Implementation of the Ranowash Model

Graphical user interface

Description generated automatically

**Definição do Problema:** To accomplish this goal, the program will develop a systematic partnership with the government, water and sanitation institutions, communities, private sector actors, civil society organizations, and beneficiaries in order to implement astrategic set of mutually supporting activities that contribute to three (3) components/strategic objectives.

We leverage a statistical analysis technique that uses machine learning models to create a profile of benefited families, identifying which families need the benefit and assistance.

The types of insights we can provide you:

•Machine Learning Model: Characteristic of family composition, number of beneficiaries who received assistance on water. The Monitoring, Evaluation, Accountability, and Learning (MEAL) Plan is a performance management tool for planning, managing and documenting how performance data are collected and used and, more broadly, tracking progress over the course of implementation

•Profile of beneficiary families: Characteristic of family composition, number of beneficiaries who received assistance on water.

•Geolocation: Terrorist identification of beneficiary families and groups with similar characteristics to optimize benefit distribution.

•Access to drinking water: Characteristics of families with access to clean water and family group definition.

**Database: The**  data used comes from the commcare system in CSV and PAI format. The information from questionnaires was grouped into text files for manipulation.

**Data Analysis:** The data will be analyzed in powerbi for the generation of graphs and information to characteristicize the strategic problems that should be solved.

**Data preprocessing:** Identification of missing data and correction by column median, when the number of missing values is more than 30% of the total column records. If the total missing values are less than 5% of the column, it is suggested that the data be deleted. Correction of outliers by the median of the column.

**Model Construction:**  The model will be built on unsupervised learning using the cluster method to identify groups and their characteristics.

**Model Deploy:**  The model can be used in powerbi data visualization tools or Flask web application tools.

**Metric You**  want: Accuracy greater than 80%.

**Necessary tools and libraries:** Python, Sklearn, Pycaret, Jupyter notebook, powerbi, numpy, pandas.

**Repository:** https://colab.research.google.com/drive/1pxwDscmGYlVv629Buwg1hyOChgz\_x3\_9?usp=sharing

Creation of the Model

1. Installation of non-native libraries:

Text, Letter

Description generated automatically

1. Loading libraries:

Text

Description generated automatically

1. Access to the cloud environment of the data repositories and selection of columns for the model:

Graphical user interface, Application

Description generated automatically

1. Dataframe inspection:

Cell phone screen with black text on white background

Description generated automatically

1. Import of libraries for Kmeans generation with Pycarert:

Graphical user interface, Text, Application

Description generated automatically

1. Cluster models that best fit the data model offered by pycaret:

Graphical user interface, Table

Description automatically generated with average trust

1. Definition of the number of clusters and result of accuracy, with measurement of accuracy the silhouette score: 81%.

Graphical user interface, Text, Application

Description generated automatically

1. Dataframe view with cluster IDENTIFICATION:

Computer screen with black text on white background

Description generated automatically

1. Distribution of clusters:

Graph, Bar Chart

Description generated automatically

1. Accuracy of clusters:

Graph, Funnel Graph

Description generated automatically

1. Model export

Graphical user interface, Text, Application

Description generated automatically