Analysis of influence of Internet inclusivity on national development

Raghav T Kesari
Department of Computer Science
PES University
Bangalore, India
raghavtkesari@gmail.com

Divya Shekar
Department of Computer Science
PES University
Bangalore, India
divyashekar39@gmail.com

Krithika Ragothaman
Department of Computer Science
PES University
Bangalore, India
krithika.ragoth@gmail.com

***Abstract*—The Internet is one of the most pivotal technological inventions in the world today. It acts as a medium for connecting people across the world to perform a wide variety of tasks ranging from using social networks to buying your daily groceries. Our aim was to study how different groups of people across the world access the Internet and to what extent the people are connected. We specifically looked at the impact of internet inclusivity on different developmental indices which show the impact of higher internet access on the overall development of a country or region. We compared different models and studied each target variable to understand how each target variable is impacted. We then drew inferences based on the models we built to understand which specific factors affect our target variables.**

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

I. Introduction

The Internet has emerged as one of the most important tools in the world today. The Internet has vast utility which helps connect the entire world. The Internet can be used to buy and sell different products, meet new people, connect with people remotely and a plethora of other applications.

We have witnessed the shift from the Internet being a commodity to it becoming a necessity due to the increasing reliance on an interconnected world. We now require products and services which are procured from around the world or exist only virtually which would make the need for an ecosystem such as the Internet much more relevant. Because the Internet has such extensive use spanning many fields, we see that studying the access to Internet across the world becomes a relevant field of study.

The increased reliance on the Internet has resulted in Internet access to different groups, or the lack thereof, being observed as a significant hindrance to the growth and development of that particular region or country. This growth could be in terms of the economic security of the people of a particular region, the net GDP of that particular region or many other things.

We surveyed different research papers on topics relevant to the subject of Internet access in different countries across the world. The papers focus on different kinds of studies into the field of internet access and their varying insights in terms of the method of study, the different means of collection of data and many other aspects. This paper provides a comparative analysis of different models to understand how different factors related to internet access affect the development of the country which is indicated by the target variables in our dataset which include the Gini Coefficient, Global Peace Index, Democracy Index, Corruption Perceptions Index and the UN e-Government Index.

II. Review of Literature/ Related Work

As part of our literature survey, we reviewed research papers which have in common with our own approach either the dataset used or the areas of interest under study. We document below the existing research in the field of studying internet inclusivity and the impacts it has on the various metrics of development for a country, the takeaways from each of these papers and shortcomings we wish to overcome in our approach.

[1] is the 2021 edition of the annual Internet Inclusivity index methodology report published by The Economist. This paper first and foremost offers legitimacy to the information present in the dataset by citing sources and nature of sampling for data collection. The features were selected based on EIU analysis, a literature review, and consultation with industry experts and specialists from academia and NGOs.

This paper provides an in-depth introduction to the basis of selection and nature of the various attributes present in the dataset. As further elaborated in section III-A, the selected features used as baseline statistics are divided into four categories. An elaborate tabulation of each attribute, the category it falls under, the type of variable, what it represents and the source it is collected from is also included.

The principle data analytical technique used in this paper is the application of weights to each attribute to measure its contribution towards an Internet Inclusivity Index. Accordingly, the influence each attribute has on the Index varies. We use this principle as the basis for our approach, in which we combine or retain attributes on the basis of their overall significance. We wish to study the influence of these attributes not on the Inclusivity Index but on various development indices mentioned in section III, which demands an adjustment of said weights.

[9] provides a general comparison of countries on basis of the four aforementioned metrics, by tabulating the best and worst performing countries for each metric. It also calculates the Pearson correlation between overall index and various attributes such as land area, population density, and urbanization percentage. It uses the Internet Inclusivity Index, which is the same dataset that we are using. This paper provided us with a decent starting point in terms of the ways we could utilize the data to further build our problem statement. One problem was that it was essentially a surface level analysis, and lacked detail.

