Forest Cover Type Prediction

Data Gate

Data Ingestion → Database Provided by USGS → Split Data Into Train and Test → **Data Ingestion Artifact**

Data Ingestion Artifact
Ingested
CSV — Train
CSV — Test

Data Validation → Scheme.yml

Scheme.yml → Validate Number of Columns → Is Data Drift Exist? → If validated — YES

Preprocessing / Feature Engineer

preprocessor.pkl

Column Transformer
- Elevation (Acumulated in 3 Critical Points)
- Consolidate Soil & Distances Features
- Circular Aspect Features + Normalized Slope (sin/cos)

Saving Objects

Data Transformation → Reduction of Feature Noise and Normalization of Geographical Features

CSV — Train
CSV — Test

Train Dataframe
Test Dataframe

Input Feature Train Dataframe → Target Feature Train Dataframe → Input Feature Test Dataframe → Target Feature Train Dataframe

Input Feature Train Array
Input Feature Test Array

Handle Imbalanced Train data → **Smooteenn** ← Handle Imbalanced Test Data

Input Feature Train Final — Train — Target Feature Train Final
Input Feature Test Final — Target Feature Test Final

train.npy
test.npy

Uncertainty / Decision / Models

Inference Core

Ensemble Loader | Probability Engine | Continuous Validation

Ensemble Loader
Load Models → XGBoost Random Forest LightGBM
Load Models → Load Artifacts → ONNX
Load Artifacts → MLflow
Transformed Features

Probability Engine
Execute Ensamble → p[7] class
Execute Ensamble → Calc of Final Predictions → np.argmax(probabilities)
Uncertainty Quantification

Continuous Validation
Seed Stability → Comparative Performance Versions

JSON File + Uncertainty Metrics

Fault Tolerancy

```
Model Training and Evaluation

                                              Hyperparameter          Train the Model        Apply ML Algorithms        Model Trainer
          Get Best                               Tunning
           Model

                      model.pkl

          Saving                                                   Model Evaluation        get_best_model from        Is model
          Model                                                                                 s3 bucket              present?

                                                      load object

                                                                                                              y_hat_best_model
          Calculate

                                                                                                              best_model_f1score
                       y_hat_trained_model
                                                         trained_model_f1score >        NO
                       trained_model_f1score              best_model_f1score
                                                                                              STOP

                                                              YES

          Model Pusher

          Push Model to S3        Fetch Zip        Load in                         Endpoint
            Bucket                                 Memory
```