

Predictive maintenance for water pumps in Tanzania

Martijn Beeks

Part of the application
process for the role
Machine Learning
Engineer at Vantage AI

12-05-2022

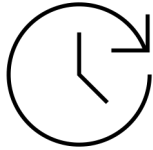
Today

Presentation of assessment

1. Introduction
2. Exploratory data analysis
3. Model training and deployment strategy
4. Modeling results
5. Recommendations

Introduction

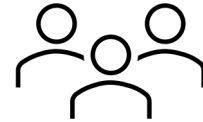
“Develop a reproducible model that can predict which pumps will fail in the future, either the moment or just whether they fail”



State prediction



Reproducible



Efficient usage

Exploratory data analysis

Sparse labels

- Functional 0.54
- Non-functional 0.39
- Functional needs repair 0.07

Categorical variables

- Missing values: funder, installer, subvillage, public_meeting, scheme_management, scheme_name, permit, construction_year

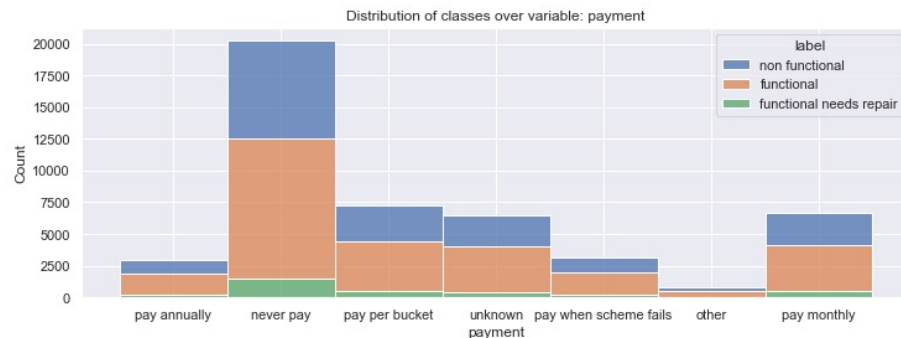
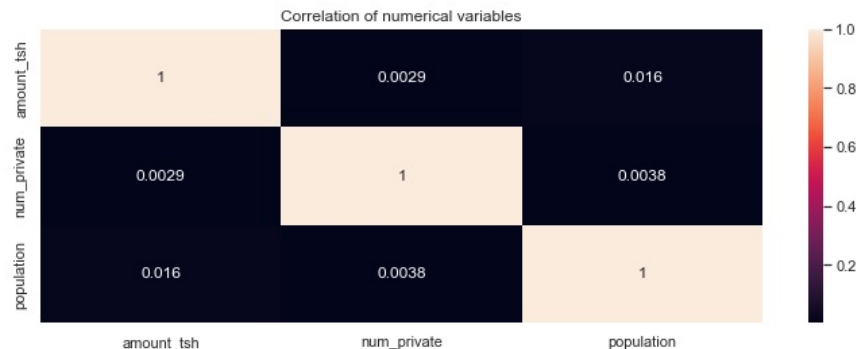
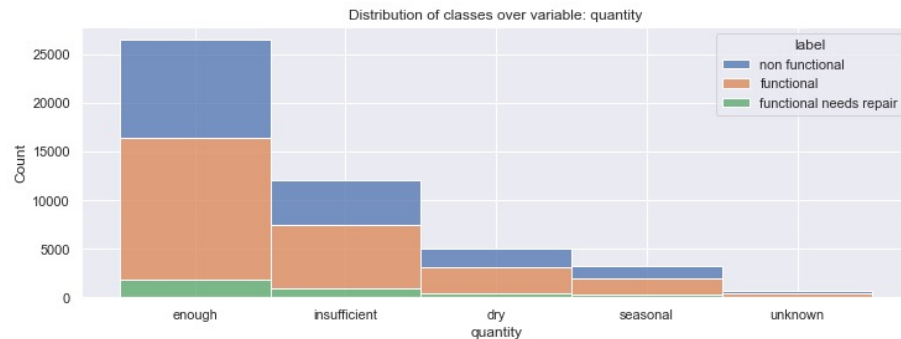
Numerical variables

- Amount_tsh, num_private, population, construction_year

Exploratory data analysis

status_group functional functional needs repair non functional

quantity			
dry	0.025136	0.005924	0.968940
enough	0.652323	0.072320	0.275357
insufficient	0.523234	0.095842	0.380924
seasonal	0.574074	0.102716	0.323210
unknown	0.269962	0.017744	0.712294



