2207 – 18
Uganda

WHITE PAPER

MACHINE LEARNING (ML) DRIVEN APPLICATION (APP) THAT PREDICTS THE UPTAKE OF FIBER OPTIC CONNECTIONS -UGANDA

TEAM 18

WHITE PAPER ON A MACHINE LEARNING (ML) DRIVEN APPLICATION (APP) THAT PREDICTS THE UPTAKE OF FIBER OPTIC CONNECTIONS

1. Introduction:

This white paper presents a Machine Learning (ML) driven application (app) that predicts the uptake of fiber optic connections in Uganda. In the world of Broadband Networks, fiber optic technologies offer significant advantages over DSL and WiMAX in terms of faster speeds, higher bandwidth, more reliability, better signal quality, longer distances, and more security. The app provides accurate predictions that can help telecommunication companies plan their resources and infrastructure accordingly, thereby overcoming challenges such as high deployment costs and limited infrastructure.

2. Problem Statement:

Despite the increasing demand for high-speed internet connectivity in Uganda, the deployment of fiber optic networks has been slow due to various challenges. The high cost of deploying fiber optic networks is a significant challenge for telecommunication companies. In addition, there are challenges related to limited infrastructure, limited funding, lack of access to accurate, updated demographic and consumption data, and a lack of a skilled workforce. As a result, telecommunication companies have difficulty prioritizing expanding their fiber network and meeting the growing demand for high-speed connectivity. The ML-driven app aims to address these challenges by providing accurate predictions that help telecommunication companies make informed decisions about where to expand their fiber networks.

3. Target Personas

Name, Age, Job Title	Background	Goals	Pain Points
John, 45, CEO	John has over 20 years of experience in the industry and is responsible for the overall growth and success of the company. John is interested in the ML-driven app as it can help his company expand its fiber network and meet the growing demand for high-speed connectivity.	• To increase the revenue of his company by expanding its fiber network. • Accurate predictions that can inform his company's decisions about where to expand its fiber network.	• The high deployment costs associated with fiber optic network deployment. • Political instability and limited funding in some regions.
Jane, 32, Network Engineer	Jane has five years of experience in the industry and is responsible for the design, implementation, and maintenance of the company's network infrastructure. Jane is interested in the ML-driven app as it can help her company plan its resources and	• to design and implement an efficient and reliable fiber network for her company. • accurate predictions that can inform her company's decisions about where to expand its fiber network.	• the technical expertise required for fiber optic network deployment. • construction disruptions and environmental factors such as natural disasters that can impact

Kamogelo Mogaba, David Mugambi, Ikaneng Jack, Omolayo Ipinsanmi, Atunima Emmanuel

Name, Age, Job Title	Background	Goals	Pain Points
	infrastructure more effectively.		network reliability.
Sarah, 28, Sales Representative	She has three years of experience in the industry and is responsible for selling the company's products and services to customers. Sarah is interested in the ML-driven app as it can help her company identify potential customers and expand its customer base.	• to increase the number of customers using her company's fiber network. • She is looking for accurate predictions that can inform her company's decisions about where to target its sales efforts.	• limited internet penetration • high costs associated with fiber optic network deployment.

4. Solution:

Our ML-driven app is designed to help telecommunication companies in Uganda overcome the challenges related to the slow uptake of fiber optic connectivity. By leveraging historical demographic data and using predictive analysis techniques, the app can provide accurate predictions about the uptake of fiber optic connections in different regions of Uganda. This information can help companies make informed decisions about where to expand their fiber network, plan their resources and infrastructure more effectively, and reduce deployment costs.

With the help of our app, telecommunication companies in Uganda can prioritize their expansion efforts, target areas with high demand, and better serve their customers. By increasing the uptake of fiber optic connectivity, the app can contribute to the economic growth of Uganda, improve access to information and communication, and enhance the quality of life for its citizens.

