Predicting User Visit in Stochastic Web Surfing

We have built and compared four models to predict a users visit in a stochastic web surf. We have used User-based Neighboring, K-means, Mean shift clustering and SVD matrix factorization

Getting Started

These instructions will get you a copy of the project up and running on your local machine for development and testing purposes. Read the steps below for instructions.

Prerequisites

Central Processing Unit (CPU) — Intel Core i5 6th Generation processor or higher

RAM — 4 GB minimum, 8 GB or higher is recommended.

Operating System — Ubuntu or Microsoft Windows 10/Mac.

Python- Latest version and required libraries.

```
Libraries:
numpy
sklearn
seaborn
collections
yellowbrick
scipy
```

Installing

Setup python environment and have all libraries installed in the environment in order to run the program. Import the required libries

```
import numpy
from sklearn.cluster import KMeans
import seaborn as sns;
from collections import Counter
from sklearn.metrics.pairwise import cosine_similarity
from yellowbrick.cluster import KElbowVisualizer
from sklearn.cluster import MeanShift, estimate_bandwidth
from scipy.sparse.linalg import svds
```

Also, have the data- training.txt and test.txt ready in the python environment.

Running the tests

In order to run the models, you will have to run the main.py file.

Models:

```
dataPreprocessing('training.txt') is used to separate user and web areas for training data.
dataPreprocessing('test.txt') is used to separate user and web areas for test data.

Mean Shift Clustering- meanShiftClustering(userValues, userInteraction, testuserValues, testLabel
SVD- svdPrediction(testuserValues, testLabel,topN)
User Based Neighboring- userBasedNeighboring(userValues,userInteraction,testuserValues,testLabel,
K means Clustering- kmeanClustering(userValues,userInteraction,testuserValues,testLabel,topN)
```

Deployment

Run the main.py after all libraries have been installed.

Built With

- [Python]

Authors

- **Himanshu Nirmal**
- **Krishna Prasad**
- **Sachin Sujir**

License

Open source.