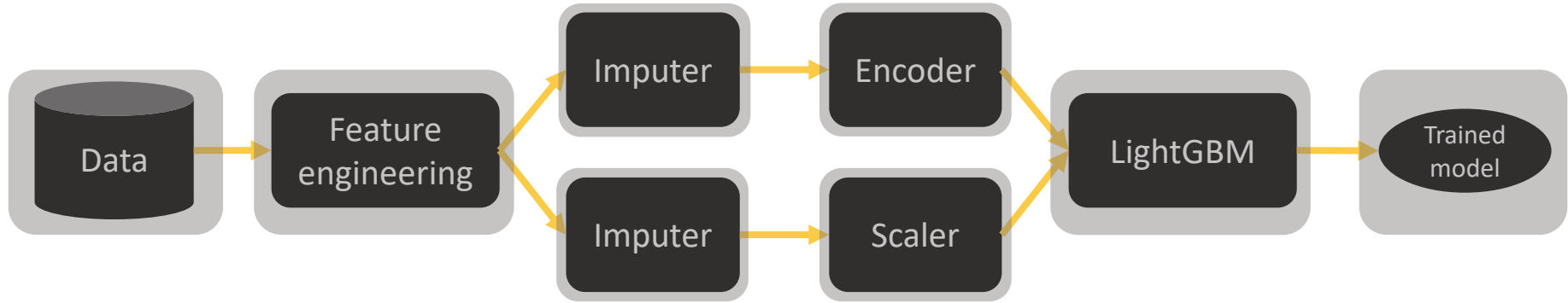
Exploratory data analysis

status_group	functional	functional needs repair	non functional
water_quality			
coloured	0.502041	0.110204	0.387755
fluoride	0.755000	0.065000	0.180000
fluoride abandoned	0.352941	0.000000	0.647059
milky	0.544776	0.017413	0.437811
salty	0.457166	0.046334	0.496499
salty abandoned	0.513274	0.212389	0.274336
soft	0.565941	0.076823	0.357236
unknown	0.140725	0.018657	0.840618

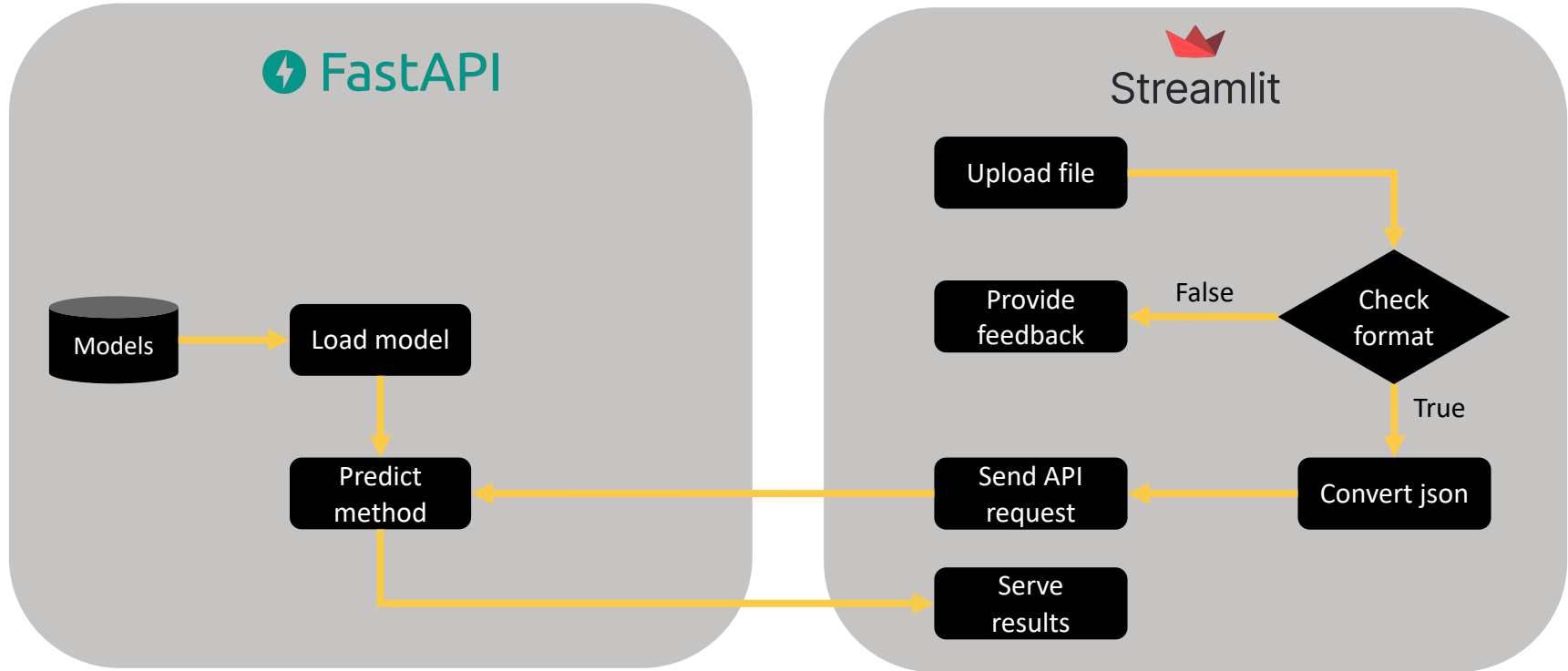
status_group	functional	functional needs repair	non functional
extraction_type_class			
gravity	0.599253	0.100859	0.299888
handpump	0.630469	0.060464	0.309067
motorpump	0.379980	0.046200	0.573820
other	0.160031	0.032037	0.807932
rope pump	0.649667	0.037694	0.312639
submersible	0.538760	0.040945	0.420295
wind-powered	0.427350	0.059829	0.512821

