Classification of Average Internet Access Costs Around the World

Aretha Kassegnin, Dr. Chantal Larose
Department of Mathematical Sciences
Eastern Connecticut State University, Willimantic, Connecticut, USA

INTRODUCTION

Can we determine how much the average person pays for internet access in other countries? In these times of COVID most people experience work or school at home, which means that we need to be connected to the internet. What financial hardship does that place on people globally? This project looks at different factors that may influence the average cost of 1 GB in across many countries worldwide. Classification and regression tree (CART) models classify countries' interest access costs with 92% accuracy, an increase in accuracy of 64% over baseline*

Now we want to find the accuracy of average cost of 1GB in 2021 with the same predictors.

OBJECTIVE

Internet prices differ around the world. India has free internet for all their citizens. How is that possible and does cost a lot to have internet there?

What can you do with 1GB?

- •Websites – 600 websites
- •Email /Texting – 350,000
- •Steaming HD -30 mins
- •Steaming music – 200 songs
- •Facetime – 4 hours
- •App download/update – 25 apps

According to CNET.com in 2022 the average internet plan has increased to 400 GB per month and that is after the COVID pandemic!

Data from USMobile .com (n = 243, complete case n = 206) shows the average U.S. internet price is $8 /GB.

Under $1	$1 to $4	$5 to $15	Over $15
17.156863	55.882353	21.568627	5.392157

Table 1: Target Variable Percentages

I made the target variable: average price of 1 GB in 2021 and the predictor variables are...

- number of internet plans per country
- most expensive 1GB
- internet population users per country
- cheapest 1GB

METHOD

Data Cleaning and Preparation. Removing incomplete records, binning the data, evaluating outliers, boxplots and a five-number summary using the variables z-scores, which shows the number of standard deviations above or below the mean that datapoints lie in the dataset to see if there would be any issues such as skewness.

Exploratory Data Analysis (EDA). Compare average of 1GB in 2021 to each predictor variable: number of internet plans per country,. One predation that a larger population would have a higher average of 1GB. Overlayed and normalized histogram (see figure 1) to better see the relationship between the target and predictors variables.

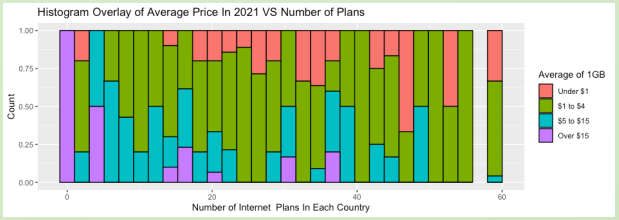
Data Partition and Validation. Now it was time to prepare the data to be used for the models. This means using the original dataset to create a testing and training dataset. Partitioned the data with 75% for training and 25% for testing dataset.

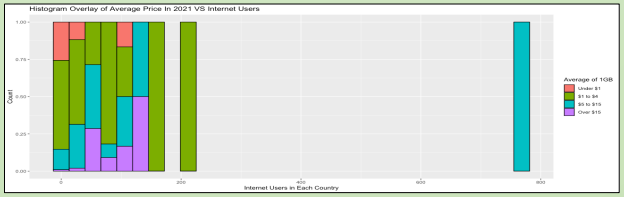
Classification and Regression Trees. The model being used is a CART model which is one of many decision trees, which consists of a set of decision nodes, connected by branches, extending downward from the root node until terminating in leaf nodes. To establish a baseline for this data I concluded that about 55% of the average price of 1 GB between $1-$4, shown in Table 1. We want our model to have a higher accuracy then the baseline.

The CART model determines branches by splitting the data into smaller sections based on different values and combinations of predictors. The CART algorithm selects the best splits and repeats this process until the optimal tree is found. The formula is called Gini criterion which is shown below

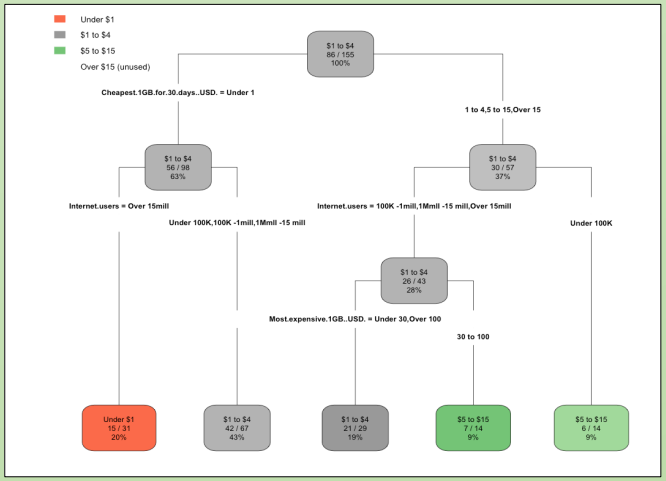
$$\Phi(s|t) = 2P_L P_R \sum_{j=1}^{\#\text{classes}} \left| P(j|t_L) - P(j|t_R) \right|$$

RESULTS





CART Model Using Training Data



Training CART Model Training Outcome

	Under $1	$1 to $4	$5 to $15	Over $15
Under $1	15	12	0	0
$1 to $4	14	63	9	0
$5 to $15	2	18	13	0
Over $15	0	3	6	0

CONCLUSION

Note: This project is a work in progress.

From EDA we can tell

- **Average of 1GB VS Number of Internet shows that as the number of plans increases, the average decreases.**
- Average of 1GB VS Cheapest 1GB shows us that there is a range between cheap and average with shows that counties have "variation" in how they charge.
- Lasty, Average of 1GB VS Internet Users shows that the more people that use the internet the average decreases.

From CART Models

- **Accuracy of 59% - a 4% increase from the baseline model.**
- The error rate for the CART training model is 58%
- The CART training model had an accuracy of 57% which is close to the training model.

NEXT STEPS

- Doing more model evaluation of this model such as
  - Sensitivity
  - Specificity
  - Precision
- Trying other models to see if there is an increase in models' accuracy compared to the baseline
- Factors pre- pandemic and post- pandemic has changes drastically and using those as predictor to better represent the world
- **Could be a bias dataset and finding more datasets to compare to this.**

REFERENCES

1. https://www.cable.co.uk/broadband/pricing/worldwide-comparison/#highlights
2. https://www.usmobile.com/blog/how-much-is-a-gigabyte/
3. Data Science Using Python and R by Daniel T. Larose and Chantal D. Larose
4. Data Mining and Predictive Analytics • Daniel T. Larose and Chantal D. Larose