Note: ReGNL: Rapid Prediction of GDP during Disruptive Events using Nightlights

Rushabh Musthyala∗
BITS Pilani, Hyderabad Campus
India
f20180433@hyderabad.bits-pilani.ac.in

Hritish Jain∗
BITS Pilani, Hyderabad Campus
India
f20190356@hyderabad.bits-pilani.ac.in

Rudrajit Kargupta∗
BITS Pilani, Hyderabad Campus
India
f20170452@hyderabad.bits-pilani.ac.in

Dipanjan Chakraborty
BITS Pilani, Hyderabad Campus
India
dipanjan@hyderabad.bits-pilani.ac.in

ABSTRACT
Policymakers often make decisions based on GDP, unemployment rate, industrial output, etc. The primary methods to obtain or estimate such information are resource-intensive. In order to make timely and well-informed decisions, it is imperative to come up with proxies for these parameters, which can be sampled quickly and efficiently, especially during disruptive events like the COVID-19 pandemic. We explore the use of remotely sensed data for this task. The data has become cheaper to collect than surveys and can be available in real-time. In this work, we present Regional GDP-NightLight (ReGNL), a neural network trained to predict GDP given the nightlights data and geographical coordinates. Taking the case of 50 US states, we find that ReGNL is disruption-agnostic and can predict the GDP for both normal years (2019) and years with a disruptive event (2020). ReGNL outperforms time-series ARIMA methods for prediction, even during the pandemic.

CCS CONCEPTS
• Computing methodologies → Neural networks; Supervised learning by regression; • Applied computing → Economics.

KEYWORDS
Neural Networks, Remote Sensing, Geographical Coordinates, Gross Domestic Product (GDP), Nightlight

ACM Reference Format:
Rushabh Musthyala, Rudrajit Kargupta, Hritish Jain, and Dipanjan Chakraborty. 2022. Note: ReGNL: Rapid Prediction of GDP during Disruptive Events using Nightlights. In COMPASS ’22, June 29–July 1, 2022, Seattle, WA, USA. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3530190.3534849

∗The authors have made equal contribution to the research

1 INTRODUCTION
During disruptive events, like the COVID-19 pandemic, it becomes crucial for policymakers to quickly estimate the event’s impact on various parameters like GDP Gross Domestic Product (GDP), unemployment rate, industrial output, etc. Existing methods to collect or estimate these parameters are resource-intensive and time-consuming. Therefore, policy-making is done in hindsight. While the existing models are time-tested and work well in “normal” times, they fall short during disruptive events, when policymakers need methods with a quick turnaround time to estimate economic parameters and subsequently make policy decisions.

In this paper, we present ReGNL (Regional GDP-NightLight), a neural networks model to predict the GDP of given geography using Nightlights (visible lights being emitted by activities on earth’s surface, remotely sensed by sensors onboard various satellites).

Off-late, nightlights have emerged as a convenient proxy used to ascertain economic activity of a region [8, 14]. We develop ReGNL to estimate the GDP of given geography using the nightlights data from the same geography. We present our technique on datasets from different states in the USA. Our techniques are general and can be extended to other geographies as well. ReGNL is trained on the quarterly nightlights and GDP data from 50 US states from 2014 to 2018.

We test our model over periods where the economy was “normal” and during the pandemic to understand whether the use of nightlights and geographical coordinates are agnostic to such events. Furthermore, we compared our predictions with ARIMA to evaluate our model against time-series predictors.

The contributions of our work are:

(1) We curate a dataset combining the quarterly nightlights values, quarterly GDP values, and latitude and longitude of 50 US states from 2014-2020.

(2) We develop the ReGNL (Regional GDP-NightLight) model, a neural network that predicts the GDP of a place using the nightlights and latitude and longitude as features.

(3) Proof that nightlights can be used to make estimations agnostic of disruptive events.

Our results validate previous claims that nightlight can be used as a proxy for GDP and further asserts that it is robust even during disruptive events, when conventional data collection methods may not be possible.
The rest of the paper is structured as follows: Section 2 contextualizes our work within existing research in the domain. Section 3 describes the dataset creation process. Section 4 provides a detailed description of the approach and model developed as well as the details of the results obtained. Finally, conclusions and future research directions are given in section 5.

