☆ Star   ⊙ Unwatch ▾

<> Code   ⊙ Issues   ⑂ Pull requests   ▷ Actions   ▥ Projects   📖 Wiki   ⊘ Security   ⬠ Insights

⑂ master ▾   ···

lddrasler final commits   ···   22 minutes ago   🕐 10

View code

☰ README.md   ✎

Tanzania Water Pumps

1. Predicting Functional Status of Water Pumps in Tanzania

In this project, I will apply Machine Learning predictor models to train the classifiers on the known labels and make predictions on the functional status of water pumps with unknown status.

2. Business Problem

The majority access to clean water in Tanzania, it is through Water Pumps. According to 2015 Tanzania Water Point Mapping Data, about 29% of these water points are non-functional.

The Goal of these predictions is to decreases the overall cost for the Tanzanian Ministry of Water.

3. Data

The data was collected from Driven Data and it is available here:

https://www.drivendata.org/competitions/7/pump-it-up-data-mining-the-water-table/

The data has about 59 thousand values and 39 variables.These data were analyzed based on parameters including:

- Construction Year;
- Funder;
- Installer;
- Etc.

4. Methods and Models

I decided to use multiple Machine Learning models to compare the results, which inclued:

- DecisionTreeClassifier
- RandomForestClassifier
- BaggingClassifier
- DecisionTreeClassifier
- AdaBoost
- GradientBoost
- Pipeline
- XGboost

I have also applied the method such as:

- Confusion Matrix
- ROC CURVE;
- Parameter Grid

5. Final Results

After carefully analyzing the results, I decided to use XGboost as my predictor

We got a final technical result as 0.79. For future predictions, we assume that 79% will be accurate on functional Water Pumps status.

6. Future Work

To have a better predictor, we possible could have more accurate data with less unknown and NaN values

We could have a better data collection of the construction year, to undarstand the water pump lifetime

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%