

Skillshed Clustering 2023 Project

Nghi Nguyen

Python Code

- `FuzzyClusteringTestRun.py`

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from os import getcwd, listdir
from fcmeans import FCM
from sklearn.manifold import TSNE

# Path to the data
data_path = getcwd() + "/data/raw"

# Read the OCC Test Run CSV into pandas
df = pd.read_csv(f"{data_path}/OCC_testrun.csv")

# Columns
print(df.columns)

# Num clusters variable
num_clusters = 5

# Instantiate a FCM model with 4 clusters
model = FCM(n_clusters=num_clusters)

# Features are everything except SOC_Code
features = df.drop(['SOC_Code', 'Title'], axis=1).values

print(type(features))
print(features.shape)

# Fit FCM Model
model.fit(features)

# Centers, Labels
centers = model.centers
labels = model.predict(features)

print(centers.shape)
print(labels.shape)

# Look at the the two artifacts as a DataFrame
df['cluster_labels'] = labels

# Value Counts of Predicted Labels with 4 Clusters

## Transform counts into a DataFrame
label_counts = df['cluster_labels'].value_counts().reset_index()

# Verify that result is a DataFrame
print(type(label_counts))

# Rename Columns

```

```

label_counts.columns = ['cluster_label', 'frequency']

# Create Pie Chart of Frequencies
label_counts.plot(x='cluster_labels', y='frequency', kind='pie', figsize=(10,10))
plt.show()

# Get DF Columns
cols = df.columns.tolist()

# FCM Labels
fcm_labels = model.u.argmax(axis=1)

# Create tSNE Components
tsne = TSNE(n_components=2, verbose=1, random_state=123)

z = tsne.fit_transform(features)

# DataFrame
result_df = pd.DataFrame()
result_df['target'] = df['cluster_labels']
result_df['comp-1'] = z[:,0]
result_df['comp-2'] = z[:,1]

plt.figure(figsize=(10,10))
sns.scatterplot(x="comp-1", y='comp-2', hue=result_df['target'].tolist(), palette=sns.c
plt.show()

```

Visualizations

- 5 clusters