[2] discusses the factors that impact the spread of the Internet and it also discusses the correlation between the development of a country and the extent of its internet connectivity. The author proceeds to create correlation and OLS regression models to conclude which factors are more important by attempting to optimize the fit of the models.

This paper is very useful to understand the kinds of models that may be utilized when analyzing different factors that impact the spread of the internet. It also gives insights on the potential impacts of the Internet and how we may use developmental indices from the UN to gauge the impact of the Internet. It specifically mentions the Gini coefficient and the and the Global Peace Index. 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.

[8] offers a better alternate by consolidating the findings and inferences that link internet inclusion to the vulnerable stakeholders within a nation. These include but are not limited to levels of digital engagement, patterns between social exclusion and digital engagement and ITC poor environments, of which internet inclusivity is a subset. This paper deals with the micro effects of digital engagement on the general population. We use these findings as a basis to scale inferences up to a macro level.

[4] studies the effects of infrastructure and internet inclusiveness on e-commerce, e-business and their revenue models on the national level. Although deviating from our area of interest of study, we takeaway the techniques used to perform data analysis.

Prior to performing the data analysis, the existing categories present in the dataset as explained in III-A are further divided into sub-categories. Each primary category is expressed as a combination of two or more sub-categories. Each sub-category, in turn consists of various baseline statistics, which are the attributes present in the dataset.

The primary takeaway from this paper is the analytical technique used in the methodology. It involves a two-level analysis. The first level studies the relationship between each of the four collective categories of features with the target variable. The second layer delves deeper into studying the relationship between each baseline statistic and the target variables. Each layer of analysis involves performing a t-test and inferring the significance of the respective feature(s) using the p-value.

From this categorisation, we draw inspiration for our own approach to re-order the categories of features to fall under factors such as economic, sociological, financial and literal. This would imply studying the collective effects of each of these factors on development metrics.

[4] aims to study the influence of internet access on different forms of well-being, activity and social development among adults of various ages in economically deprived communities. It aims to answer the question: Does being able to use the internet (via mobile or broadband connection) positively influence both individuals and communities, particularly those among the lower spectrum of development? The paper also aims to specifically study the effects of internet use on those aged 65 years and older.

In [5], the sections of the study were Internet Access, Social Contact and Support, Use of Amenities, Sense of Community, Well-being and Physical Activity.The paper selected records with specific answers to each question, thus selecting a subsample of 3833 records.

A logistic regression model is used for target variables with binary outcomes, with the odds ratio corresponding to the likelihood of a positive outcome (e.g., frequent social contact, not feeling lonely). A regression model using OLS was used to analyse WEMWS scores. Measures of Interest, Loneliness and Well-being were regressed, and adjusted for covariates such as the social, amenities, and community outcomes. Note that all regression models were built against the Internet Access measure.

The study conclusively shows the positive effect of internet access for people in deprived communities, specifically when it comes to social contact, financial social support, and use of social amenities and shops. It also concludes that internet access has a positive effect on the mental well-being of an individual. Finally the paper notes the disparity of internet access along the lines of age, education, and physical ability.

This paper provides us with a great idea of the statistical models to use whilst analysing our data. It also inspired us to develop our problem statement to study the effects of Internet Inclusivity on other factors or measures of a nation, instead of studying Internet Inclusivity alone. This once again points to the macro-level scaling of our approach as opposed to the effects of internet inclusivity in the daily lives of the stakeholders.

[3] discusses the correlation between certain factors and the extent to which a country is developed and it also discusses the correlation between the development of a country and the extent of its internet connectivity. The paper studies a dataset of 18 countries, and uses attributes in many categories such as economic situation, human capital, the legal environment and existing technological infrastructure. The author proceeds to create correlation and OLS regression models to conclude which factors are more important by attempting to optimize the fit of the models. This paper is very useful to understand what kinds of models may be utilized when analyzing different factors that impact the spread of the internet. It also gives insights on the potential impacts of the Internet and how we may use developmental indices from the UN to gauge the

impact of the Internet. 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.

