# Analysis of Influence of Internet Inclusivity On Development of Countries

P Dinesh Kumar
*Department of Computer Science*
*PES University*
Bangalore, India
PES2UG20CS236

Reshmi Pradeep
*Department of Computer Science*
*PES University*
Bangalore, India
PES2UG20CS270

Rimzim Sanghvi
*Department of Computer Science*
*PES University*
Bangalore, India
PES2UG20CS273

ABSTRACT

*Internet access is not a luxury, but a basic necessity for economic and human development in both developed and developing countries. Our aim is to study the impact of internet inclusivity on different developmental indices which show theimpact of higher internet access on the overall development ofa country or region. In this initial report, the preliminary understanding of this topic after having reviewed relevant research papers has been included as well as the initial approach to cleaning and processing the data and our basic approach to solving the problem.*

*Index Terms—Access, Affordability, Development, Inclusion, Index, Internet, Gender Gap, Relevance*

## I. Introduction and problem statement

The internet is easily one of the most important tools in the world today. It is a powerful tool for the delivery of essential services such as education and healthcare, offers increased opportunities for women's empowerment and environmental sustainability, and contributes to enhanced government transparency and accountability. It also helps foster the social development of communities, including within the broader global context.

The increased reliance on the internet has been observed to have increased the rate of development in some regions and it's lack has been observed as a significant barrier development and betterment of that country. Access to the internet is essential for the overall growthand development of a country or for individual gain.

From our analysis of the dataset and the review of some other relevant literature, we note that a large number of factors affect the access to internet by different groups of people. These factors could be economic factors such as GDP or per capita income, human factors such as age, ICT skills, education levels or simply a lack of awareness. It could also be an issue with policy where countries with more free markets tend to have higher percentages of people with access to the internet or statistics on access to electricity in rural and urban areas. Our dataset has columns similar to these that can be used to determine an index of internet access in that region.

Our goal is to understand the impact of Internet inclusivity and accessibility on the aforementioned developmental indices to understand how increased access to the Internet can impact different spheres of life in a particular country.

## II. Review of Literature

As part of our literature survey, we reviewed research papers which have in common with our own approach either thedataset used or the areas of interest under study. The takeaways from each of these papers and shortcomings we wish to overcome in our approach and a brief explanation are given below.

[1] is the 2021 edition of the annual Internet Inclusivity index methodology report published by The Economist. This paper offers legitimacy to the information present in the dataset that we are using for this problem statement by citing the relevant sources and researches and how the sampling was done to collect the enormous amount of data. This report assesses the progress that countries have made and looks at the obstacles still preventing roughly half of the world's population from using the internet. The main technique used in this paper to analyze the data is by assigning weights to each attribute to measure its contribution towards the index. The influence of each attribute on the index varies based on the weight. Using this approach, we will be using it as a base principle to analyze the influence of these attributes on different development indices that are relevant to our problem statement.

[2] discusses the impact the spread of the Internet on the economic growth and development of a country. According to the paper, the Internet accounted for 21 percent of the GDP growth in mature economies over the past 5 years. They used various graphs to explain the correlation between the increase in the use of the internet and how it impacted the economy of that particular country. This paper is very useful to understand the kinds of models that could be utilized when analyzing different factors that impact the spread of the internet. It also highlights the need to access and analyze "big data"—the large datasets generated from every customer interaction, every wired object, and every social network.

[3] talks about how the introduction of internet into society is expected to moderate inequality by ensuring that everyone has access to inexpensive knowledge without restriction. The network is expected to significantly impact scholarship by improving global cooperation. The network's global dissemination has increased at a rate that exceeds 50% annually. The OECD is the perfect example that led to variations in international Internet access among nations with fairly similar socio-economic development. Analyzing the level of connectivity across long-standing democratic societies is a first step in understanding the potential global impact of the Internet. It specifically mentions the

Gini coefficient and the Global Peace Index as some of the major factors. One drawback of the paper is that it was published in the year 1999 during the early stages of the Internet and it may not be as relevant in today's context.