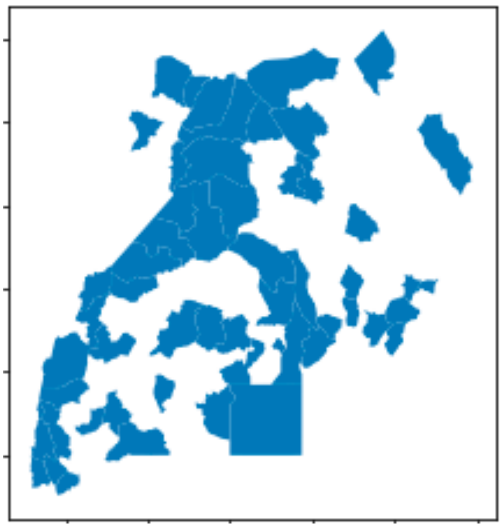
5. Key Features:

Feature	Description
Predictive Analysis	The app uses predictive analysis techniques to forecast the uptake of fiber optic connections in different regions of Uganda. The app considers various factors such as population density, economic growth, and competition among service providers to provide accurate predictions.
Interactive Dashboard	The app provides an interactive dashboard that allows telecommunication companies to explore the predicted uptake of fiber optic connections in different regions of Uganda. The dashboard provides real-time data visualization and allows users to interact with the data and adjust parameters to simulate different scenarios.
Data Visualization	The app uses data visualization techniques to present the predicted uptake of fiber optic connections in an easy-to-understand format. The data is presented in charts, graphs, and maps, making it easy for telecommunication companies to interpret and use in their decision-making processes.

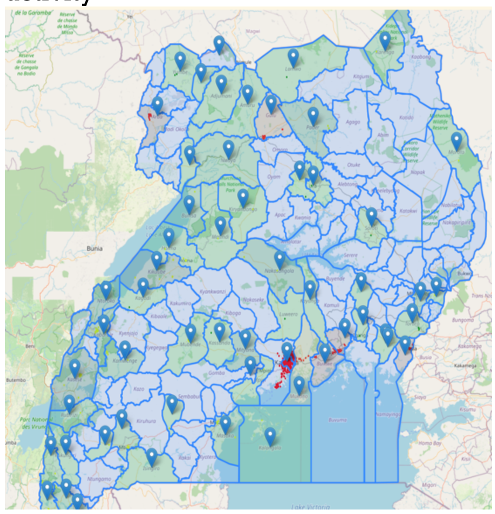
Customization	The app allows users to customize their predictions based on their specific needs. Users can adjust parameters such as population density, economic growth, and competition among service providers to tailor the predictions to their unique circumstances.
Integration	The app integrates with existing telecommunication infrastructure and management systems, making it easy for telecommunication companies to incorporate the predictions into their existing workflows.

6. Data and Methodology

Districts estimated to have speed test activity



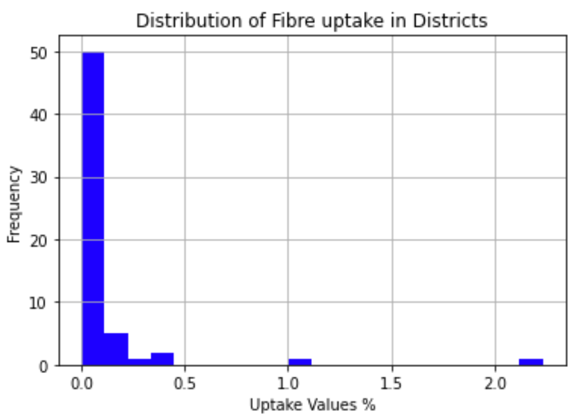
Districts with various levels of speed test activity



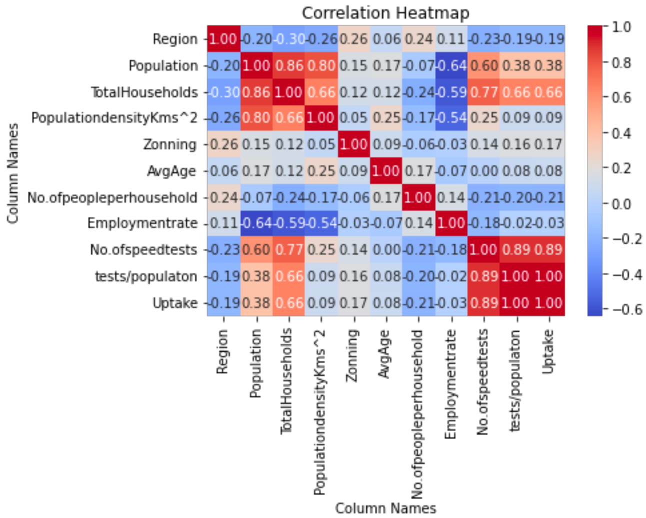
Through the use of available data, our team was able to predict the uptake of fiber optic connectivity in Uganda. We utilized Ookla test data as a proxy for uptake and measured it against demographic data. Our analysis of Ookla data revealed that a significant number of districts in Uganda lack access to fiber optic connectivity, as depicted in the left image above. Furthermore, among the districts that have fiber optic access, only highly urban areas such as Kampala, Jinja, and Gulu, indicated by the red regions in the left image above, have a relatively large population that uses fiber optic broadband.