III. Proposed Solution

Access to the internet is essential for the overall growth and development of a country or for individual gain. From our initial 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.

When we look to measure the development of a country, we look at different spheres of development which could be in economic terms or human terms or the extent of inequality there is in the population. There exist different indices to measure this such as the World Peace Index, the Gini Coefficient, the World Democracy Index etc.

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. These could include economic growth, the degree of inequality in a country, the happiness of the people and in many other ways.

Parameters considered are the attributes mentioned under the dataset section, which, after data cleaning and preprocessing, totals to 52 parameters. The target variables to be estimated are: 'GINI coefficient', 'Global Peace Index', 'Democracy Index','Corruption Perceptions Index', and 'UN E-Government Development Index'. Since there are 52 parameters to choose from, we will use the correlation matrix heatmap generated in order to select specific variables, as well as the Variance Inflation Factor as a metric to judge each model for presence of multicollinearity.

Our initial solution approach is to determine the impact of Internet inclusivity on development by subjecting the inclusivity measures and the target variables (the developmental indices) through a Multiple Linear Regression Model, using Ordinary Least Squares to estimate their coefficients. Diagnostic tests to check for accuracy, multicollinearity, and normality of residuals will be conducted. The standardized beta coefficients of the parameters will be analysed to understand the influence of each particular aspect of internet inclusivity on the development of a nation.

A. Dataset

The Inclusive Internet Index contains data about access to internet by different sections of society and it also contains a large number of indicators that help generate an index for the extent of access to the internet in that particular region. The dataset contains 85 columns and 600 entries. The information spans 4 years from the year 2017 to 2021 represented from E1 to E4 respectively. The index contains 57 indicators organised across four categories as well as 24 background indicators.

The four categories are,

Availability: This category refers to the quality and breadth of available infrastructure required for access. Insufficient or unavailable infrastructure can limit Internet connectivity.

Affordability: This category examines the Internet and considers initiatives to lower costs or other ways to promote access. This category includes factors that focus on price.

Relevance: This category considers the value of being connected, in terms of useful services and the availability of locally relevant content. If consumers do not find value in an Internet connection, then adoption of the Internet as a service is less likely. Relevance measures the availability of local content, such as whether basic information or government services are available online in the local language.

Readiness: This category measures the capacity among Internet users to use the Internet to the fullest extent and obtain maximum benefit from it. The category looks at measures such as the level of literacy and educational attainment, the level of privacy regulations and web accessibility.

120 countries are selected - 80 core and 20 non-core. Together, they form a diverse selection of high, low and middle-income countries which represent 96 percent of the world's population.

To ensure diversity within the population surveyed from each country, certain criteria were met. These included quotas for gender, income, community, and age.

Columns such as the Gini Coefficient and the Global Peace Index are indexes which indicate different developmental factors. These may be used to calculate the impact of the Internet on development as a whole.

B. Data Preprocessing

Data pre-processing is done to convert the data into a format that would be suitable for further analysis. The following undesirable characteristics of data were taken into account during the process of data cleaning. Below we address the methods by which these were either corrected or accounted for:

1) Incomplete data: We noticed that a disproportionally large amount of values corresponding to the year 2007 (Edition 1) were missing. The imputation of these values would lead to distorted inferences due to high randomisation, hence all the tuples for Edition 1 were dropped.

Similarly, certain columns contained a higher proportion of missing values. Imputing or aggregating values for these columns would distort inferences, so these columns were also dropped.

2) Noisy data: Outlier analysis involves the identification of anomalous observations within the dataset. We performed outlier analysis by data visualisation using a box-plot. For each attribute, we verified whether the outliers present were within the rationally permissible range of the attribute and format of

the datatype. This held true for all attributes, which signified that the outliers were not present as inconsistent or incorrect data, rather they are accurate deviations from the general trend which are integral in drawing accurate inferences.

