

```
@tracer(cat_col = ['race', 'occupation', 'education'], numerical_col = ['age', 'hours-per-week'])
def adult_pipeline_normal(f_path = '../pipelines/adult-sample_missing.csv'):
    raw_data = pd.read_csv(f_path, na_values='?')
    data = raw_data.dropna()

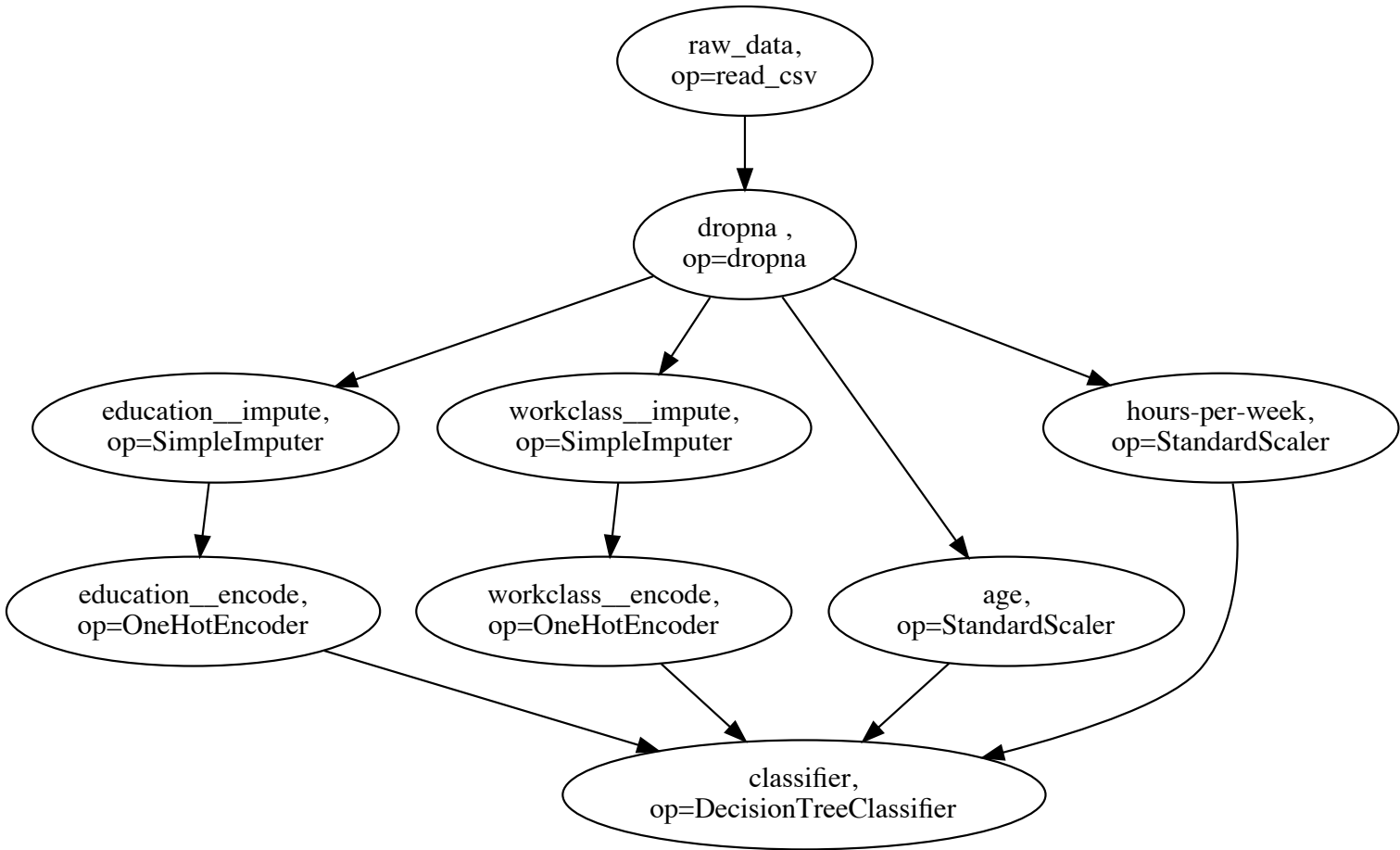
    labels = label_binarize(data['income-per-year'], ['>50K', '<=50K'])

    nested_categorical_feature_transformation = Pipeline(steps=[
        ('impute', SimpleImputer(missing_values=np.nan, strategy='most_frequent')),
        ('encode', OneHotEncoder(handle_unknown='ignore'))
    ])

    nested_feature_transformation = ColumnTransformer(transformers=[
        ('categorical', nested_categorical_feature_transformation, ['education', 'workclass']),
        ('numeric', StandardScaler(), ['age', 'hours-per-week'])
    ])

    nested_pipeline = Pipeline([
        ('features', nested_feature_transformation),
        ('classifier', DecisionTreeClassifier())
    ])

    return nested_pipeline
```



