*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 9 Issue 7 July 2021

Economic Quotient Estimation to Optimise Small City Connectivity to Boost Development

First Author: Yashil Vora – JamnabaiNarsee International School

Second Author - Reetu Jain Chief Mentor On My Own Technology Pvt Ltd

Third Author – Shekhar Jain CEO On My Own Technology Pvt Ltd

Abstract-Poverty is a huge issue in India and one of the major factors contributing to this is the inaccessibility of a region. Lack of connectivity of remote places gives rise to issues like high poverty levels, limited opportunities of employment, lack of medical and transport facilities, poor education and infrastructure, inaccessibility to resources during an emergency, slow response from nearby cities, and ill-equipped infrastructure to face an issue. Moreover, there is slow movement of goods and services from the village/town leading to slow economic development. By conducting research on the connectivity, infrastructure, and poverty levels of a region and the correlation between them, the paper aims to help solve this issue by creating a model and estimation of a coefficient for the target area. To solve the problem, the algorithm takes into account factors like highway density, internal road density, and the night light data of a region to predict an estimation of the economic quotient of that region. In addition to this, using the night light data, a connectivity model is built which gives the optimum connections of every village/town in the region to its nearest city. This has been done using machine learning, data analysis, and image processing on satellite images of that region with the help of computer vision. This data can then be used by the government or other private organizations to improve the welfare of the area.

Index Terms- Image Processing, Artificial Intelligence, Connectivity Model, Economic Quotient

I. Introduction

Poverty due to the inaccessibility of a region is a prevalent issue in India. The aim of this study is to help improve the connectivity of a region by coding an algorithm that analyzes satellite images of the target area, extracts data from them, and analyzes this data. With this, the algorithm then predicts an economic quotient for that area and builds a connectivity model for the region.

This pandemic has had a negative impact, especially on remote villages and towns, where there was limited health infrastructure and delays in doctors and appropriate equipment in arriving. Approximately 75 million people have been pushed into poverty because of this. Better connectivity could have helped reduce the impact of the pandemic on these regions and helped in their faster recovery. [1]

Research proved that the growth of highway lengths has only doubled between 1990 and 2013. This slow growth and poor infrastructure is the basis of many of the problems we face today.

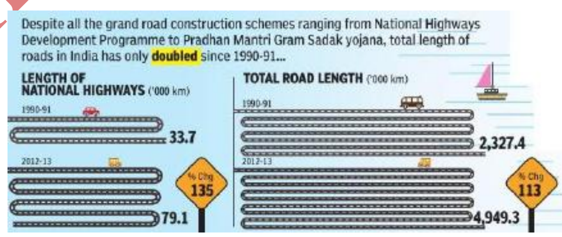

Fig 1. Length of highways between 1990 and 2013 [2]

Although the rural roads (roads within the rural regions) have grown in length over the past few decades, the total length of national highways has hardly changed and this is the major concern as it is these highways that connect places to each other and allow trade and movement of goods to occur].

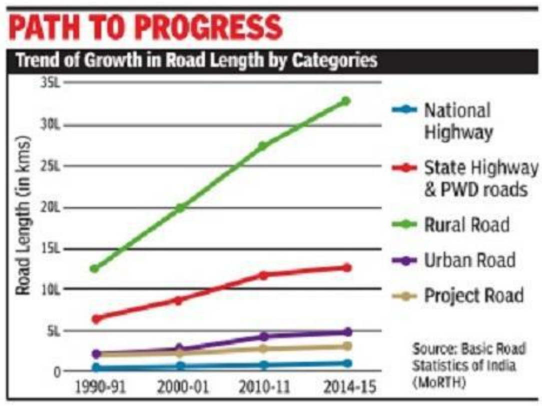

Fig 2. Growth of road lengths over the years [2]

*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 9 Issue 7 July 2021

Both of these figures are concerning as it indicates the faults made in the development of these regions and how inefficient decision-making can affect the connectivity of an area. It also proves that progress is slow due to dependence on surveys done manually by labour which are time-consuming and costly. This made me think of a more efficient solution that can help the government make better decisions to improve the welfare of these target regions.[3]