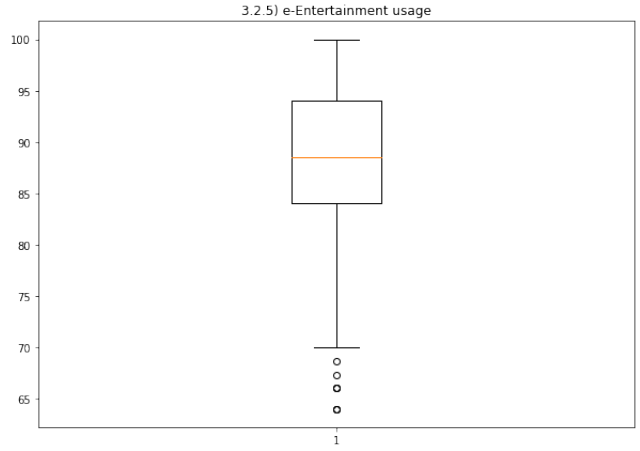

Fig. 1. Box plot with outliers for the 'e-Entertainment usage' feature

3) Irrelevant and Redundant Data: First, we dropped irrelevant columns - the columns that provided background information regarding the nations, but not necessarily about internet inclusivity within the nation itself. We also dropped redundant attributes. Multiple attributes were present in the dataset that described internet inclusivity along the lines of gender (i.e., gender gap in internet access) or the size of offline and online populations within the country.

4) Transformation of Data: First, divide the attributes that counted a subset of the population that satisfied a certain condition (say, Total Male Users within a country) by the total population of the country in order to find the proportions of the attribute. The data is then standardised.

The final cleaned dataset consisted of 57 attributes, with 429 records.

C. Model Building

Four different Multiple Linear Regression Models were built to fit each of the five dependent variables (Gini Coefficient, Global Peace Coefficient, Democracy Index, Corruption Perceptions Index, UN E Government Development Index) : MLR (including all attributes), Lasso Regression, Ridge Regression and ElasticNet Regression. All models were trained and tested on a 80-20 train-test split of the dataset. For Lasso Regression, Grid Search was used to find the optimal values for the hyperparameters (in this case, alpha), and to perform K-fold cross validation. /add part about ridge and elastic net/. Each model was evaluated using metrics such as the Root Mean Squared Error (RMSE), Coefficient of Determination (R-square), adjusted R-square, p-value for significance.

For each target variable, the regression model with the best performance is selected. From an interpretation of the correlation matrix, attributes are selected for each target variable such that they have a high coefficient of correlation. The attributes are used to build another model for that particular variable. This functions as a technique of dimensionality reduction to bring down the total of 53 attributes available to the model after data cleaning. It also enables the easier study of autocorrelation and multi-collinearity among the explanatory variables within the model.

Model	RMSE	Adjusted R-square
Multivariate	5.55	0.13
Lasso	5.26	0.315
Ridge	4.84	0.34
Elastic-Net	4.82	0.344

TABLE I
EVALUATION METRICS FOR GINI COEFFICIENT

Model	RMSE	Adjusted R-square
Multivariate	0.35	0.33
Lasso	0.35	0.41
Ridge	0.278	0.588
Elastic-Net	0.277	0.591

TABLE II
EVALUATION METRICS FOR WORLD PEACE INDEX

IV. MODEL EVALUATION

A. Evaluation metrics

We tabulate below the various metrics being used to evaluate the four regression models on the test dataset:

As can be observed from the data, the best-performing models for each of the tagrets variables are as listed below:

1) Gini coefficient: ElasticNet regression
2) World peace Index: ElasticNet regression
3) Democracy Index: Ridge regression
4) Corruption Perceptions Index: Multivariate regression
5) UN e-Government Development Index: ElasticNet regression

The Gini Coefficient is best estimated by the Elastic-Net and Ridge regression models. The performance of all models however, were generally poor for this indicator, as seen by the high RMSE scores and the low adjusted R-square values. This is consistent with the observation made from the correlation matrix developed during Exploratory Data Analysis, wherein

Model	RMSE	Adjusted R-square
Multivariate	1.35	0.56
Lasso	1.42	0.57
Ridge	1.58	0.676
Elastic-Net	1.199	0.653

