time-series data produces a significant improvement over the vanilla time series forecasting algorithms. Further research can be conducted to make the ensemble procedure less error prone. The data before year 2016 can be interpolated by other means other than Newton's method that has been utilized in the current research, that might make the data more reliable and increase its information content. The proposed procedure proved to be good in unemployment rate forecasting, other than this specific application, the procedure can be utilized in other domains where normal vanilla models does not produce significantly good results.

References

[1] "Home — Ministry of Health and Family Welfare — GOI". mohfw.gov.in. Retrieved 4 July 2020.

[2] Sengupta, Sohini, Sareeta Mugde, and Garima Sharma. "Covid-19 pandemic data analysis and forecasting using machine learning algorithms." medRxiv (2020). doi: https://doi.org/10.1101/2020.06.25.20140004

[3] Gupta, Rajan, and Saibal Kumar Pal. "Trend Analysis and Forecasting of COVID-19 outbreak in India." medRxiv (2020). doi: https://doi.org/10.1101/2020.03.26.20044511

[4] Poddar and Yadav. "Impact of COVID-19 on Indian Economy- A Review" Journal of Humanities and Social Sciences Research, Horizon Journals (2020). doi: https://doi.org/10.37534/bp.jhssr.2020.v2.nS.id1033.p15

[5] Paul, Dhritabrata. (2020). "Covid 19 Impact on Indian economy". doi: 10.13140/RG.2.2.27275.23846.

[6] Katris, C. "Prediction of Unemployment Rates with Time Series and Machine Learning Techniques". Comput Econ 55, 673–706 (2020). https://doi.org/10.1007/s10614-019-09908-9

[7] Meron, D. "Modeling and Forecasting Unemployment Rate in Sweden using various Econometric Measures". Diss. M. SC. Thesis, Örebro University School of Business, Department of Applied Statistics, `https://www.diva-portal.org/smash/get/diva2:949512/FULLTEXT01.pdf`, 59-68, 2016.

[8] Fajar, Muhammad and Prasetyo, Octavia Rizky and Nonalisa, Septiarida and Wahyudi, Wahyudi (2020): Forecasting unemployment rate in the time of COVID-19 pandemic using Google trends data (case of Indonesia). Published in: International Journal of Scientific Research in Multidisciplinary Studies , Vol. 6, No. 11 (30 November 2020): pp. 29-33.

[9] Barbate V, Gade RN, Raibagkar SS. COVID-19 and Its Impact on the Indian Economy. Vision. 2021;25(1):23-35. doi:10.1177/0972262921989126

[10] Chand, Khem & Tiwari, Rajesh & Phuyal, Manish. (2018). Economic Growth and Unemployment Rate: An Empirical Study of Indian Economy. PRAGATI : Journal of Indian Economy. 4. 10.17492/pragati.v4i02.11468.

[11] Dumičić, Ksenija. (2015). Developing Forecasting Models for Unemployment Rate by Gender: Cross Countries Comparison.

[12] Centre for Monitoring Indian Economy Pvt. Ltd. https://www.cmie.com/

[13] J. J. More, "The Levenberg-Marquardt Algorithm: Implementation and Theory," Numerical Analysis, ed. G. A. Watson, Lecture Notes in Mathematics 630, Springer Verlag, pp. 105-116, 1977.

[14] Bisong E. (2019) Google Colaboratory. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4470-8_7

[15] Hyndman, Rob J; Athanasopoulos, George. 8.9 Seasonal ARIMA models. Forecasting: principles and practice. oTexts. Retrieved 19 May 2015.

[16] David A. Dickey & Wayne A. Fuller (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74:366a, 427-431, DOI: 10.1080/01621459.1979.10482531

[17] Prophet: Forecasting at scale. Facebook Open Source. https://facebook.github.io/prophet/

[18] Deb, Chirag, et al. "A review on time series forecasting techniques for building energy consumption." Renewable and Sustainable Energy Reviews 74 (2017): 902-924.

[19] Nguyen, Hieu M., Philip J. Turk, and Andrew D. McWilliams. "Forecasting COVID-19 Hospital Census: A Multivariate Time-Series Model Based on Local Infection Incidence." JMIR Public Health and Surveillance 7.8 (2021): e28195.

[20] Shankar, Sonali, et al. "Forecasting container throughput with long short-term memory networks." Industrial management & data systems (2019).