**Feature Class Prediction using Machine Learning**

**Problem statement:**

To predict feature class from the feature class element and feature class geometry type.

**Requirements based on my assumptions:**

1. Independent features:
2. Feature class element – Individual feature layer that is required to model utility network in ArcGIS
3. Geometry type – Describes the type of vector data model like point, line, polygon and attributes
4. Dependent features:
5. Feature class – This is the target feature that is required to predicted. It is the layer that is required to represent the data in ArcGIS UN model.

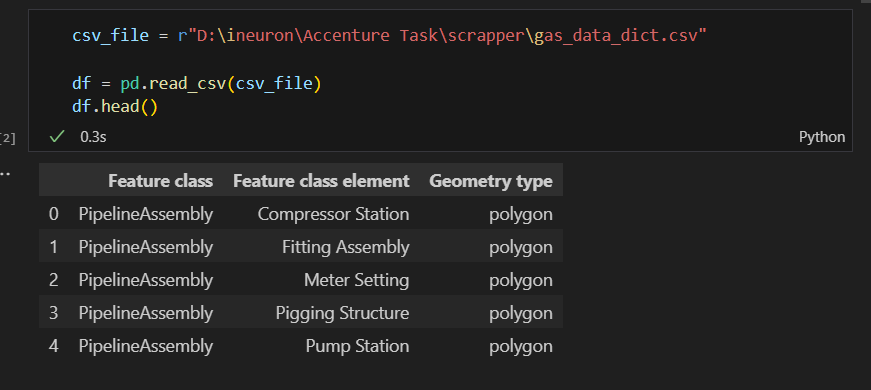
**Data Collection:**

I created a sample data from the [Gas Utility Network Configuration Data Dictionary](https://solutions.arcgis.com/gas/help/gas-utility-network-configuration/DataDictionary/DataDictionary/) . Following table was created:

