

Assignment 3: Split sample data into training and test sets. (Use suitable data set).

```
Code =>
import pandas as pd

from sklearn.preprocessing import LabelEncoder, StandardScaler

# Load Dataset
train_path = r"C:\Users\Shreyash Musmade\Desktop\Practical\AICS\AICS_Prac-4\train.csv"
test_path = r"C:\Users\Shreyash Musmade\Desktop\Practical\AICS\AICS_Prac-4\test.csv"

train_data = pd.read_csv(train_path)
test_data = pd.read_csv(test_path)

print("Train Data Shape:", train_data.shape)
print("Test Data Shape:", test_data.shape)

# Combine Train and Test for Consistent Feature Engineering
data = pd.concat([train_data, test_data], axis=0, ignore_index=True)

# 1. Handle Missing Values
data['Age'].fillna(data['Age'].median(), inplace=True)
data['Embarked'].fillna(data['Embarked'].mode()[0], inplace=True)
data['Fare'].fillna(data['Fare'].median(), inplace=True)

# 2. Drop Columns with Excessive Missing Data
data.drop(['Cabin', 'Ticket', 'Name'], axis=1, inplace=True)

# 3. Encode Categorical Features
label_encoder = LabelEncoder()
data['Sex'] = label_encoder.fit_transform(data['Sex'])
data['Embarked'] = label_encoder.fit_transform(data['Embarked'])

# 4. Feature Scaling
scaler = StandardScaler()
data[['Age', 'Fare']] = scaler.fit_transform(data[['Age', 'Fare']])

# 5. Create New Features
data['FamilySize'] = data['SibSp'] + data['Parch'] + 1
data['IsAlone'] = (data['FamilySize'] == 1).astype(int)

print("Feature Engineering Completed.")
print(data.head())

# Split back to train and test
train_final = data[:len(train_data)]
test_final = data[len(train_data):]
print("Final Train Shape:", train_final.shape)
print("Final Test Shape:", test_final.shape)
```

Assignment 3: Split sample data into training and test sets. (Use suitable data set).

Output =>

```
Running] python -u "c:\Users\Shreyash  
Musmade\Desktop\Practical\AICS\AICS_Prac-4\Practical.py"
```

Train Data Shape: (891, 12)

Test Data Shape: (418, 11)

Feature Engineering Completed.

	PassengerId	Survived	Pclass	Sex	...	Fare	Embarked	FamilySize	Is Alone
0	1	0.0	3	1	...	-			
0.503291	2		2		0				
1	2	1.0	1	0	...	0.734744	0	2	
0									
2	3	1.0	3	0	...	-			
0.490240	2		1		1				
3	4	1.0	1	0	...	0.383183	2	2	
0									
4	5	0.0	3	1	...	-			
0.487824	2		1		1				

[5 rows x 11 columns]

Final Train Shape: (891, 11)

Final Test Shape: (418, 11)

```
[Done] exited with code=0 in 9.585 seconds
```