II. Research

According to a survey conducted by the Government of India in 2017 with evidence from 6 lakh villages - Remoteness continues to be a major and persistent predictor of low living standards. Each additional 10 kilometers from a town is associated with a 3.2% reduction in mean earnings, a 5.2% reduction in non-farm employment, and a 1.9 percentage points decrease in the proportion of literate residents. Moreover, 75% of healthcare infrastructure is concentrated in urban areas which cater to just 27% of the population.[4] This causes inefficiency and most of the time the people who actually require the resources and services provided by this health infrastructure are left helpless. This was clearly seen during the pandemic when India became the global hotspot for the virus and had the greatest number of infections and deaths in the whole world. Apart from healthcare, rural road connectivity is poor as well. In places like Ludhiana, children have to walk up to 12 km just to go to school every day. These kinds of barriers are seen in all aspects of their lives, ranging from getting a clean water supply to good sanitation facilities.[5]

All of these problems can be solved simply by improving the connectivity of that place and that is what this research paper aims to do. Improving the connectivity of a region can tremendously help improve the welfare of the towns/villages. They can receive aid from nearby cities more quickly during disasters, there is greater movement of goods and, materials in and out of that place, there are more opportunities of employment for the people living there as they can easily travel to the nearby cities to get a job and eventually this leads to lower poverty levels for that region too. With lower poverty levels, better infrastructure can be built within the target area which will eventually make them better equipped to face any challenge and give them greater financial independence in the long term. [6]

The 2 diagrams below further help to prove that this problem exists and that there is a direct correlation between the welfare of a region and its connectivity. Fig 3. and 4. indicate that in places with low highway density (East/ North-East of the country) the poverty levels are the greatest. Places with high highway densities have lower poverty levels (Northern and Western regions of the country). Thus, this correlation is a valid assertion, and the research to improve the connectivity by giving an economic quotient will help solve the problems that arise because of it.

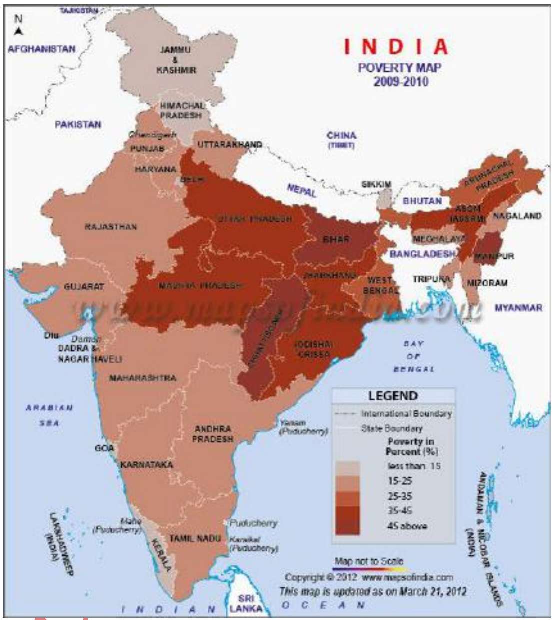

Fig 3. Poverty levels in India **[7]**

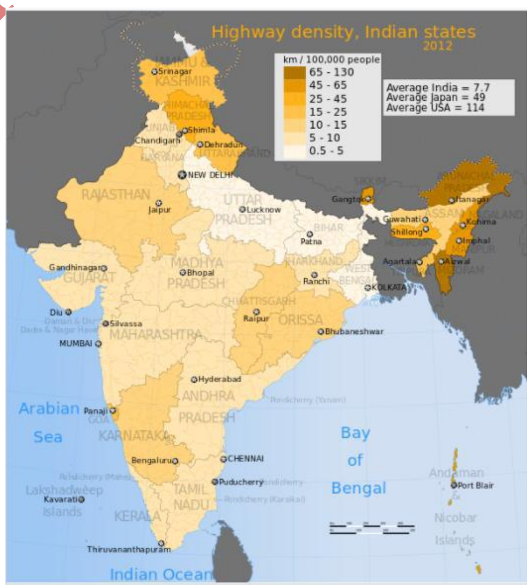

Fig 4. Highway density in Indian states **[8]**

There have been previous attempts to solve this issue using computer vision and satellite imagery as well. However, this research is unique due to the fact that it focuses on road infrastructure and connectivity to give the economic quotient. Moreover, it builds a connectivity model as well which can potentially help in government decision making, improve the allocation of resources and reduce costs.