Density of tests correlated with potential fiber optic uptake.
It sounds like you're discussing a correlation between the density of tests being carried out and the potential for fiber optic uptake in different districts. Based on the data collected, it seems that only one district has a density of tests above 2.2%, while most districts have little or no activity.

Overall, more research would be needed to fully understand the relationship between test density and fiber optic uptake, and to determine whether this correlation holds up in other regions or countries.



Further, the Correlation heatmap shows that indicators such as region (proximity to the capital), population size, total households, Zoning, and number of people in the households.



Data Extraction:
The extraction of Ookla data was performed using the geopandas library within a Jupyter Notebook hosted on Kaggle's IDE. The code snippet provided a streamlined process for reading the Ookla data file, which was in the shapefile format (.shp). Geopandas, a powerful library for working with geospatial data, was utilized to load the Ookla data into a geopandas GeoDataFrame. This data structure facilitated efficient manipulation and analysis of the spatial data, ensuring ease of use and accuracy.

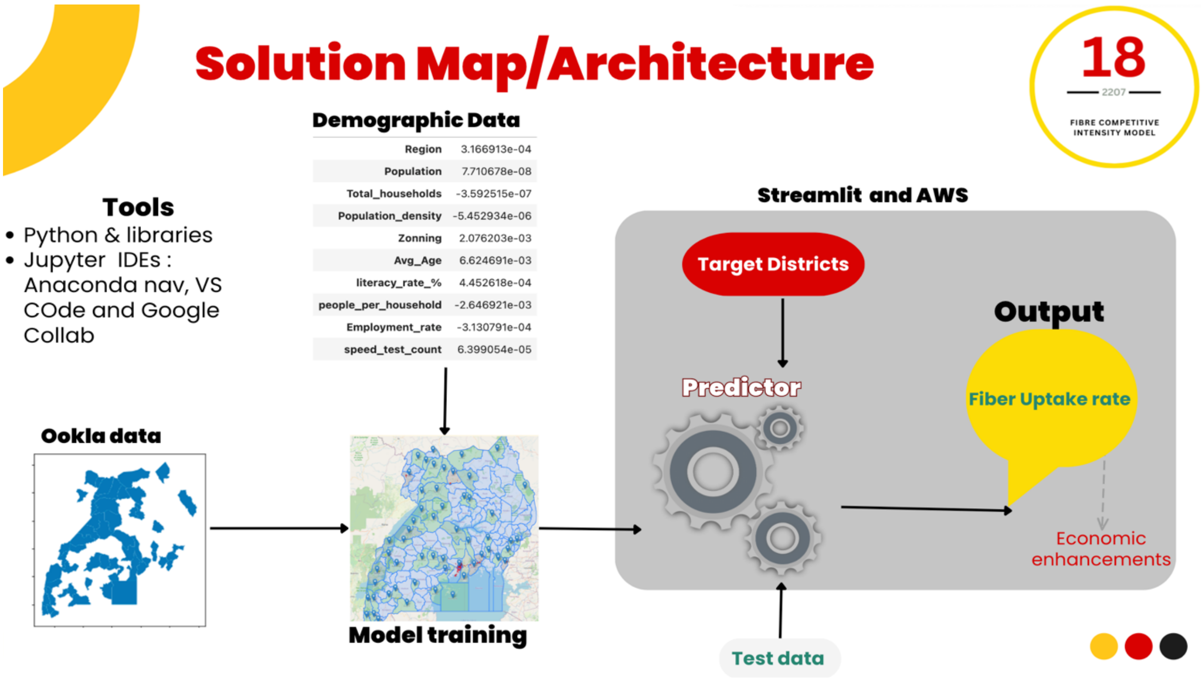
Calculation of Speed Test Proportions and Fiber Infrastructure:
To assess the adoption of fiber optic connectivity, the proportion of speed tests per household (HH) was calculated for each district. The analysis incorporated information on the total number of households with fixed broadband, which amounted to 70,000. Out of these households, 26,000 were equipped with FFTX (Fiber to the X) infrastructure.

Estimating the population with access to fiber involved using upload speeds as a proxy, assuming that 55% of the population in Uganda had internet access. This estimation enabled the identification of approximately 1,000,000 households with fiber infrastructure. To gauge the national uptake rate, the proportion of households with fiber connectivity relative to the total number of households was calculated, resulting in a rate of 2.4%.