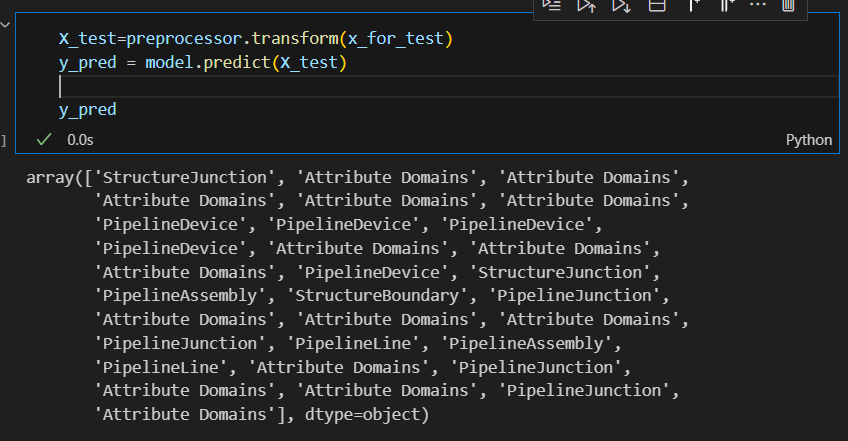
|  |  |  |
| --- | --- | --- |
| **Feature class** | **Feature class element** | **Geometry type** |
| PipelineAssembly | Compressor Station | polygon |
| PipelineAssembly | Fitting Assembly | polygon |
| PipelineAssembly | Meter Setting | polygon |
| PipelineAssembly | Pigging Structure | polygon |
| PipelineAssembly | Pump Station | polygon |
| PipelineAssembly | Regulator Station | polygon |
| PipelineAssembly | Rural Tap | polygon |
| PipelineAssembly | Town Border Station | polygon |
| PipelineAssembly | Unknown | polygon |
| PipelineAssembly | Wellhead | polygon |
| PipelineDevice | Anode | point |
| PipelineDevice | Compressor | point |
| PipelineDevice | Controllable Tee | point |
| PipelineDevice | Controllable Valve | point |
| PipelineDevice | Excess Flow Valve | point |
| PipelineDevice | Flow Valve | point |
| PipelineDevice | Lamp | point |
| PipelineDevice | Meter | point |
| PipelineDevice | Pressure Monitoring Device | point |
| PipelineDevice | Pump | point |
| PipelineDevice | Rectifier | point |
| PipelineDevice | Regulator | point |
| PipelineDevice | Relief Valve | point |
| PipelineDevice | Scrubber | point |
| PipelineDevice | Short Stop | point |
| PipelineDevice | Strainer | point |
| PipelineDevice | Test Point | point |
| PipelineDevice | Unknown | point |
| PipelineDevice | Wellhead Source Flange | point |
| PipelineJunction | Bond Junction | point |
| PipelineJunction | Connection Point | point |
| PipelineJunction | Cooling System | point |
| PipelineJunction | Coupling | point |
| PipelineJunction | Coupon | point |
| PipelineJunction | Dehydration Equipment | point |
| PipelineJunction | Drip | point |
| PipelineJunction | Elbow | point |
| PipelineJunction | Electro Stop | point |
| PipelineJunction | End Cap | point |
| PipelineJunction | Expansion Joint | point |
| PipelineJunction | Flange | point |
| PipelineJunction | Line Heater | point |
| PipelineJunction | Odorizer | point |
| PipelineJunction | Plastic Fusion | point |
| PipelineJunction | Plug | point |
| PipelineJunction | Reducer | point |
| PipelineJunction | Saddle | point |
| PipelineJunction | Screw | point |
| PipelineJunction | Tank | point |
| PipelineJunction | Tee | point |
| PipelineJunction | Transition | point |
| PipelineJunction | Union | point |
| PipelineJunction | Unknown | point |
| PipelineJunction | Weld | point |
| PipelineLine | Bond Wire | line |
| PipelineLine | Customer Pipe | line |
| PipelineLine | Distribution Pipe | line |
| PipelineLine | Gathering Pipe | line |
| PipelineLine | Rectifier Cable | line |
| PipelineLine | Riser Pipe | line |
| PipelineLine | Service Pipe | line |
| PipelineLine | Station Pipe | line |
| PipelineLine | Transmission Pipe | line |
| PipelineLine | Unknown | line |
| StructureBoundary | Pipeline Processing Facility | polygon |
| StructureBoundary | Pipeline Station Structure | polygon |
| StructureBoundary | Pipeline Vault Boundary | polygon |
| StructureBoundary | Unknown | polygon |
| StructureJunction | Pipeline Anchor | point |
| StructureJunction | Pipeline End Connection | point |
| StructureJunction | Pipeline Muffler | point |
| StructureJunction | Pipeline Pipe Hanger | point |
| StructureJunction | Pipeline Pipe Support | point |
| StructureJunction | Pipeline Valve Box | point |
| StructureJunction | Pipeline Vault | point |
| StructureJunction | Unknown | point |
| StructureLine | Pipeline Casing | line |
| StructureLine | Unknown | line |
| Attribute Domains | AnnotationStatus | attribute |
| Attribute Domains | Asset Manager | attribute |
| Attribute Domains | Asset Owner | attribute |
| Attribute Domains | BooleanSymbolValue | attribute |
| Attribute Domains | CP ASTM Anode Manufacture Component | attribute |
| Attribute Domains | CP Anode Weight | attribute |
| Attribute Domains | CP Color Code | attribute |
| Attribute Domains | CP Rectifier Cable Gauge | attribute |
| Attribute Domains | CP Rectifier Manufacturer | attribute |
| Attribute Domains | CP Test Point Wire Count | attribute |
