CISC 499 RESEARCH PROPOSAL

Project Name: How are you served (by the Internet)?

Group member: [Yuehan Qi] [Alex Wang] [Peiyang Huang]

Student No. [20167259] [20144392] [20171385]

[2022/11/10]

Background

Since the Internet became popular throughout Canada in the 1990s, it has changed the way Canadians learn and work, buy products and services, communicate and consume entertainment. The Internet makes information more accessible than ever before. However, it does not benefit all Canadians equally, and certain risks accompany its advantages. For example, Statistics Canada found in 2018 that 1.2% of families with children do not have the Internet. Among the families with the lowest income, this figure increased to 4.2%. In addition, private Internet service providers avoid building infrastructure in rural and remote areas of Canada. Therefore, visits to these areas may be more expensive and slower than in urban areas. Therefore, the background of our project is based on the data released by National Broadband Data for analysis. Just as Rajabiun and Adiandari did, (Rajabiun, 2020) investigated how the emergency of broadband Internet connectivity as an essential utility has influenced the development of public policies that aim to promote universal access to essential communications services in Canada. (Adiandari et al., 2020) Aim to analyze and process proposals for improvement in service quality using internet banking services based on the Quality Function Deployment (QFD) method through the House of Quality (HoQ) preparation.

Objective

By analyzing the national broadband data in Canada, we will find out the potential social factors that affect the Internet coverage in different regions such as the household income, education levels of people living in the area and its relationships with the economic growth. We will also analyze the speed distribution of broadband based on geographic distribution. Then by stating a specific scenario, we intend to provide advice on broadband investment to achieve efficiency for the government.

In addition, we also apply multiple machine learning models to different datasets. For example, for closer analysis, unsupervised learning K-means Clustering is used to automatically classify them into the appropriate k-clusters with variables such as age, income, education, and geography. For supervised learning, multiple linear regression can find complex relationships between different variables and Internet speed/Internet usage and predict trends that may be able to be improved in the future.Eventually, we will compare the performance of all the different ML models and analyze them in a more accurate scheme.

Moreover, we will present data visualization to illustrate the network/internet speed/internet usage penetration in different provinces of Canada. In addition, the data visualization also serves as an essential tool to show the availability of broadband/cellular services in urban/rural areas and the relative level of LTE coverage in major population centers/significant roadways.

Then visualize the data and make it possible for the user to make the decision. Like Chawla et al., 2020, it is to present a generalized collaborative framework between different production facilities located at different geographic locations to realize an energy-saving and effective cyber-physical production system for the production of different types of jobs in the context of industry 4.0 and beyond. We will start from various angles, such as network speed, family income, infrastructure improvement or government finance.

Technically, we may use stratified random sampling or non-sampling stratified (example)through the differentiation of household income, such as the rich in rural areas, the poor in rural areas, the rich in urban areas and the ordinary in urban areas. Based on the test of the incremental effect of Internet speed, usability data is further divided into four categories, such as superfast and fast. The unemployment rate may also be introduced. The fixed effect estimator might be used to establish a primary model, detect whether the increase of broadband availability is related to the change of county employment rate and income, and then conduct visualization processing of python map cloud ANA.

Timeline

Dec 6–Jan 1	Group members doing individual research on the data and project. Then meet up to exchange ideas and tools used for the project.
Jan 1–Jan 31	Start to have meetings with professor Chen. Update the proposal according to plan changes. Explore and clean the data.
Feb 1–March 1	Perform preprocessing on the raw data and extracting features from it. Get preliminary results.
March 1– March 15	Visualize the data and record the findings in the report along the way.
March 15 – April 1	Build on a website to illustrate the findings and finalize the report.

Reference

Adiandari, A., Winata, H., Fitriandari, M & Hariguna, T. (2020). Improving the quality of Internet banking services: An implementation of the quality function deployment (QFD) concept.Management Science Letters , 10(5), 1121-1128.

Chawla, V., Angra, S., Suri, S & Kalra, R. (2020). A synergic framework for cyber-physical production systems in the context of Industry 4.0 and beyond.International Journal of Data and Network Science, 4(2), 237-244.

Reza Rajabiun, Technological change, civic engagement and policy legitimization: Perspectives from the rise of broadband Internet as an essential utility in Canada, Government Information Quarterly,Volume 37, Issue 1,2020,101403,ISSN 0740-624X,