2 RELATED WORK

Liu, et al. [16] use a transfer learning framework on the premise that nightlights are an effective proxy for the economic activity of a region (Mainland China). The authors use satellite images and nightlight intensity levels as the input set of features to train their model. Otchia, et al. [17] use Machine Learning methodologies on nighttime data to study the industrial progress in Africa. Asher, et al. [5] found a statistically significant relationship between nightlight data and economic variables such as per-capita income and consumption, electricity coverage, population growth, and density. This working paper [15] found associations between differences in GDP values in different regions in North Korea and their respective nightlights. Lee also commented that satellite data is of great utility when it comes to assessing the economies of cities and regions. These papers provide motivation testing whether such models are robust during disruptive events.

Gallup, et al. [12] examined the role that the geography of a region plays in its economic progress. Their work highlights how the location and climate of a region impacts its economy through different facets such as agricultural productivity, transportation costs, and the burden of disease. Bickenbach, et al. [6] highlight that the relationship between the growth in nightlight and the GDP growth varies significantly, economically as well as statistically, across regions. They concluded that trying to draw a relationship between the two without accounting for the positional information (such as longitude and latitude) may lead to poor results.

After reviewing existing literature, we decided to start by evaluating the performance of a model for GDP prediction that takes in nightlights data and compare its performance when geographical coordinates are considered too.

3 DATASET

In this section we describe the data collection and processing pipeline. There are 3 components to the dataset: nightlight values, GDP values and geographical coordinates (latitudes and longitudes).

3.1 Nightlights Data and Geographic Coordinates

The dataset that we use in this work for obtaining the remote-sensing data is provided by “Earth Observation Group, Payne Institute for Public Policy, Colorado School of Mines” [10] and is accessed using the Google Earth Engine [3]. The dataset contains nighttime data in the form of Average Day Night Band (DNB) radiance. The average DNB radiance band, part of the VIIRS dataset, is a metric that ranges from -1.5 to 193565 units and is reported in nanoWatts cm$^{-2}$ sr$^{-1}$. Shapefiles contain the geometric information about the boundaries of a region like country or state.

We obtained the same from DIVA-GIS [1] and the US National Weather Service [4].

Using the VIIRS dataset and the shapefiles we obtain the nightlight information mapped to each of corresponding state along with the state’s geographic coordinates in the form of latitudes and longitudes.

3.2 GDP Data

The quarterly GDP values for the United States of America were obtained from the Bureau of Economic Analysis [2]. The monthly nightlight values were averaged into quarters and mapped to their corresponding GDP values. Furthermore, GDP Data is reported for a particular base year. This is to allow for purchasing power comparisons and to report the growth of a region while adjusting for inflation. To address this, the base year was chosen to be 2011 and adjusted the GDP values of all the regions accordingly.

We show the pipeline for dataset creation and modelling in Figure 1.

4 EXPERIMENTS, MODEL AND RESULTS

We aim to answer the following with regards to ReGNL:

(1) Does it work for both regular (economy-wise) years and years with disruptive events?

(2) Does the inclusion of geographical coordinates result in improvement in its performance?

4.1 Timeframe with Disruptive Event

The United States experienced a significant drop in the national GDP during the 2nd quarter of 2020, attributed to the COVID - 19 pandemic. To test our method across a regular timeframe and one with a disruptive event, we trained ReGNL on data from 2014-2018 and used it to make predictions for 2019 (regular year) and 2020 (year with a disruptive event).

4.2 Evaluating the Incorporation of Geographic Coordinates

To evaluate the value added by incorporating geographical coordinates in the dataset, ReGNL was trained under two different scenarios - one containing just the mean nightlight of a region and the other consisting of the mean nighttime and the latitude-longitude of the region. The metric used to compare different predictions was a weighted error. The error of each state was weighted by its contribution to the country’s total GDP to ensure that adequate importance is given to states in accordance with how high or low their GDP contribution is to the country. The error is computed using Equation 1.

$$\text{error} = \sum_{\text{states}} \left( \frac{\text{GDP}_{\text{actual}} - \text{GDP}_{\text{predicted}}}{\text{GDP}_{\text{actual}}} \right) \times \left( \frac{\text{GDP}_{\text{state}}}{\text{GDP}_{\text{national}}} \right) \times 10$$

The results are shown in Table 2. As observed in Table 2 we were able to obtain a significant increase in ReGNL’s performance by adding the latitude and longitude to our feature set. This is in line with previous works that have highlighted the role geography plays in a region’s economy.
Furthermore, to establish that latitude and longitude gave us a comparative advantage over knowing GDP trends in a state itself, we evaluated the performance of a model trained giving a categorical encoding to each state versus one with latitude and longitude. There was a random train-test split and MSE was used to evaluate performance as we omitted a few states from our training set to see if the model could learn from nearby states. The results are shown in Table 3.