| Attribute Domains | Gas Regulatory Type | attribute |
| Attribute Domains | Gas Valve Condition | attribute |
| Attribute Domains | HorizontalAlignment | attribute |
| Attribute Domains | Pipeline ASTM Coupling Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM EFV Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Elbow Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM End Cap Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Manufacturer | attribute |
| Attribute Domains | Pipeline ASTM Material | attribute |
| Attribute Domains | Pipeline ASTM Meter Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Pipe Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Reducer Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Regulator Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Saddle Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Strainer Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Tee Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Transition Manufacture Component | attribute |
| Attribute Domains | Pipeline ASTM Valve Manufacture Component | attribute |
| Attribute Domains | Pipeline Casing Diameter | attribute |
| Attribute Domains | Pipeline Casing Fill Type | attribute |
| Attribute Domains | Pipeline Cooling Method | attribute |
| Attribute Domains | Pipeline Device Status | attribute |
| Attribute Domains | Pipeline EFV Diameter | attribute |
| Attribute Domains | Pipeline Fitting Diameter | attribute |
| Attribute Domains | Pipeline Girth Weld Type | attribute |
| Attribute Domains | Pipeline Install Method | attribute |
| Attribute Domains | Pipeline Length Source | attribute |
| Attribute Domains | Pipeline Life Cycle Status | attribute |
| Attribute Domains | Pipeline Lining Type | attribute |
| Attribute Domains | Pipeline Meter Diameter - Customer | attribute |
| Attribute Domains | Pipeline Meter Type | attribute |
| Attribute Domains | Pipeline Odorant Type | attribute |
| Attribute Domains | Pipeline Over Pressure Protection Type | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Customer | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Distribution | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Gathering | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Riser | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Services | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Station | attribute |
| Attribute Domains | Pipeline Pipe Diameter - Transmission | attribute |
| Attribute Domains | Pipeline Pipe Outside Diameter | attribute |
| Attribute Domains | Pipeline Pipe Outside Diameter - Riser | attribute |
| Attribute Domains | Pipeline Pipe Outside Diameter - Station | attribute |
| Attribute Domains | Pipeline Pipe Outside Diameter - Transmission | attribute |
| Attribute Domains | Pipeline Pipe Wall Thickness | attribute |
| Attribute Domains | Pipeline Pipe Wall Thickness - Riser | attribute |
| Attribute Domains | Pipeline Pipe Wall Thickness - Station | attribute |
| Attribute Domains | Pipeline Pipe Wall Thickness - Transmission | attribute |
| Attribute Domains | Pipeline Pipe Wellhead Pressure Rating | attribute |
| Attribute Domains | Pipeline Pipeline Operator Number | attribute |
| Attribute Domains | Pipeline Pressure Range | attribute |
| Attribute Domains | Pipeline Reducer Fitting Size | attribute |
| Attribute Domains | Pipeline Regulator Casing Material | attribute |
| Attribute Domains | Pipeline Regulatory Type | attribute |
| Attribute Domains | Pipeline Seam Orientation | attribute |
| Attribute Domains | Pipeline Seam Type | attribute |
| Attribute Domains | Pipeline Supplier | attribute |
| Attribute Domains | Pipeline Valve Diameter | attribute |
| Attribute Domains | Pipeline Valve Position | attribute |
| Attribute Domains | Pipeline Valve Type | attribute |
| Attribute Domains | Symbol Rotation | attribute |
| Attribute Domains | VerticalAlignment | attribute |
| Attribute Domains | Yes No | attribute |
| Attribute Domains | Yes No Not Applicable | attribute |