Unit of Analysis: District Level
The unit of analysis in this study was the district. By focusing the data and analysis at the district level, a localized understanding of fiber optic uptake was achieved. This approach allowed for the examination of variations and patterns within each district, unveiling valuable insights into the distribution and penetration of fiber infrastructure across different geographic areas.

By employing these data extraction techniques, calculations, and district-level analysis, the ML-driven app provides telecommunication companies with comprehensive and actionable information. The code snippet utilizing geopandas ensures a smooth data processing pipeline, while the calculated proportions and uptake rates offer a clear understanding of the current state of fiber optic connectivity in Uganda at a granular level. This robust methodology empowers companies to make informed decisions, allocate resources effectively, and drive the expansion of high-speed connectivity in the country.

7. Application



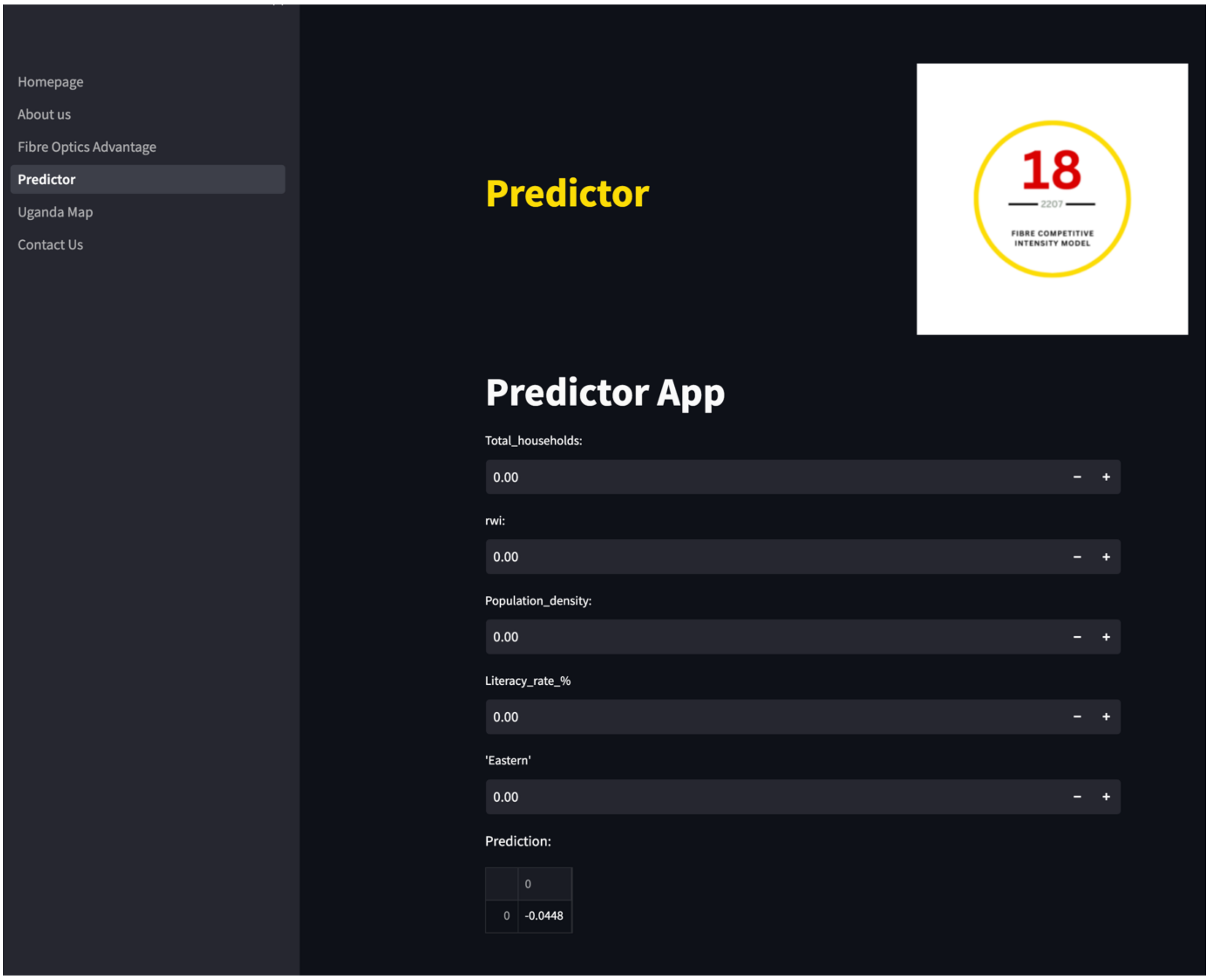
The back end of the application was developed using Python and its associated libraries. The application seamlessly integrates data from Ookla, which serves as the source for the dependent variable (uptake rate), with relevant demographic data, serving as the independent variables.