TABLE III
EVALUATION METRICS FOR DEMOCRACY INDEX

Model	RMSE	Adjusted R-square
Multivariate	6.27	0.83
Lasso	7.39	0.859
Ridge	43.018	-4.42
Elastic-Net	6.414	0.879

TABLE IV
EVALUATION METRICS FOR CORRUPTION PERCEPTIONS INDEX

Model	RMSE	Adjusted R-square
Multivariate	0.054	0.88
Lasso	0.069	0.897
Ridge	0.062	0.904
Elastic-Net	0.057	0.9212

TABLE V
EVALUATION METRICS FOR UN E-GOVERNMENT DEVELOPMENT INDEX

Model	RMSE	Adjusted R-square
Gini coefficient	6.45	-0.029
Global Peace Index	0.377	0.3259
Democracy Index	4.2e-16	1
Corruption Perceptions Index	0.118	0.7
UN e-Government Development Index	0.068	0.089 height

TABLE VI
EVALUATION METRICS CHOSEN REGRESSION MODEL OF EACH TARGET VARIABLE

we noted that the Gini Coefficient was not strongly correlated with any of the variables.

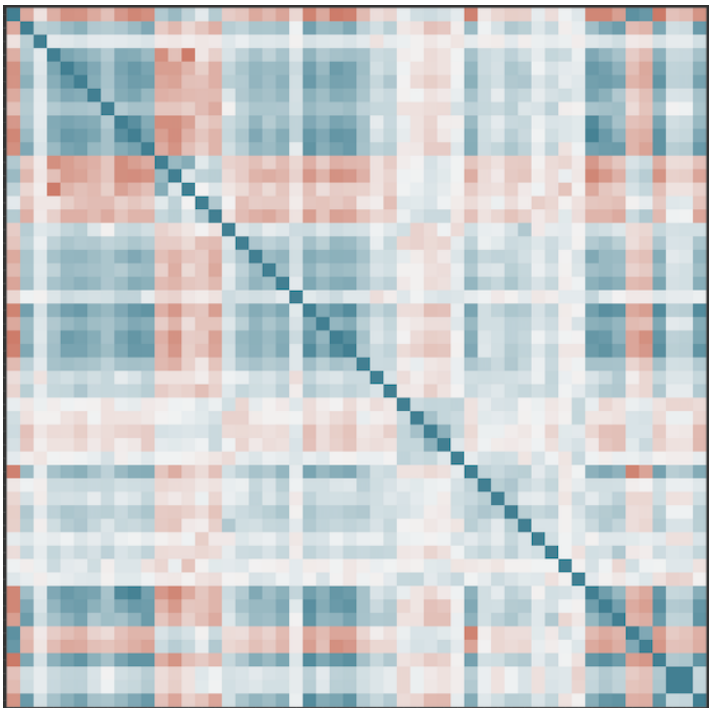

Fig. 2. Heatmap of correlation matrix

The World Peace Index was best estimated by the Ridge and Elastic-Net models (once again). We note that although all of the models yield a fairly low RMSE score, the adjusted R-Squared values are not very high, with the best model performing slightly better than chance. Thus, this particular indicator may possibly be better estimated with a simpler model involving less independent variables.

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.

The Corruption Perceptions Index was best estimated by the standard MLR model, yielding the lowest RMSE score along with a significant Adjusted R-square value (larger than any of the values observed for the previous indicators). We note that the Ridge regression model performs especially poorly here, with an exceptionally high value of RMSE, and a negative Adjusted R-squared value, suggesting that the model performs worse than chance.

The UN E-Government Development Index is best estimated by the Elastic-Net model, yielding exceptional values for both the metrics used. We note that all of the models perform an exceedingly good job in learning this particular indicator, with extremely low RMSE values and very high Adjusted R-squared values.