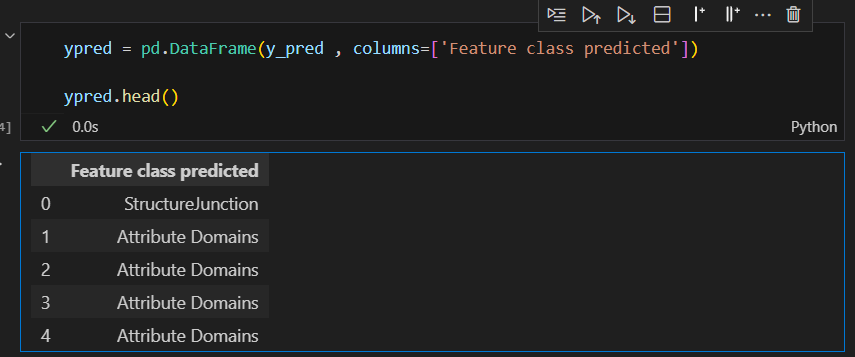
**Implementing ML model:**

1. Data loaded in pandas data frame as df.



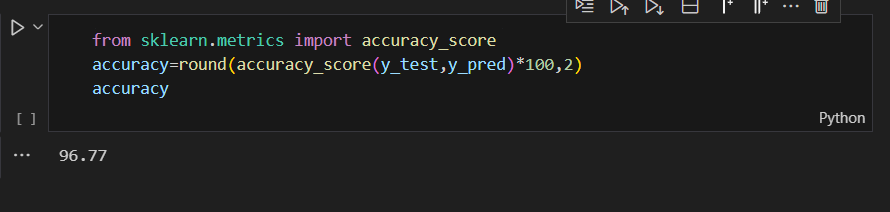
1. Split data in dependent and independent variable
   1. # Split data into features and target variable
   2. X = df[['Feature class element', 'Geometry type']]
   3. y = df[['Feature class']]
2. exporting sample data to test the model later.
   1. X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)
   2. X\_test.to\_csv('X\_test.csv', index=False)
   3. y\_test.to\_csv('y\_test.csv',  index=False)
3. Applied One hot encoding for converting the categorical variables to numerical values.
   1. # One-hot encode categorical variables
   2. cat\_x = ['Feature class element', 'Geometry type']
   3. target\_column = ['Feature class']
   4. # data\_encoded = pd.get\_dummies(df, columns=cat\_x)
   5. cat\_pipeline=Pipeline(
   7. steps=[
   9. ("imputer",SimpleImputer(strategy="most\_frequent")),
   10. ("encoder",OneHotEncoder())
   11. ]
   12. )
   13. preprocessor=ColumnTransformer([
   14. ('cat\_x',cat\_pipeline,cat\_x)
   15. ])
4. Preparing training data - X\_train and y\_train for training the model.
   1. x\_train = preprocessor.fit\_transform(X)
   2. y\_train = y
5. Applying Random Forest ML algorithm.
   1. model = RandomForestClassifier(oob\_score=True)
   2. model.fit(x\_train , y\_train)
6. Loading test data – X\_test and y\_test.
   1. # loading x\_test and y\_test data
   2. x\_for\_test = pd.read\_csv(r"D:\ineuron\Accenture Task\model\X\_test.csv")
   3. y\_test = pd.read\_csv(r"D:\ineuron\Accenture Task\model\y\_test.csv")
7. Testing prediction and exporting the prediction to new CSV.





* 1. df\_for\_export = pd.concat([x\_for\_test ,y\_test, ypred], axis=1, ignore\_index=False)
  2. df\_for\_export.to\_csv("pred\_output.csv" , index=False)

1. Checking accuracy



1. Final prediction assessment
   1. Table below represents the comparison of predicted and actual values.
   2. Actual values are in column **Feature Class**
   3. Predicted values are in column **Feature Class predicted**.
   4. Result column is comparison of both predicted and actual value. The field value TRUE represents that predicted and actual both are same/equal; while FALSE represents the vice versa.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature class element** | **Geometry type** | **Feature class** | **Feature class predicted** | **Result** |
| Pipeline Anchor | point | StructureJunction | StructureJunction | TRUE |
| Pipeline Valve Type | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline ASTM Material | attribute | Attribute Domains | Attribute Domains | TRUE |
| CP ASTM Anode Manufacture Component | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Pipe Wall Thickness - Transmission | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Reducer Fitting Size | attribute | Attribute Domains | Attribute Domains | TRUE |
| Test Point | point | PipelineDevice | PipelineDevice | TRUE |
| Controllable Tee | point | PipelineDevice | PipelineDevice | TRUE |
| Pressure Monitoring Device | point | PipelineDevice | PipelineDevice | TRUE |
| Flow Valve | point | PipelineDevice | PipelineDevice | TRUE |
| Pipeline Pipe Wall Thickness - Station | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Pipe Outside Diameter | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Cooling Method | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pump | point | PipelineDevice | PipelineDevice | TRUE |
| Pipeline End Connection | point | StructureJunction | StructureJunction | TRUE |
| Wellhead | polygon | PipelineAssembly | PipelineAssembly | TRUE |
| Pipeline Station Structure | polygon | StructureBoundary | StructureBoundary | TRUE |
| Connection Point | point | PipelineJunction | PipelineJunction | TRUE |
| Pipeline Pipe Diameter - Station | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Life Cycle Status | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Install Method | attribute | Attribute Domains | Attribute Domains | TRUE |
| Cooling System | point | PipelineJunction | PipelineJunction | TRUE |
| Distribution Pipe | line | PipelineLine | PipelineLine | TRUE |
| Unknown | polygon | StructureBoundary | PipelineAssembly | FALSE |
| Service Pipe | line | PipelineLine | PipelineLine | TRUE |
| Pipeline Pressure Range | attribute | Attribute Domains | Attribute Domains | TRUE |
| Elbow | point | PipelineJunction | PipelineJunction | TRUE |
| CP Rectifier Manufacturer | attribute | Attribute Domains | Attribute Domains | TRUE |
| Pipeline Girth Weld Type | attribute | Attribute Domains | Attribute Domains | TRUE |
| Odorizer | point | PipelineJunction | PipelineJunction | TRUE |
| CP Rectifier Cable Gauge | attribute | Attribute Domains | Attribute Domains | TRUE |