Table 3: MSE when using a state as a variable v/s geographical coordinates

| Input Data                        | MSE  |
|----------------------------------|------|
| Nightlights + State              | 0.0173 |
| ReGNL (Nightlights + Lat, Long)  | 0.0034 |

These results support the idea that geographical coordinates in addition to nightlights can help in making more accurate predictions.

4.3 Model

Several Machine Learning algorithms such as Support Vector Regressors [9], Linear (and Polynomial) Regression, XGBoost [7] as well as Neural Networks [11, 13] were tested. We saw that the Neural Networks provided better and more consistent results when compared to the others. The results obtained using the different models are presented in Table 4. We also felt that NNs were more scalable if more features were to be added that have a non-linear relationship. Additionally, the architecture could be modified in the future to take advantage of temporal data too.

After running various tests with different architectures, based on train and test loss, the final model developed and trained was a feed-forward neural network consisting of 8 hidden layers. The ReLU activation function was used after each layer (barring the last one) to introduce non-linearity. Regularization techniques such as dropout...
Table 4: Comparisons of different algorithms in terms of their average weighted error of 2019-20 GDP predictions

| Method       | Error     |
|--------------|-----------|
| SVR          | 6.8221    |
| Linear Regression | 5.1546  |
| Neural Network      | 0.7066    |
| XGBoost       | 5.1436    |

and weight decay were used to avoid over-fitting the training data. The model was trained for $5 \times 10^6$ epochs with a learning rate of $10^{-6}$.

Predictions for the year 2020 using all 3 features (nightlight, latitude and longitude) were made using ReGNL and the results of the same can be visualised in Figures 3, 4 and their respective errors are shown in Table 5.

Table 5: Weighted error of predicted GDP values (2020)

| Quarter   | ReGNL (Nightlight, Lat, Long) |
|-----------|-------------------------------|
| 2020 Q1   | 0.7243                        |
| 2020 Q2   | 0.6879                        |
| 2020 Q3   | 0.6983                        |
| 2020 Q4   | 0.6478                        |
| Average   | 0.6895                        |

The error obtained across the year was consistent. The biggest takeaway is the low error from 2020 Q2. As shown in Figure 3, despite the GDP being hit drastically during this period, our model did not produce any anomalous result and continued to perform as hypothesized, in a similar fashion to as it was before the pandemic struck. These results provide evidence to support our claim that nightlight and geographical data can act as a reasonable proxy to determine the GDP for the United States, even during disruptive events such as the COVID-19 pandemic.

We employ ARIMA to predict the GDP in USA states for 2019 and 2020 as a baseline prediction estimate. We can also use this information to gain insight into whether the nightlight and geographical features of a region were able to serve as proxies to estimate the GDP more accurately. The results obtained through the ARIMA model are reported in Table 6. A comparison of the actual and predicted GDP values using ARIMA and our model (ReGNL) is shown in Figure 3. The ARIMA model overestimated the GDP values in 2020 Q2, whereas the estimations from the ReGNL model were closer to the ground truth.

While ARIMA performs well in the first quarter, it cannot take into account the impact of the pandemic and gives us predictions that are far off in the second quarter. ReGNL can perform much better.

Table 6: Weighted error of predicted GDP values using ARIMA (2020) vs error of predicted GDP values using ReGNL

| Quarter | ARIMA Error | ReGNL Error |
|---------|-------------|-------------|
| 2020 Q1 | 0.1235      | 0.7243      |
| 2020 Q2 | 1.4976      | 0.6879      |

Figure 3: The predicted GDP (ARIMA and ReGNL model) values vs actual values for USA Q1 2020
better during this period and provides more consistent results even in the quarters affected by the disruptive event of the pandemic.