Overall, the Elastic-Net model performs the best, yielding good values for both metrics (other models often failed in one metric or the other). The Lasso model yields the lowest performances, with it being unable to best predict any of the indicators. Both Ridge and MLR perform decently well, with certain glaring exceptions.

We tabulate below the evaluation metrics for the models built exclusively on the strongly correlated variables corresponding to each target attribute:

From table IV-A, we observe the following:

- In the case of Gini coefficient, Global Peace Index and UN e-Government Development Index, there is an increase in RMSE value and a decrease in adjusted R-square value.
- Note worthily, the Democracy Index achieves a perfect adjusted R-square value of 1 and an infinitesimally small RMSE value in the order of negative sixteenth power of 10.
- In the case of the Corruption Perceptions Index, we make a trade-off between a small decrease in adjusted R-square value for a significant reduction of the RMSE value.

B. Homoscedasticity

We plot a graph of the residual patterns to verify whether the model is homoscedastic. This is tested by plotting residuals as predicted values vs the residuals or error in prediction. The absence of discernible patterns in the residual plot indicates homoscedasticity. We observe homoscedasticity in all 5 of the regression models, as can be seen in IV-B.

V. INFERENCES

A. Gini coefficient

The Gini index is a measure of income inequality. It quantifies the amount by which income earned per household deviates from an income which is equally distributed. The regression model for Gini coefficient has a high RMSE score and a low adjusted R-square value. An evaluation of the correlation matrix revealed that there is only one variable that is strongly correlated with the Gini coefficient - 'BG24) Total fixed line broadband subscribers'. This is attributed to be the reason for the under-performance of this model. We are thus able to conclude that income inequality does not depend on internet inclusivity to a large extent.

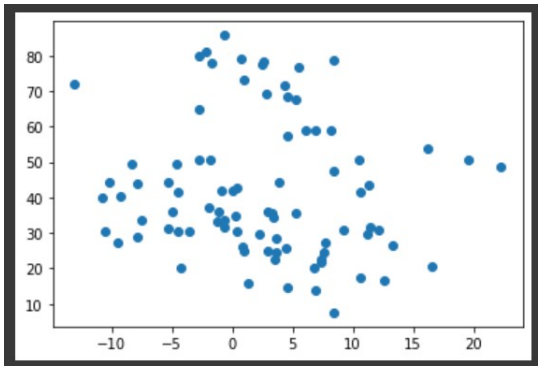

Fig. 3. Residual plot for model built to estimate CPI

B. Global Peace Index

The global peace index is a measure of sustainable peace within a region. A higher value of Global Peace Index indicates a higher level of violence in a country.

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. For example, Male Internet Users has a coefficient of -0.335 which would mean that the higher the number of male users, the lower the index and therefore, the more peaceful the country is. This draws attention to the implication that engagement on the internet, at least among male users, does not leads to disharmonious consequences within society. Following a similar reasoning, 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 a consistent inference.

C. Democracy Index

The democracy index is measure of the quality of democracy and aims to understand the biggest threats against a sustainable democracy.

The Democracy Index is highly impacted by the total number of internet users, the number of male internet users and the number of female internet users. 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 counter-intuitive at the face value, it can be explained by the vast amount of misinformation propagating on social media, both in developed and developing countries. This conclusion is backed by the negative correlation between Democracy Index and trust in social media. However, the data also suggests 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.

We also observe a high impact created by Business Environment Rankings, which implies that healthy market competition and investment leads to a better quality of democracy.

D. Corruption Perceptions Index

The corruption perceptions index is the measure of public-sector corruption perceived by the general populace worldwide. The higher the value of CPI, the more ”clean” the sector is viewed by the public.

The Corruption Perceptions Index is highly impacted by the number of female and male users; having a negative relation with the latter and a positive relation with the former. The Business Environment Rankings also possess a significant coefficient, suggesting a strong positive relationship between CPI and this variable. It also depends on the Electricity Access, the Educational attainment, the Network Coverage, the Gender Gap in Internet Access, the cost of a Smartphone, the Broadband operators' market share, the trust in Government Websites and Apps, as well as the e-Entertainment usage in the country. We note that the CPI is 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. The difference in relationship between male and female users is confusing, and it is unclear as to why this is the case, particularly the strong negative relationship between female users and CPI. Countries that wish to be publicly viewed as “clean” and “not corrupt” may do well to focus on removing the barriers that exist to Internet access, whether they be gender (gaps), electricity or non-lucrative business environments.