[4] The author makes the case that using the Internet in a way that will benefit society as a whole, and vulnerable individuals in particular, is the only way it can be used as a tool for social and economic growth. These include making it possible for important knowledge to be created on the Internet and for developing nations to acquire the knowledge, skills, and talents necessary to use it efficiently. These results imply that increased connectivity among developing nations would enhance the information infrastructure as a whole. Alternative aspects of development like human autonomy, equity, sustainable development, empowerment, and cultural identity are increasingly being considered in the UNDP Reports on Human Development.

[5] highlights the effects of infrastructure and internet inclusiveness on e-commerce, e-business and their revenue models on the national level between Internet infrastructure and the growth of E-Commerce does exist is not a matter of conclusion in this study. They assumed that there is a relationship between these variables such that better Internet infrastructure should facilitate E-Commerce revenue growth based on the results of the 2019 Network Readiness Index analysis. This part of the analysis shows the relationship between Network Readiness Index (NRI) and GDP Per Capita. This paper helps identify analytical techniques that could be used based on their methodology. It involves a studying the relationship between the four collective categories of features with the target variable and between each baseline statistic and the target variables. Each layer of analysis involves performing a t-test and finding the significance of the feature of interest using the p-value.

[6] is very useful to directly derive the relationship between national development and the increasing use of internet and its reliability. The cost structure of various internet businesses gives rise to various types of scale economies. The massive data collected by internet platforms have created a new branch of economics— nano-economics-which studies individual and computer-mediated transactions. The paper also has various graphs showing the relations between platforms on the internet, users and vendors and the transactions and interactions that take place between them. They also highlight the main three features: inclusion, efficiency and innovation which show how the internet impacts development.

## III. PROPOSED SOLUTION

### A. Dataset

The dataset we have used is called the Inclusive Internet Index that contains information spanning from 2017 to 2021 represented by different versions or editions. It contains data about access to internet by different people in different regions with different socio-economic situations. It also contains a large number of indicators that are used to generate an index for how much the internet is accessed in that particular region. The dataset contains 85 columns and 600 entries. 120 countries are selected form a diverse selection of high, low and middle-income countries which represent 96 percent of the world's population. Indices like the Gini Coefficient and the Global Peace Index are show different developmental factors. These may be used to calculate the impact of the Internet on development as a whole.

### B. Data cleaning

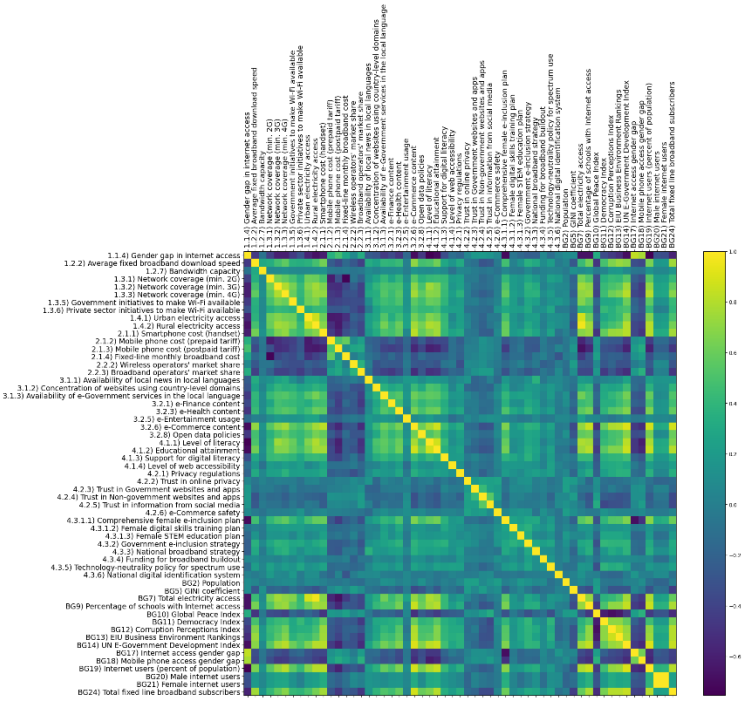
Incomplete data were taken care off by dropping them off the dataset as imputing or aggregating values for these columns would distort inferences, so these columns were also dropped.