METHODOLOGY

Primarily, I needed to find a reliable source from where I could get satellite images of the different locations with up-to-date data. After a lot of research, the website https://www.lightpollutionmap.info/ proved to have the most useful and latest data. This website used VIIRS 2020 data of satellite images and the night-light data and road and connectivity data could be found in one place by altering the transparency levels. Below are the 2 images taken of the same map at different transparency levels.

Fig 5. Night light data of satellite images

Fig 6. Roadways data of satellite images

After this, I needed to determine which factors to take into account while making the prediction and whether it is possible to extract the data from a satellite image of an area. Taking into account the impact the factor has on the connectivity (and thus poverty levels) and whether it could be extracted from a satellite image of the region, the 3 main factors chosen were the highway density of the region (in length of highway in kilometers per 1000 sq. kilometers of are), internal roads density (in length of highway in kilometers per 1000 sq. kilometers of area), and the night light data of the map which gives the areas with most and least activity (further in helping decide whether it is a major or minor region and how it can aid in the connectivity of the target area). These factors allow us to approach the problem holistically and make a well-balanced prediction that would give reliable results.

Now that a reliable data source was there and the factors to take into account were finalized, the actual coding began. Different parts of the algorithm were coded separately and compiled together at the end. I divided the algorithm into the following sections:

- Capturing and saving screenshots
- Building a connectivity model
- Calculating highway and midway densities
- Converting data extracted into a CSV file
- Use of artificial intelligence to calculate the economic quotient

Once each of these sections was coded individually, they were compiled together to produce a final algorithm that gives the relative economic quotient of the target area and a connectivity model for the target location. Below are the specifics of how each of these sections was coded and the role they play in the final algorithm.

Capturing and saving screenshots

The aim of this code was that when an input location is given, it would automatically go to the website (https://www.lightpollutionmap.info/), enter the location name, alter the zoom and transparency levels to take 2 screenshots- one for the night-light data (Fig 5) and one or the roadways data (Fig 6). On taking this screenshot using the snip screen feature, it would save it as “location1.jpg” and “location2.jpg”. This was coded using pyautogui, a module in python where the computer automatically performs actions based on what you code it for.

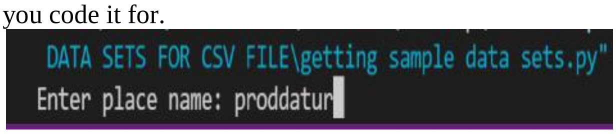

Fig 7. Input location to initiate algorithm

Building a connectivity model

This section essentially analyzes the night-light data image and extracts data from it using image processing. The following indicates the step-by-step procedure of how this works:

1. Using the night-light data satellite image, it converts the image into an HSV image, isolates the warmer colours (indicating human activity) using the threshold feature, and creates a black and white image of the areas with human activity and without human activity.

2. The contours feature in OpenCV is used to detect the contours and using the minimum enclosing circle function, the radius and x and y coordinates for each of the regions s obtained.

3. Using this, they are divided into major, minor, and tiny/insignificant regions.

4. Each of the regions is labelled with numbers and sorted in ascending order. The argsort feature is used here to sort the x and y coordinates with respect to the radius values.

5. A function called connecting_areas (which I coded) is applied in which with the input of the image, x, y, and radius values of the major and minor regions respectively, a connectivity model is build in which every minor region is connected to its nearest and largest major region. [This is done by calculating the distances between each region, finding the minimum distance and drawing a line between the 2 regions]
6.