During the algorithm development phase, careful consideration was given to selecting the most influential variables. The following key variables were identified: "GDP_per_capita," "relative wealth index," "people_per_household," "Total_households," "literacy_rate_%," "poverty_index," and "Zonning." These variables were chosen based on thorough analysis of model coefficients derived from linear regression, ridge, LASSO, regression tree, and random forest models. Through this comparative analysis, these features were determined to have the highest impact on the dependent variable.

To optimize model performance, the selected variables were used to refine the LASSO and random forest models, as they demonstrated the most favorable Mean Squared Error (MSE) and R-squared indicators. This deliberate feature selection process helped enhance the accuracy and predictive capabilities of the models.

The application itself was developed using Streamlit, a powerful framework for building interactive web applications. The predictor within the application utilizes a random forest machine learning model, which consistently achieved the lowest MSE and demonstrated the best R-squared performance. This choice of model ensures reliable and accurate predictions of the uptake rate in the context of fiber optic connectivity.

By leveraging the strengths of Python libraries, carefully selecting influential variables, and implementing advanced machine learning algorithms, the ML-driven application achieves optimal predictive performance. The integration of Streamlit as the application framework and the utilization of the random forest model provide users with a powerful tool for informed decision-making and resource allocation in expanding fiber optic connectivity.



8. <u>Future Improvements</u>

a) Usability Improvements:

1. **Enhanced user interface (UI) design:** Improving the visual design, layout, and navigation of the app to make it more intuitive and user-friendly. For example Unifying the predictor and the map visualization.
2. **Customization options:** Allowing users to adjust parameters and scenarios within the app to tailor the predictions to their specific needs and preferences.
3. **Incorporating user feedback:** Collecting feedback from users and incorporating their suggestions for improving the app's usability.
4. **Performance optimization:** Optimizing the app's speed, scalability, and resource utilization to ensure a smooth user experience.

b) Usefulness Improvements:

1. **Enhanced data accuracy and coverage**: Improving the accuracy and coverage of demographic and consumption data used for predictions to enhance the reliability of the app.
2. **Incorporating real-time data:** Integrating real-time data sources to provide more accurate and timely predictions.
3. **Expansion to make the data more granular:** Expanding the app's scope to cover the ward or neighborhood as the unit of analysis allowing telecommunication companies in different areas to benefit from accurate predictions.
4. **Integration with more systems:** Expanding the app's integration capabilities with various telecommunication infrastructure and management systems for seamless incorporation into existing workflows.

Lessons learnt

a) **Understanding the challenges:** The team gained a deep understanding of the challenges faced by telecommunication companies in Uganda regarding the slow uptake of fiber optic connectivity. This knowledge helps them identify the pain points and address them effectively.
b) **Domain knowledge:** The team developed a strong understanding of the broadband network industry, specifically fiber optic technologies, their advantages over other options, and their importance in meeting the growing demand for high-speed internet connectivity.
c) **Problem-solving skills:** The team acquired problem-solving skills related to deploying fiber optic networks, such as overcoming high deployment costs, limited infrastructure, limited funding, lack of accurate data, and a shortage of skilled workforce. They explored solutions to tackle these challenges and provide value to telecommunication companies.
d) **Data analysis and predictive modeling:** The team gained expertise in analyzing historical demographic data and using predictive analysis techniques to generate accurate predictions of fiber optic uptake. They learned how to leverage data to make informed decisions and provide valuable insights for telecommunication companies.
e) **Application development:** The team acquired skills in developing an ML-driven application that incorporates predictive analysis, interactive dashboards, data visualization, customization options, and integration with existing systems. They learned how to design and develop a user-friendly and valuable tool for telecommunication companies.

9. Conclusion:

Our ML-driven app is a powerful tool for telecommunication companies in Uganda to overcome the challenges associated with fiber optic network deployment. By providing accurate predictions of the uptake of fiber optic connections, the app helps companies plan their resources and infrastructure more effectively, reduce deployment costs, and increase the uptake of fiber optic connectivity in Uganda. The app's predictive analysis, interactive dashboard, data visualization, customization, and integration features make it a valuable asset for any telecommunication company looking to expand its fiber network and meet the growing demand for high-speed connectivity in Uganda.