

MI Result Result #101 — Analysis Report

Generated 2026-09-21 02:36:21 · 3 run(s) · CAT-G Data Analyzer

1. pipeline stage — preprocessing (#101)

2026-09-21 02:12:52 · ml_result · status: completed

Pipeline stage preprocessing: missing values: mean.

preprocessing

Field	Value
stage	preprocessing
detail	missing values: mean

Method & Provenance

method	Random Forest
engine	Python · scikit-learn 1.8.0
datasets	gallery — cat_g_field_data_plots.csv
started	2026-09-21 02:12:52.641122
permanent link	https://cat-g-omxpert.wishartlab.com/results/by-token/a316fb7a42834890a899475de983

Parameter	Value
model_type	random_forest
target_column	PC_SHOOTMASS
n_splits	3

Command history

```
# python
# Models compared via app.services.ml_model_registry

# python
scores = sklearn.model_selection.cross_val_score(model, X, y, cv=3)
```

2. pipeline stage — feature selection (#101)

2026-09-21 02:12:52 · ml_result · status: completed

Pipeline stage feature selection: kbest_f (k=5); 40 → 5 features.

feature selection

Field	Value
stage	feature_selection
detail	kbest_f (k=5); 40 → 5 features

Method & Provenance

method	Random Forest
engine	Python · scikit-learn 1.8.0
datasets	gallery — cat_g_field_data_plots.csv
started	2026-09-21 02:12:52.641122
permanent link	https://cat-g-omxpert.wishartlab.com/results/by-token/a316fb7a42834890a899475de983

Parameter	Value
model_type	random_forest
target_column	PC_SHOOTMASS
n_splits	3

Command history

```
# python
# Models compared via app.services.ml_model_registry

# python
scores = sklearn.model_selection.cross_val_score(model, X, y, cv=3)
```

3. random forest (#101)

2026-09-21 02:12:52 · ml_result · status: completed

The model explains 27.2% of the variance in PC_SHOOTMASS under cross-validation.

r2_mean	0.271886
r2_std	0.016556
mse_mean	63.288757
mse_std	5.592898
task_type	regression

pipeline stages

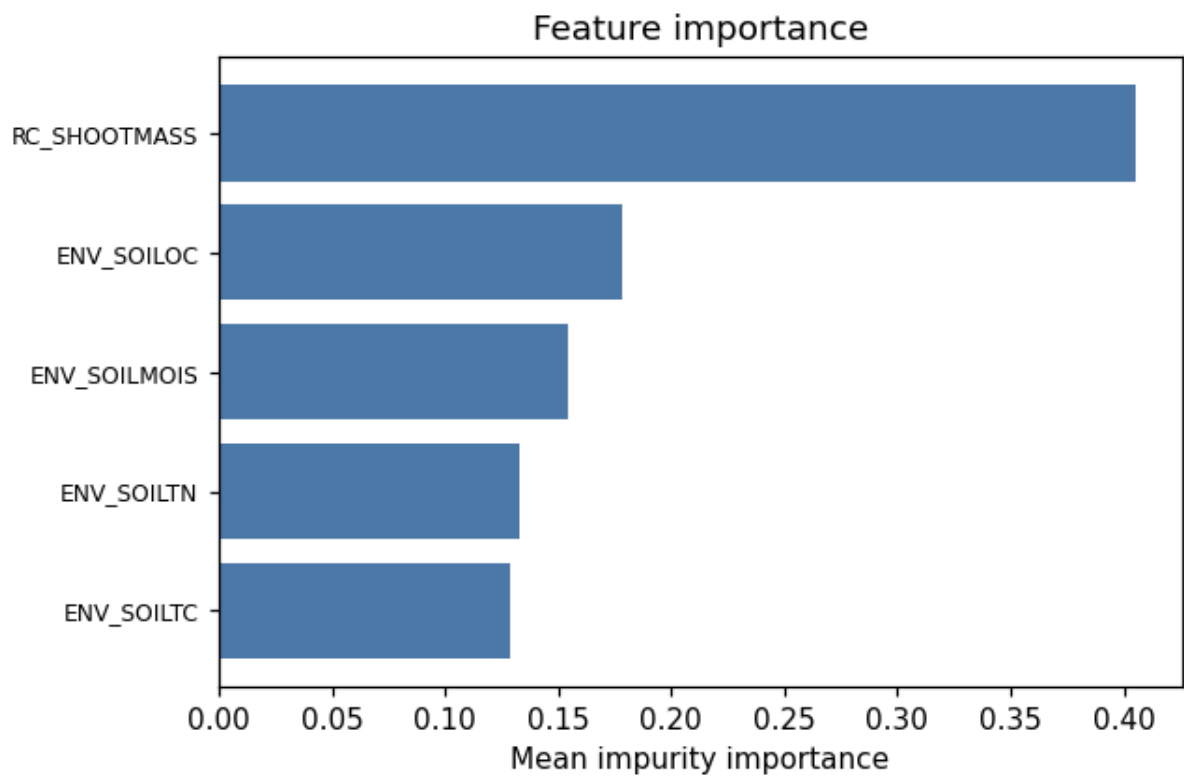
stage	detail
preprocessing	missing values: mean
feature_selection	kbest_f (k=5); 40 → 5 features
model	random_forest (regression), 3-fold cross-validation