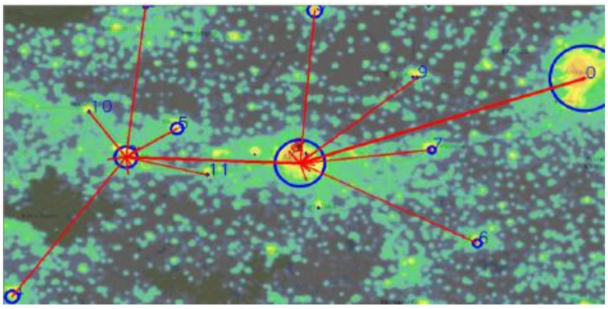

Fig 8. Connectivity model for Akola

Calculating highway and midway densities

This section calculates the highway and midway/internal roads densities by doing the following:

1. The roadways satellite image is taken, pixels with the colours wanted (purple for highways and white for midways) are isolated using the threshold function, and

Fig 9. Road data extracted from satellite image

2. The number of white pixels in this black and white image is counted for each of these respectively.

3. Using the scale of the map in pixels (105 pixels is 1 km) and the constant area of the picture (2028 X 1085 pixels), the area of the map in sq. kilometers is calculated.

4. Using the ratio of the length of the highway in the area, the highway length per 1000 sq kilometers is calculated so that a standard threshold can be used while calculating the economic quotient
5. The same is done for midways and both values are stored in the variables midway_density and highway_density respectively.

```
DATA SETS FOR CSV FILE\getting sample data sets.py"
Highway= 30.599999999999998 KM
Midway= 346.2 KM
```

Fig 10. Highway and midway length output

Converting data extracted into a CSV file

Now all of this data extracted from the images needs to be converted and stored in a CSV file. The following parameters are stored:

- Location name
- Highway density
- Midway density
- Number of major regions
- Average radius of major regions
- Number of minor regions
- Average radius of minor regions

This data is stored by using the CSV module in Python. Each of these parameters is stored in columns and every time a new location is targeted, the data is stored in a new row in the location_data file. Each row represents the data for an individual location. This helps the machine learning model learn as new locations are targeted and new data is obtained.

	A	B	C	D	E	F	G
1	location	highway_density	midway_density	no_of_major_regions	avg_radius_major_regiond	no_of_minor_regions	avg_radius_minor_regions
2	agar	57.85	142.69	2	48	13	4
3	akola	65.56	741.73	3	40	9	6
4	ambikapur	0	43.71	6	31	21	4
5	amreli	79.7	308.52	7	16	12	3
6	balangir	0	23.14	7	13	8	2
7	beohari	0	28.28	7	11	15	2
8	chanderi	41.14	10.28	7	8	11	1
9	darsi	34.71	44.99	2	22	15	3
10	khargone	60.42	86.13	9	19	16	3
11	kolkata	325.23	303.38	6	31	27	3
12	mumbai	241.67	280.24	5	28	16	3
13	nashik	173.54	737.87	4	32	19	5
14	pokaran	70.7	41.14	5	10	8	2
15	pune	287.95	419.07	3	30	16	3
16	renukut	59.13	111.84	5	29	32	3
17	sironj	28.28	66.85	4	31	48	3
18	ujjain	116.98	176.11	6	17	12	1
19	shivpuri	83.56	124.69	9	14	10	2
20	sidhi	71.99	47.56	10	10	17	2
21	shirpur	43.70663772	102.8391476	12	14	16	3

Fig 11. Stored data in the CSV file

Use of artificial intelligence to calculate the economic quotient

The CSV file is passed through a machine learning and data analysis model to calculate the economic quotient of the target area. The correlations are found using linear regression models and heatmaps. With the help of these correlations graphs are plotted and the relationships are used to estimate the quotient.