status_group	functional	functional needs repair	non functional
region			
Arusha	0.684776	0.052239	0.262985
Dar es Salaam	0.572671	0.003727	0.423602
Dodoma	0.458428	0.094957	0.446615
Iringa	0.782206	0.023234	0.194560
Kagera	0.520808	0.091677	0.387515
Kigoma	0.484020	0.214134	0.301847
Kilimanjaro	0.602877	0.073533	0.323590
Lindi	0.297542	0.060155	0.642303
Manyara	0.623500	0.060644	0.315856
Mara	0.449975	0.030472	0.519553
Mbeya	0.499892	0.108644	0.391464
Morogoro	0.528957	0.074888	0.396156
Mtwara	0.302890	0.072832	0.624277
Mwanza	0.484204	0.058994	0.456802

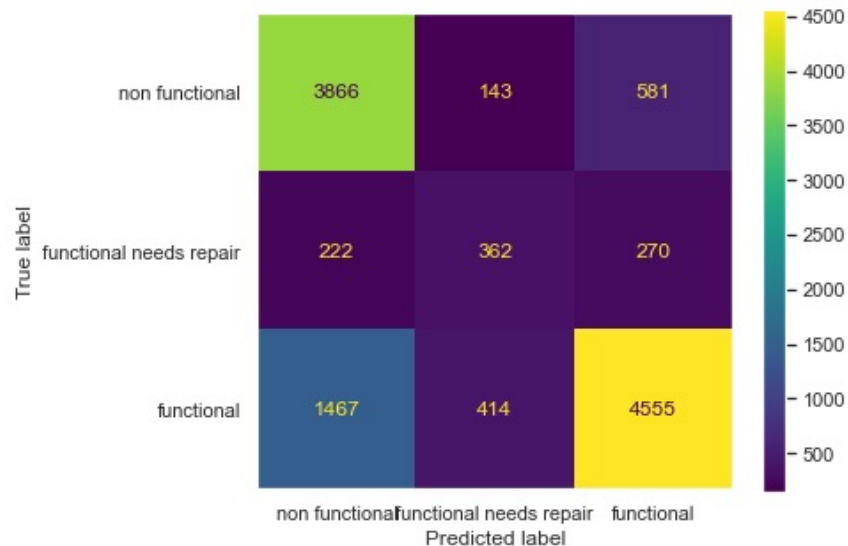