feature importance

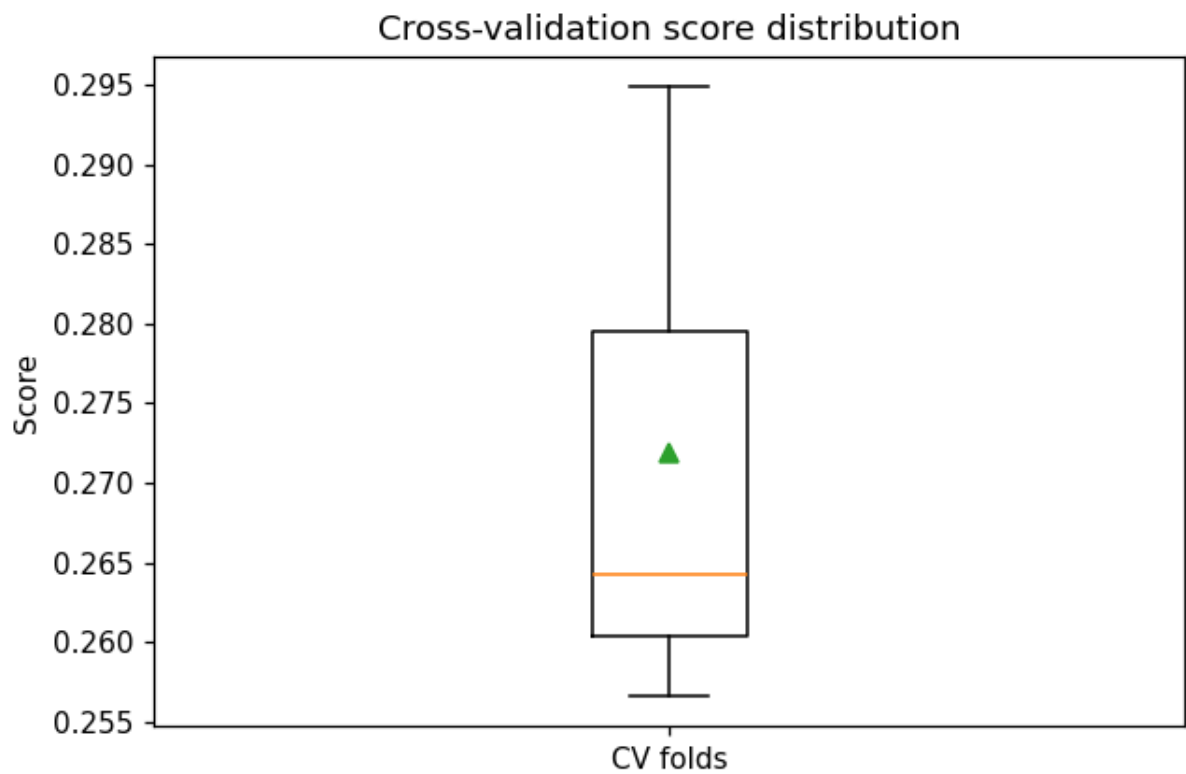
name	value
ENV_SOILOC	0.177931
ENV_SOILTN	0.133153
ENV_SOILTC	0.128963
ENV_SOILMOIS	0.154279
RC_SHOOTMASS	0.405673

split metrics

r2	mse
0.294885	65.113089
0.264187	69.041771
0.256587	55.711411



importance bar



cv boxplot

Method & Provenance

method	Random Forest
engine	Python · scikit-learn 1.8.0
datasets	gallery — cat_g_field_data_plots.csv
started	2026-09-21 02:12:52.641122
permanent link	https://cat-g-omxpert.wishartlab.com/results/by-token/a316fb7a42834890a899475de983

Parameter	Value
model_type	random_forest
target_column	PC_SHOOTMASS
n_splits	3

Command history

```
# python
# Models compared via app.services.ml_model_registry

# python
scores = sklearn.model_selection.cross_val_score(model, X, y, cv=3)
```