## 5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose ReGNL, a neural networks-based method, to estimate the GDP of a geography using the nightlights data (obtained via remote sensing) in combination with geographical coordinates in the form of latitude and longitude. We demonstrate that adding latitude and longitude as features for our predictive model improves the performance. Using the USA as an example, we have also shown that the model is agnostic to the COVID-19 disruptive event. We used ARIMA as a baseline for comparison and found that our model outperformed ARIMA for the timeframe when the disruptive event affected the economy.

We hope this work opens the doors to further analysis on features that could be added to improve predictions. Some ideas include GDP data of nearby regions, historical data (using an RNN instead of a simple feed-forward network), and other remotely sensed indicators such as Carbon Monoxide emissions, NDVI (Normalized Difference Vegetation Index), etc. While this was a preliminary study, we aim to further it with more recent data, while expanding the methodology to different countries with the aforementioned changes to come up with more a more robust way to construct a GDP estimation model that is agnostic of disruptive events.

## REFERENCES

[1] [n.d.]. GDP by State | U.S. Bureau of Economic Analysis (BEA). https://www.bea.gov/data/gdp/gdp-state
[2] [n.d.]. Download data by country | DIVA-GIS. http://diva-gis.org/gdata
[3] [n.d.]. Google Earth Engine. https://earthengine.google.com/
[4] [n.d.]. US States and Territories. https://www.weather.gov/gis/USStates
[5] Sam Asher, Tobias Lunt, Ryu Matsunura, and Paul Novosad. 2021. Development Research at High Geographical Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the SHRUG Open Data Platform. The World Bank Economic Review (03 2021). https://doi.org/10.1093/wber/lhab003 arXiv:https://academic.oup.com/wber/advance-article-pdf/doi/10.1093/wber/lhab003/36643310/lhab003.pdf lhab003.
[6] Frank Bickenbach, Eckhart Bode, Peter Nunnenkamp, and Mareike Söder. 2016. Night lights and regional GDP. Review of World Economics 152, 2 (2016), 425–447.
[7] Tiang Chen and Carlos Guestrin. 2016. XGBoost: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Aug 2016). https://doi.org/10.1145/2939672.2939785
[8] Dave Donaldson and Adam Storeygard. 2016. The View from Above: Applications of Satellite Data in Economics. Journal of Economic Perspectives 30, 4 (November 2016), 171–98. https://doi.org/10.1257/jep.30.4.171
[9] Harris Drucker, Chris J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. 1996. Support Vector Regression Machines. In Proceedings of the 9th International Conference on Neural Information Processing Systems (Denver, Colorado) (NIPS’96). MIT Press, Cambridge, MA, USA, 155–161.
[10] Christopher D Elvidge, Kimberly Baugh, Mikhail Zhirin, Feng Chi Hsu, and Frank Bickenbach. 2016). https://doi.org/10.1145/2939672.2939785
[11] Terrence L Fine. 2006. Feedforward neural network methodology. Springer Science & Business Media.
[12] John Luke Gallup, Jeffrey D Sachs, and Andrew D Mellinger. 1999. Geography and economic development. International regional science review 22, 2 (1999), 179–232.
[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep learning. MIT press.
[14] Sungwon Han, Donghyun Ahn, Sungwon Park, Jeaark Yang, Susang Lee, Jiheer Kim, Hyoungjo Jung, Sangyoon Park, and Meeyoung Cha. 2020. Learning to Score Economic Development from Satellite Imagery. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (Virtual Event, CA, USA) (KDD ’20). Association for Computing Machinery, New York, NY, USA, 2970–2979. https://doi.org/10.1145/3394486.3403347
[15] Yong Suk Lee. 2014. Countering Sanctions: The Unequal Geographic Impact of Economic Sanctions in North Korea.
[16] Haoyu Liu, Xianwen He, Yanbing Bai, Xing Liu, Yilin Wu, Yanyun Zhao, and Hanfang Yang. 2021. Nightlight as a Proxy of Economic Indicators: Fine-Grained GDP Inference Around Mainland China via Attention-Augmented CNN from Daytime Satellite Imagery. Remote Sensing 13, 11 (2021), 2067.

[17] Christian Otchia and Simplice Asongu. 2020. Industrial growth in Sub-Saharan Africa: Evidence from machine learning with insights from nightlight satellite images. Journal of Economic Studies (2020).