Presence of outliers constitutes noisy data. We performed outlier analysis by data visualization using box-plots. The data points present outside the minimum and maximum range of the box-plot are considered to be outliers and were removed from the dataset.

### C. Data Pre-processing

Irrelevant data were all dropped from the dataset like columns that gave information on a nation but not related to internet within the nation.. We also dropped attributes that were redundant. There were multiple attributes in the dataset that described internet inclusivity with respect to gender and were also repeated with very little variations. Also noticed that data from the year 2017 , i.e., the first edition of the dataset, had a lot of missing and inconsistent data, so we dropped it from the dataset.

We also used the corr() function to generate a correlation matrix, which is a table used to show the coefficients of correlation between variables. Noted certain strong positively and negatively correlated variables from this graph. This can be used to avoid multi-collinearity in our models. It is to be noted that one of our target variables, GINI coefficient is weakly correlated to most of the other attributes. This could pose as a difficulty while building a model for this target variable.



*D. Data visualization*

For each of the categorical variables, we plotted a bar chart with frequency on the Y- axis which shows the relative spread of data. For numeric/ quantitative data, we plotted a histogram for each attribute which show the shape of the distribution such as number of peaks, uniformity, symmetry and skewness.

*E. Building Models and Evaluation*

We chose four target, dependent variables: Gini Coefficient. Corruption Perceptions Index, Democracy Index and Global Peace Coefficient and built four different models on them based on multiple linear regression – ElasticNet, Lasso and Ridge Regression and MLR and then trained and tested them on the split dataset in the ratio 80-20%.

For each target variable, the model that gave us the best performance is chosen. The performance of the models were evaluated based on metrics like R-square, RMSE, adjusted R-square, etc. Grid search is also used to find the best hyperparameters for lasso regression.

Attributes with high correlation with the target variables were chosen and used to build the models based on the correlation matrix. We also plotted graphs of the residuals against the errors in prediction to check if the models are homoscedastic and according to the test, all the models are homoscedastic.

## IV. COMPONENTS AND INFERENCES

*A. Gini Coefficient*

The Gini coefficient or index is used to measure the amount by which income earned per household is different from an income which is equally distributed. All the regression models tried for this target variable gave poor performances as can be seen from the low MLE and give him my blood.

The elasticnet regression model we used for Gini coefficient has a little high RMSE value and a low adjusted R-square values. Like we had noted earlier, the Gini Coefficient is only correlated to total fixed line broadband subscribers and is weakly correlated to every other variable. This is probably why the model underperforms on this variable. Hence you can say that inequality does not really depend on internet inclusivity.

| Model | RMSE | Adjusted R-square |
|---|---|---|
| Elastic Net | 6.45 | -0.02 |
| Lasso | 6.58 | -0.03 |
| Ridge | 6.73 | -0.05 |
| Multivariate | 6.89 | -0.03 |

*B. World Peace Index*

The world peace index is used to measure sustainable peace within a region. It is directly proportional to the peace within the country.

We used ElasticNet Regression model for this as well as it was the best performing one. All the models that were tried for this variable gave low RMSE values but high or average All the attributes with high coefficients have negative values which would mean that the country has a lower level of violence when there is a higher value of specific attributes. It can be noticed that the higher the number of internet users, the lower the index which means that country is more peaceful then. Similarly, the Global Peace Index is also dependent on the Level of Web Accessibility and Percentage of schools with Internet Access. Hence, a larger engagement with the internet leading to a more peaceful society is again inferred. The Democracy Index was best learnt by the Ridge model, but all models yield a low RMSE score. The adjusted R-square value, however, while being better than those of the models built for the Peace Index, are still not very high.

| Model | RMSE | Adjusted R-square |
|---|---|---|
| Elastic Net | 0.37 | 0.32 |
| Lasso | 0.41 | 0.41 |
| Ridge | 0.45 | 0.55 |
| Multivariate | 0.54 | 0.58 |

C. *Democracy Index*

The democracy index is used to measure of the quality of democracy and to understand the biggest threats against a sustainable democracy. The data tells us that a decrease in the number of total users and number of female users leads to an increase in the Democracy Index. Although this might seem like something outdated and ancient, it can be explained by the large amount of misinformation coming from social media. This conclusion can further be solidified by the negative correlation between Democracy Index and trust in social media.