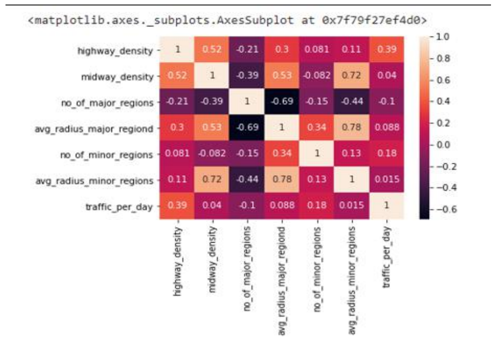

Fig 12. Heatmap for the correlations

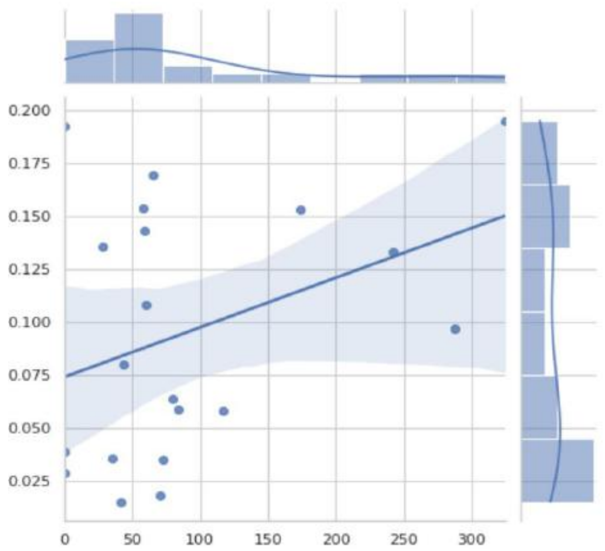

Fig 13. Highway density VS Economic quotient jointplot

The graph in Fig 13. shows a positive correlation between highway density and the economic quotient further reinforcing the assumption that they are directly related. The highway density with highest frequencies lie in around 50km per 1000 sq. km and the economic quotient with highest frequency is about 0.025. This is because most of the locations selected were relatively isolated from major cities to prove the effectiveness of the model.

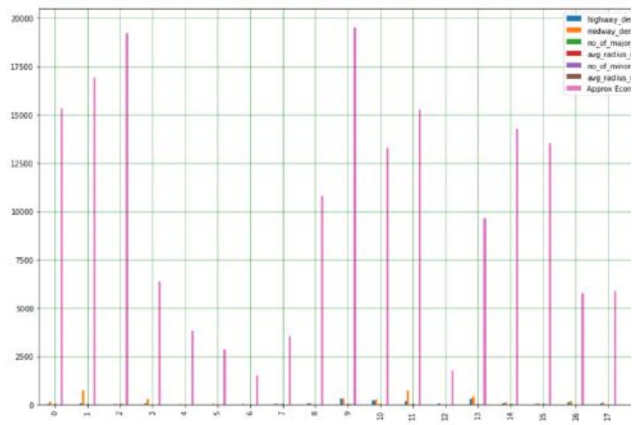

Fig 14. Graphs for recognizing the importance of each factor

By analyzing these graphs (Fig. 11 and Fig. 12), I could determine which factors contribute the most to the quotient and then create a formula to estimate it. I used Fig. 13 to understand the importance of each factor and it's the difference between the magnitude of the numerical values of the factors. This would help in creating the formula to estimate the quotient.

The aim was to estimate the quotient for the twenty locations for which I had collected the data by creating a formula which makes use of all the factors. This would help train the machine learning model to give an estimation of the quotient. The formula wasn't used to give an estimation directly for any locations in the future as the use of machine learning is essential in giving a quotient which can be used to compare the data between locations reliably. The formula was just used to give the model some data which would then keep on using data given to it in the future to train itself. The formula used is below:

Economic Quotient *= Highway density + Midway density + Area of all minor regions + Area of all major regions*

Area of all minor and major regions was calculated by using the below formula:

Area of all major/minor regions *= Number of major/minor regions * π * (Average radius of major/minor regions)^2*

This formula is a reliable estimation as, keeping in mind the relative importance of each factor and their correlations using the graphs plotted earlier, all of them are directly proportional to the economic prosperity of any location. If the highway and midway density is greater, there is greater movement of goods in and out of the location. Better transport allows the residents to have better education and employment opportunities. The location is quick to receive aid from nearby cities and will have greater access to healthcare facilities as well. If the regions around the target location consist of a large number of villages/towns, there is greater connectivity and again, greater movement of goods and more opportunities.

As most of the values were in the order of magnitude of 10^5 or lower, they were all multiplied by 10^{-5} so that they can be compared on a scale out of 1.