E. UN e-Government Index

The UN e-Government index measures the trends in e-governments or the digitisation of governing bodies across the world.

The UN E-Government Development Index is mainly dependent on Trust in Non-Government websites and apps, Wireless Operators Market Share, Open Data Policies, and Technology-Neutrality Policy. It is also dependent on e-Health content and Level of web accessibility to a lesser extent. Almost all variables are positively correlated, with Wireless operators' market share being the only exception. In general, governments may better use the Internet to include its citizens by remaining transparent and following policies that keep the public in the loop. They will also benefit by providing various forms of content on the Internet that may be of use to the public, such as health content, and making sure that the Internet is accessible in the first place. The high performance of all models with this indicator suggests that Internet Inclusivity is a key factor here, and that governments must work on the same to see good results.

CONCLUSION

The novelty of approach stems from the analysis of macro impacts of internet inclusivity on national development, as opposed to micro impacts on the lives of the population. We thus provide a framework that can form the basis of developing centralised policies towards the enhancement of national development indicators that fall under the categories

of social (Global Peace Index), political (Democratic Index) and economic (Gini coefficient) factors.

We see the beginnings of a solution being bridging the gender divide in terms of accessibility of internet as well as education. This is reflected in 4 out of the 3 target indices chosen. Our regression model performs best on the Democracy index with a perfect adjusted R-square.

In final conclusion, internet inclusivity plays a large role in the national development of a country.

ACKNOWLEDGMENTS

We would like to acknowledge our Data Analytics Course Professor Dr. Gowri Srinivasa for providing constant guidance during each phase of our project. We would also like to thank PES University for granting us the opportunity to undertake this project. We would also like to acknowledge our assistant professors who have prepared the course content and also the teaching assistants who have been constantly providing resources to practice the learnt concepts.

REFERENCES

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

[2] INFRASTRUCTURE, INTERNET INCLUSIVENESS AND E-COMMERCE: AN EXPLORATORY STUDY

[3] Weaving the Western Web: explaining differences in Internet connectivity among OECD countries, E. Hargittai, http://www.sscnet.ucla.edu/soc/groups/scr/hargi.pdf

[4] Kearns, A., Whitley, E. Associations of internet access with social integration, wellbeing and physical activity among adults in deprived communities: evidence from a household survey. BMC Public Health 19, 860 (2019).

[5] GoWell Community Survey. $http://www.gowellonline.com/about/components/survey$

[6] Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, Stewart-Brown S.The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health Qual Life Outcomes. 2007;5:63

[7] Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE. International physical activity questionnaire: 12 country reliability and validity. Med Sci Sports Exerc. 2003;35:1381e–1395.

[8] Ellen Johanna Helsper. Digital Inclusion: An Analysis of Social Disadvantage and the Information Society

[9] Sankara Moorthy K, Krishnaraj R, Celina P, Arunkumar C Analysis of internet inclusion among the populous countries. International Journal of Pure and Applied Mathematics 119, 15 (2018)

[10] Towards Digital Inclusion: Barriers to Internet Access for Economically and Socially Excluded Communities

VI. APPENDIX

A. Contributions

Raghav T Kesari (PES1UG19CS363): Compilation and writing of Phase-1 report and building different OLS regression models to optimize our choice of models. Also worked on inferences for one target variable.

Divya Shekar (PES1UG19CS148): Data cleaning, EDA and compilation and writing of Phase-2 report and evaluation of models based on specific performance metrics. Additionally worked on inferences for one target variable.

Krithika Ragothaman (PES1UG19CS231): Data cleaning, EDA and building of all OLS regression models, and inferences for two of the target variables