Water Pump Functionality Prediction

A machine learning project that predicts the operational status of water pumps in rural Tanzania using metadata. The goal is to support proactive maintenance and ensure sustainable water access.

📊 Project Overview

This project uses the Tanzania Water Pump dataset from DrivenData. It applies data preprocessing, feature engineering, and classification models to predict whether a pump is:

- Functional
- Functional but needs repair
- Non-functional

Achieved **81% accuracy** using a Random Forest Classifier.

🧰 Tech Stack / Tools

- Python (Pandas, scikit-learn, matplotlib, seaborn)
- Jupyter Notebook
- XGBoost / RandomForest
- SMOTE (imbalanced-learn)

🚀 How to Run the Project

1. Clone the repo:

```
git clone https://github.com/yourusername/water-pump-predictor.git
cd water-pump-predictor
```

2. Install dependencies:

```
pip install -r requirements.txt
```

3. Run the notebook:

```
jupyter notebook notebooks/Water_Pump_Model.ipynb
```

4. (Optional) Predict on new data:

```
model.predict(new_data)
```

🧠 Project Structure

```
data/
    └── Training_Set_Values.csv
    └── Training_Set_Labels.csv
    └── Test_Set_Values.csv

notebooks/
    └── Water_Pump_Model.ipynb

README.md
requirements.txt
```

📈 Results

- Accuracy: 81%
- Confusion Matrix:
  - Functional: Precision 0.81, Recall 0.89
  - Needs Repair: Precision 0.55, Recall 0.33
  - Non-functional: Precision 0.84, Recall 0.78

Visualizations and model performance plots available in the notebook.

📢 Business Impact

- Enables data-driven pump maintenance in rural communities
- Reduces downtime and repair costs
- Improves infrastructure planning and public service delivery

🔮 Future Work

- Improve "Needs Repair" classification using better features
- Deploy model as a REST API
- Integrate with GIS dashboards

🤝 Contributing

Pull requests are welcome. For major changes, please open an issue first to discuss what you would like to change.

To contribute:

- Fork this repo
- Create a feature branch
- Commit your changes
- Open a pull request

License

MIT License

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%