	location	highway_density	midway_density	no_of_major_regions	avg_radius_major_regiond	no_of_minor_regions	avg_radius_minor_regions	Calculated quotient	Usable Ec
2	agar	57.85	142.69	2	48	13	4	15330.004	0.
3	akola	65.56	741.73	3	40	9	6	16904.336	0.
4	ambikapur	0	43.71	6	31	21	4	19213.143	0.
5	amreli	79.7	308.52	7	16	12	3	6357.07	0
6	balangir	0	23.14	7	13	8	2	3840.0625	0.0
7	beohari	0	28.28	7	11	15	2	2877.6205	0.0
8	chanderi	41.14	10.28	7	8	11	1	1493.3685	0.0
9	darsi	34.71	44.99	2	22	15	3	3544.7745	0.0
10	khargone	60.42	86.13	9	19	16	3	10805.6595	0.1
11	kolkata	325.23	303.38	6	31	27	3	19505.8835	0.1
12	mumbai	241.67	280.24	5	28	16	3	13288.966	0.
13	nashik	173.54	737.87	4	32	19	5	15271.2065	0.1
14	pokaran	70.7	41.14	5	10	8	2	1783.118	0.
15	pune	287.95	419.07	3	30	16	3	9641.446	0.
16	renukut	59.13	111.84	5	29	32	3	14285.7295	0.1
17	sironj	28.28	66.85	4	31	48	3	13528.184	0.
18	ujjain	116.98	176.11	6	17	12	1	5778.149	0.
19	shivpuri	83.56	124.69	9	14	10	2	5875.516	0.
20	sidhi	71.99	47.56	10	10	17	2	3474.672	0.
21	shirpur	43.70663772	102.8391476	12	14	16	3	7987.729785	0.0

Fig 15. Final data for machine learning algorithm

This data could now be used in the machine learning algorithm to build a reliable model.

I used the Pytorch library of python for machine learning. The concepts used Gradient Descent and Linear Regression. The following were carried out in the algorithm:

1. Extract data from the csv file as a data frame
2. Divide the data into inputs and target tensors. The target tensor is the 'Usable Economic Quotient' column, whereas the inputs tensor consists of the 6 columns for highway density, midway density, number of major regions, average radius of major regions, number of minor regions and average radius of minor regions.
3. Use a data loader to randomly shuffle the data set in batches of 5 so that the model predicts values as accurately as possible.
4. Use the nn.Linear() action to generate random matrices for the weights and biases which can easily be used with the data set.
5. Create a loss function and optimizer for the data set using torch.nn.functional.mse_loss() and torch.optim.SGD() actions respectively
6. Define a function which automatically trains itself and uses gradient descent input data of the number of epochs, model, loss function, optimizer, and data loader are given to it.

To further elaborate on how exactly this model works, it uses the formula below to calculate the economic quotient:

$$Economic\ quotient = W_1*I_1 + W_2*I_2 + \ldots W_n*I_n + Bias$$

$I_1 \ldots I_n$ are the 6 inputs used here. Each input is multiplied with a weight and a bias is added at the end to make the final answer. The algorithm randomly creates a matrix for weights and biases of shape that can be multiplied by the inputs. The answers initially are not even close to the real value, but as the number of epochs (iterations) increase, the model uses gradient descent o move up and down the loss graph to minimize it by using the fact that if the gradient is positive, it should move to the left and if the gradient is negative, it should move to the right to minimize the loss.

As the data I had was limited (data for 20 locations), the predicted values after training the model were showing low accuracy as shown below.

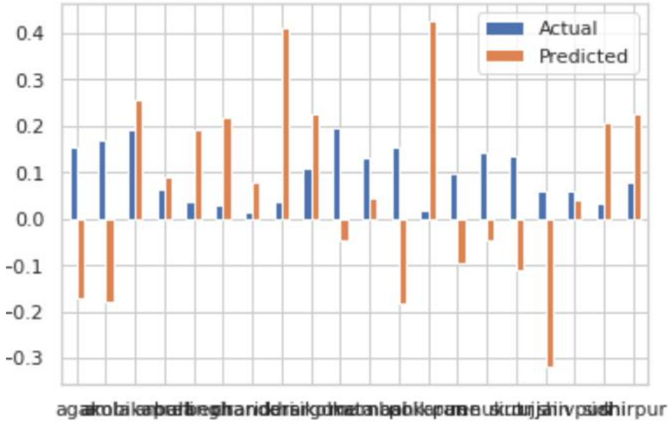

Fig 16. Difference between the predicted and actual values for the 20 locations

However, this deviation is expected and as data for more locations is collected in the future, the model will predict the quotient and use that value to further train itself.