Model training



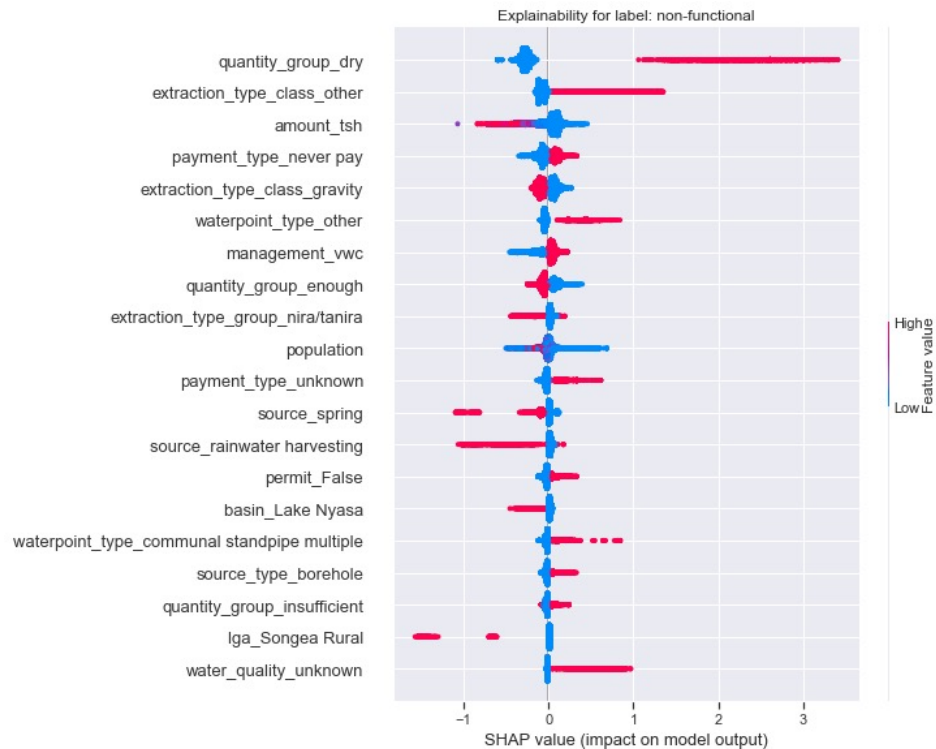
Deployment strategy



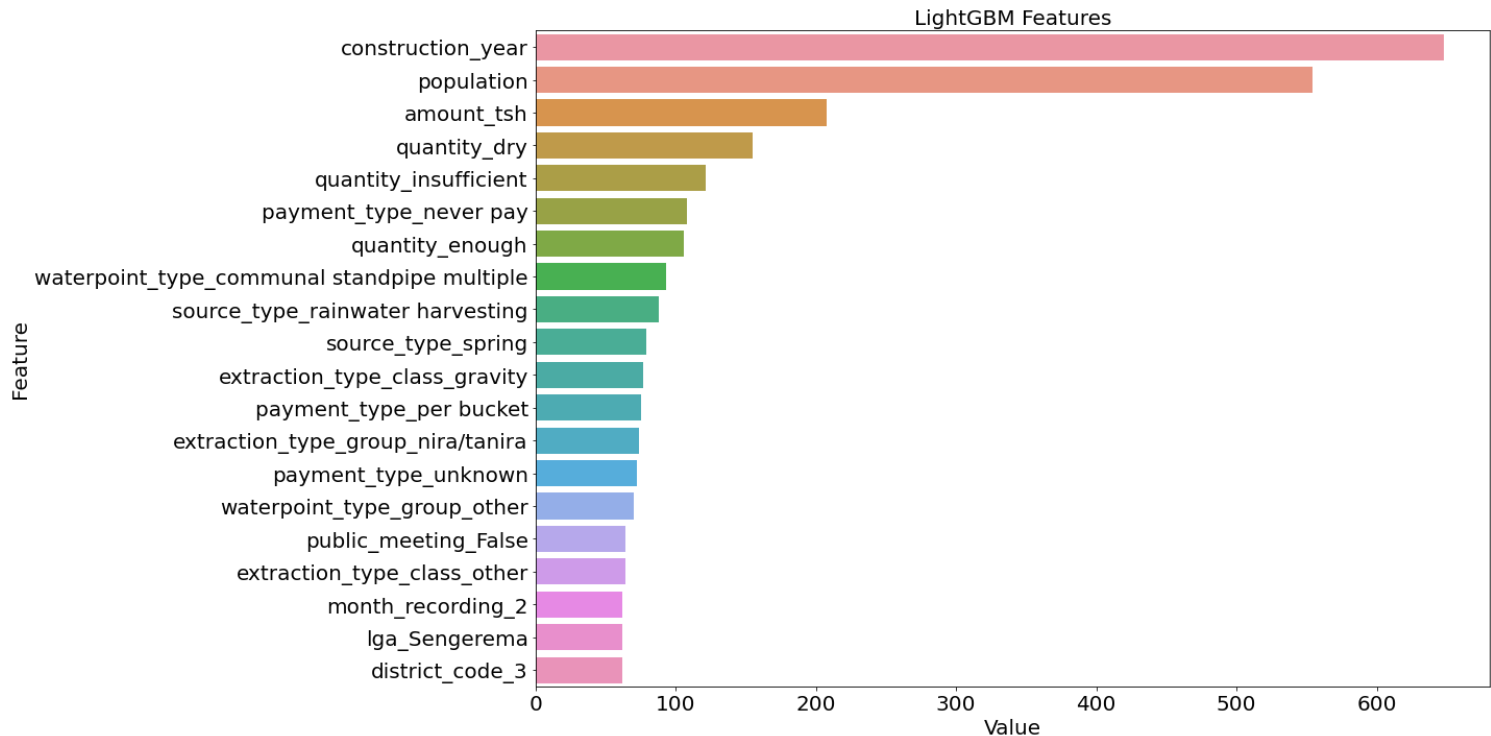
Modeling results



f1_score: 0.741




Modeling results



Recommendations

1. Feature engineering of new and existing features
 - For now, only basic feature engineering has been performed
 - Use subset of features
2. Verify data with business and discuss explainable variables
3. Extensive hyper-parameter optimization with multiple (boosting) methods, parameters and class weights
4. Save data in database and include ingestion



Predictive maintenance for water pumps in Tanzania

Thank you!