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 the impact of higher internet access on the overall development of a 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

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 internet has changed from being a mere commodity to an absolute necessity due to the increasing reliance on an interconnected world. It provides a capability so powerful and general that it can be used for almost any purpose that depends on information. Due to the huge impact of internet in a person's daily life and how it is being used frequently on a daily basis for various purposes by most individuals, studying the access to Internet across the world has become an important field of study.

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. This paper provides a survey of different research papers on topics related to the access of internet in different countries across the world with respect to their different methodologies of study and the means of collection of data and many other aspects.

II. Review of Literature

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.

A. The Inclusive Internet Index 2021: Methodology report

[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.

This paper only documents the features related to the internet and the factors affecting it but does not perform any data analytics on the collected data. It provides a detailed explanation on the basis of selection of attributes and their nature and relevance to the inclusivity index. Hence, it acts as a good starting point to perform data analysis by looking at the importance of the features and how they contribute to the internet inclusivity of any region. An elaborate tabulation of each attribute(feature), the category it falls under, the type of variable and necessary metadata is included.

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.

B. The impact of the Internet on economic growth and prosperity

This paper 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. Around two billion people are connected to the Internet, and almost $8 trillion exchange hands each year through e-commerce. the Internet accounts for, on average, 3.4 percent of GDP across the large economies that make up 70 percent of global GDP.

This paper also 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, specially G8-developed and developing countries and between internet maturity and better living standards.

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.

C. Weaving the Western Web - explaining differences in Internet connectivity among OECD countries

Despite the Internet's growing significance, little sociology research has been done to examine its spread. The most significant indicators of a country's internet connectivity are economic prosperity and telecommunications policy.

A major technological advancement of the 20th century with massive political, social, and economic implications is the Internet. In terms of society, the new medium 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 Internet's broad effects are undeniable, but research into how it spread, particularly internationally, has received little attention.

The internet makes numerous forms of connection - regardless of location proximity - quicker than ever before, which notably aids in the convergence of space and time. Understanding which countries will be able to advance these industries the most requires examining the growth of the Internet, as knowledge-intensive activities are becoming a more significant part of Economies. The network's global dissemination has increased at a rate that exceeds 50% annually. In the United States, usage of the system increased by more than fourfold between 1994 and 1998. The OECD is the perfect example that led to variations in international Internet access among nations with fairly similar socio-economic development.

Monopoly in the telecom sector has restricted competition and kept Internet access charges high. 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 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.

D. The Internet and Socio-economic development: Exploring the interaction

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.

Governments and international organizations are aware of the value of increasing developing nations' access to the Internet. These results imply that increased connectivity among developing nations would enhance the information infrastructure as a whole. We believe that development must prioritize reducing poverty and creating a social climate that supports universal access to basic welfare systems.

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. Over the past few decades, there has been a lot of discussion over the definition, justification, and conduct of development.

Less focus should be placed on how many people are connected to the Internet than on how accessible it is and how it advances society. Governments in developing nations will need to put extra effort into addressing these problems by enacting laws that improve the poor's ability to access information online.

E. Infrastructure, Internet inclusiveness and e-commerce: An exploratory study

This paper 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. We assume that there is a relationship between these variables such that better Internet infrastructure should facilitate E-Commerce revenue growth. One of the bases of this presumption is shown in Figure which depicts 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.

The Trust and Safety is one of the sub-component of the Index that is not at all correlated with E-Commerce revenue is especially not worthy. This variable aims to measure Internet safety and cultural acceptance of the Internet based upon a variety of factors, including social trust in websites, level of

privacy regulation, social media, and a measure of E-commerce safety.

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.

F. Enabling digital development – how the internet promotes development

This paper 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. Supply-side scale economies, where costs drop with an increase in number of transactions, favor the emergence of natural monopolies. Water and electric utilities operate in similar environments. Because of entry costs are high, such sector tends to be regulated. Many internet-based markets—such as mobile payments, web searches, or online bookstores—are dominated by a few firms. Initially, entry costs are low but such websites can scale up extremely quick, even with few resources.

The massive data collected by internet platforms have created a new branch of economics— nano-economics- which studies individual and computer-mediated transactions. The benefit to the user is that services can be tailored to individual preferences and needs-but at the cost of giving up their privacy. For the seller, it allows more targeted advertisement and even price discrimination, when automated systems can analyze user behavior to understand the willingness to pay and offer different prices to different users.

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. 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.

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

IV. Problem Statement and Approach

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.

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.

Our initial solution approach is to determine the impact of Internet inclusivity on development by subjecting the inclusivity measures and the target variables through a Multiple Linear Regression Model by using Ordinary Least Squares to estimate their coefficients.

Multiple Linear Regression models will be run to estimate these target variables. Diagnostic tests to check for accuracy, multicollinearity, and normality of residuals will be conducted. The standardized beta coefficients of the parameters could be analyzed to understand the influence of each attribute of internet inclusivity on the development of a nation.

V. Exploratory Data Analysis

A. Data cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. This prevents inaccurate inferences being drawn from the model owing to the presence of inaccurate data.

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 perfomed outlier analysis by data visualisation using a box-plot. The data points present outside the minimum and maximum range of the box-plot are considered to be outliers.

B. Data Pre-processing

Data pre-processing refers to the process of converting data into an understandable format. It involves manipulation or dropping of data to ensure or enhance performance. Irrelevant and Redundant Data were all dropped from the dataset. We also dropped redundant attributes. There were multiple attributes in the dataset that described internet inclusivity along the lines of gender and were also repeated with very little variations.

C. Correlation

We used the corr() function associated with Pandas dataframes to generate a correlation matrix, which is a table used to show the coefficients of correlation between variables. We noted certain strong positively and negatively correlated variables from this graph. This information will be used during the model-building phase in order to avoid multi-collinearity in our models.

D. Data visualization

Data visualization is the process of generating graphical representation of the data. It provides insights as to the shape of distribution of the data. We use different visualization techniques depending on the type of each attribute.

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 tendency to skew.

ACKNOWLEDGMENT

We would like to acknowledge our Data Analytics Course Professor Dr. Prajwala for providing 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.

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