III. CONCLUSION

It can be concluded that there is a direct relationship between the connectivity of a region and its economic welfare and poverty levels. The output of this algorithm produces reliable output. The connectivity model can be used by the government to make quicker and less costly decisions to develop road infrastructure for remote locations. The economic quotient can be used as well to determine which places require developments in infrastructure to improve their welfare. This will give the following benefits in the short run:

- A correlation matrix of economic coefficients can be used by the government to make resource allocation decisions.
- Government doesn't have labour costs to conduct surveys (which are very time-consuming) as the whole process is autonomous. Thus, time and costs are saved
- Efficient fund allocation by the government in targeted areas.
- Better disaster management abilities as the areas will become well connected to major cities if they need aid in case of emergencies.
- Greater financial independence as there is movement of goods in and out of the region, there are greater employment opportunities.
- City-wise cost-benefit analysis can be done using the economic quotients.
- Easier access to basic necessities like education

and clean water.

- In the future, there can be monitoring of how the welfare levels of the targeted region change with time and improve due to the greater connectivity.
- Improvements in the standard of living for locals and reduction in poverty levels.
- This can help the government do nation-wide development by providing resources to regions on a need-basis.

This solution can be implemented in other countries where such problems exist and a more efficient system for allocating resources to develop infrastructure is required. [9]

ACKNOWLEDGMENT

I would like to thank my mentors and On My Own Technology (OMOTEC) for providing me with the opportunity to work on this project. It allowed me to research numerous topics, through which I gained immense knowledge.

REFERENCES

[1] Y. Nee Lee, "Covid pandemic pushes 75 million more people into poverty in India, study shows," *CNBC*, 19-Mar-2021. [Online]. Available: https://www.cnbc.com/2021/03/19/covid-pandemic-pushes-75-million-more-people-into-poverty-in-india-study.html. [Accessed: 03-Jun-2021].

[2] K. Sharma, "National Highways: India," *Indpaedia*, 30-May-2021. [Online]. Available: http://indpaedia.com/ind/index.php/National_Highways:_India. [Accessed: 03-Jun-2021].

[3]A. Dangra, "The Missing Piece in India's Economic Growth Story: Robust Infrastructure," *Accelerating Progress*, 02-Aug-2016. [Online]. Available: https://www.spglobal.com/en/research-insights/articles/the-missing-piece-in-indias-economic-growth-story-robust-infrastructure#:~:text=We%20believe%2C%20for%20India%2C%20investments,on%20economic%20growth%20across%20sectors. [Accessed: 03-Jun-2021].

[4]*Rural Healthcare Sector : A Challenge yet to be resolved*, 02-Jul-2018. [Online]. Available: https://www.smilefoundationindia.org/Media/rural-healthcare.html. [Accessed: 03-Jun-2021].

[5]P. Tajane, "Covid-19: Managing the disaster in rural India," *India Water Portal*, 17-May-2020. [Online]. Available: https://www.indiawaterportal.org/articles/covid-19-managing-disaster-rural-india. [Accessed: 03-Jun-2021].

[6] S. John, "Improving Connectivity Across Rural India," *World Bank*, 10-Apr-2104. [Online]. Available: https://www.worldbank.org/en/results/2014/04/10/improving-connectivity-roads-rural-india. [Accessed: 03-Jun-2021].

[7] "India Poverty Map," *Maps of India*, 18-Dec-2019. [Online]. Available: https://www.mapsofindia.com/maps/india/poverty.html. [Accessed: 03-Jun-2021].

[8] "Roads in India," *Wikipedia*, 22-May-2021. [Online]. Available: https://en.wikipedia.org/wiki/Roads_in_India. [Accessed: 03-Jun-2021].

[9]Indpaedia.com. 2021. *National Highways: India - Indpaedia*. [online] Available at: <http://indpaedia.com/ind/index.php/National_Highways:_India> [Accessed 31 May 2021].

AUTHORS

Yashil Vora, Student, JamnabaiNarsee International School, yashil2004@gmail.com

Reetu Jain, Chief-Mentor and Co-founder of On My OwnTechnology Private Limited Mumbai, Email:reetu.jain@onmyowntechnology.com

Shekhar Jain, Chief Executive Officer and Co-founder of On My OwnTechnology Private Limited Mumbai, Email:shekhar.jain@onmyowntechnology.com