Start Pandas Opeation

Inpected raw_data = pd.read_csv(f_path, na_values='?')

Changes in numerical features!

	count	missing_count	median	mad	range
age	-14.0	0.0	0.0	-0.7413	-23.0
hours-per-week	-14.0	0.0	0.0	0.0000	0.0

Changes in categorical features!

	missing_count	num_class	class_count	class_percent
race	-4.0	0.0	{'White': -6, 'Black': -2, 'Amer-Indian-Eskimo': -2, 'Other': 0, 'Asian-Pac-Islander': 0}	{'White': 0.0271, 'Black': -0.0111, 'Amer-Indian-Eskimo': -0.0184, 'Other': 0.0012, 'Asian-Pac-Islander': 0.0012}
occupation	-8.0	0.0	{'Exec-managerial': 0, 'Adm-clerical': 0, 'Craft-repair': -1, 'Sales': -1, 'Other-service': 0, 'Prof-specialty': -2, 'Transport-moving': -1, 'Machine-op-inspct': 0, 'Farming-fishing': 0, 'Handlers-cleaners': 0, 'Tech-support': 0, 'Protective-serv': -1}	{'Exec-managerial': 0.0106, 'Adm-clerical': 0.0099, 'Craft-repair': -0.0018, 'Sales': -0.0033, 'Other-service': 0.0068, 'Prof-specialty': -0.0157, 'Transport-moving': -0.0056, 'Machine-op-inspct': 0.0046, 'Farming-fishing': 0.0023, 'Handlers-cleaners': 0.0015, 'Tech-support': 0.0008, 'Protective-serv': -0.0101}
education	-2.0	0.0	{'HS-grad': -3, 'Bachelors': -1, 'Some-college': -4, '11th': -2, 'Masters': -2, '7th-8th': 0, '10th': 0, 'Assoc-voc': 0, 'Prof-school': 0, 'Assoc-acdm': 0, '12th': 0, '5th-6th': 0}	{'HS-grad': 0.0078, 'Bachelors': 0.0183, 'Some-college': -0.0138, '11th': -0.0133, 'Masters': -0.0147, '7th-8th': 0.0043, '10th': 0.0028, 'Assoc-voc': 0.0028, 'Prof-school': 0.0014, 'Assoc-acdm': 0.0014, '12th': 0.0014, '5th-6th': 0.0014}

Inpected data = raw_data.dropna()

Start Sklearn Pipeline

Operations SimpleImputer on education

Operations OneHotEncoder on education

Changes in categorical features!

	education
missing_count	0
num_class	-10
class_count	{0.0: 84, 1.0: 2}
class_percent	{0.0: 0.9767, 1.0: 0.0233}

Operations StandardScaler on age

Changes in numerical features!

	age
count	0.0000
missing_count	0.0000
median	-36.0972
mad	-12.8320
range	-44.6418

Operations StandardScaler on hours-per-week

Changes in numerical features!

	hours-per-week
count	0.0000
missing_count	0.0000
median	-40.1126
mad	-1.3509
range	-63.7813