The data also shows that an increase in number of male internet users leads to an increase in Democracy Index. This discrepancy between the usage of internet by males and females points to the gender gap in mobile phone usage, education attainment and internet access. The model used in this case is Ridge model as it had the best performance and it.

| Model | RMSE | Adjusted R-square |
|---|---|---|
| Ridge | $4.34 \times 10^{-15}$ | 1.0 |
| Lasso | 1.45 | 0.76 |
| Elastic Net | 1.73 | 0.64 |
| Multivariate | 1.89 | -0.03 |

*D. Corruption Perception Index*

The corruption perceptions index is used to measure the public sector corruption perceived by the general populace worldwide. The best model for this target variable is the standard MLR model which gives the bets RMSE score and a pretty good r-squared value. The greater the value of CPI, the better the sector as viewed by the public and has a negative relation with the number of female users and negative relation with and male users.

It can also be noted that largely affected by variables related to gender, the availability of facilities to access the Internet and media, the awareness of the general public and the relationship it has with the Government. Countries that wish to be known as good and less corrupted should focus on removing the barriers that exist

to Internet access (even gender gaps) electricity or non-lucrative business environments.

| Model | RMSE | Adjusted R-square |
|---|---|---|
| Multivariate | 10.6 | 0.703 |
| Lasso | 11.56 | 0.72 |
| Elastic Net | 11.89 | 0.81 |
| Ridge | 40.98 | -5.56 |

On the whole, the model that has the best performance is elastic net as it gives both good RMSE and r-squared values, Ridge and MLR have pretty average/good performance as well with a few exceptions. If the models were build only on the variables that have high correlation, we would notice that RMSE values increased and adjusted R-squares decreased for both the variables using elastic net regression, i.e., Gini coefficient and World Peace Index. For the corruptions perception index, the RMSE value decreases quite a bit by decreasing the r-squared value a little.

## V. CONCLUSION

We can see from all the analysis done in this project that bridging the gap between the genders in terms of accessibility of internet and education can actually cause a big impact on the society and in turn the national development. We can also enhance the national development indicators used in the project – Gini coefficient (economic), World Peace Index (social) and Democratic Index (political). The regression model we built performs the best with the democracy index, giving us almost zero RMSE and exactly 1 adjusted r-squared value. Hence, internet inclusivity plays a major role in the development of a nation.


## ACKNOWLEDGMENT

We would like to acknowledge our Data Analytics Course Professor. Prajwala T R for providing help and clearing up concepts. We would also like to thank PES University for granting us the opportunity to undertake this project and help us acquire more knowledge of the subject in the process.


## CONTRIBUTIONS

Rimzim Sanghvi: Data cleaning, writing Phase 1 review, literature review papers 3 and 4, model evaluation

Reshmi Pradeep: EDA, literature review papers 1 and 2, building regression models, inferences

P Dinesh Kumar: Data visualization, literature review papers 5 and 6, inferences, initial model builds

## REFERENCES


[1] The Inclusive Internet Index 2021: Methodology report, The Economist Intelligence Unit

[2] The impact of the Internet on economic growth and prosperity - James Manyika and Charles Roxburgh, October 2011

[3] Weaving the Western Web - explaining differences in Internet connectivity among OECD countries - Eszter Hargittai, Princeton University

[4] The Internet and Socio-economic development: Exploring the interaction – Shirin Madon, London School of Economics

[5] Infrastructure, Internet inclusiveness and e-commerce: An exploratory study - Authors: Frederick Augustine, Stetson University, John Rasp, Stetson University, Giao P. Nguyen

[6] Enabling digital development – how the internet promotes development - WORLD DEVELOPMENT REPORT 2016