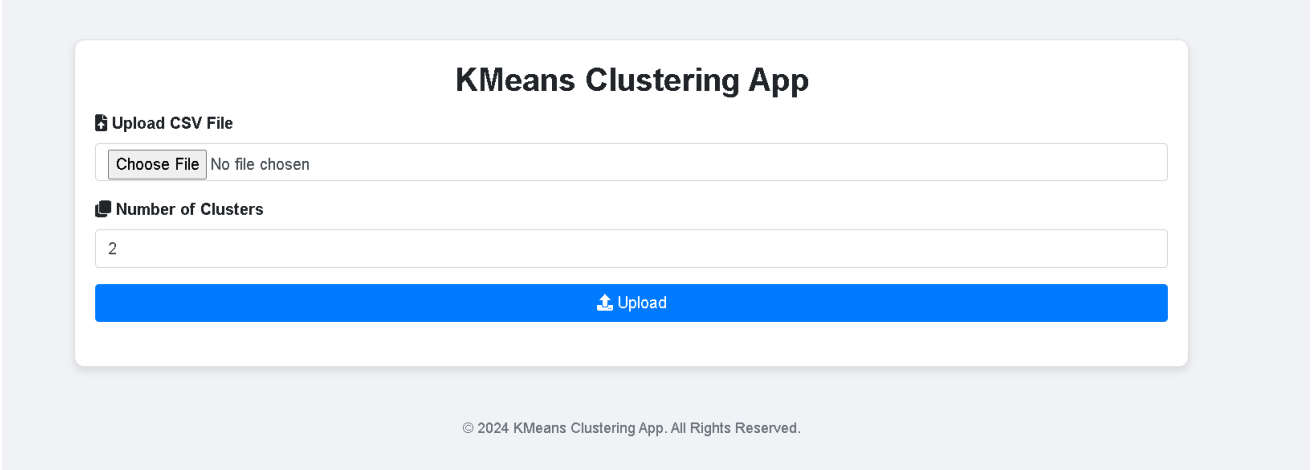
Kmeans Clustering Customer Segmentation

Wireframe Documentation

Application Overview

The KMeans Clustering Application is a web-based tool designed to facilitate exploratory data analysis through clustering. Leveraging the KMeans algorithm and Principal Component Analysis (PCA), this application allows users to upload CSV datasets, preprocess the data by handling missing values and scaling features, and then apply KMeans clustering to identify patterns and groupings within the data. Users can specify the number of clusters and receive immediate feedback, including a data preview, silhouette scores, and visual representations of clusters using interactive PCA plots. This application serves as a valuable resource for data scientists and analysts seeking to gain insights into their datasets and explore underlying patterns and structures efficiently.



Flowchart Overview

The flowchart of the KMeans Clustering Application illustrates a clear sequence of operations from the user's interaction to the display of results. The process begins with the user uploading a CSV file containing the dataset to be clustered. The application checks the file type and reads the data, handling any potential errors such as non-numeric columns or missing values. Once the data is processed, it undergoes preprocessing steps including imputation and scaling. The user specifies the number of clusters and submits the form. The application then loads a pre-trained KMeans model, updates the number of clusters, and fits the model to the preprocessed data. After clustering, it calculates the silhouette score and reduces the dimensionality of the data using Principal Component Analysis (PCA). The results are visualized through a scatter plot of the clusters and displayed alongside a summary of the data's shape and cluster centers. Flash messages provide feedback throughout the process. This flowchart ensures a streamlined user experience and efficient handling of the clustering process.

flowchart.fun

Detailed Page Wireframes

The Detailed Page Wireframes for the KMeans Clustering Application include several key components. The Home Page features a prominent header with the application's name and a form for uploading a CSV file and specifying the number of clusters. Flash messages provide feedback on file upload status and errors. Upon submission, the Result Page displays a data preview table, the shape of the dataset, the silhouette score for the chosen number of clusters, and a scatter plot of the clustered data. Additionally, the cluster centers and a back-to-home button are presented. The pages are designed to be responsive and intuitive, guiding users through the clustering process with clear visualizations and informative summaries.

Data Preview

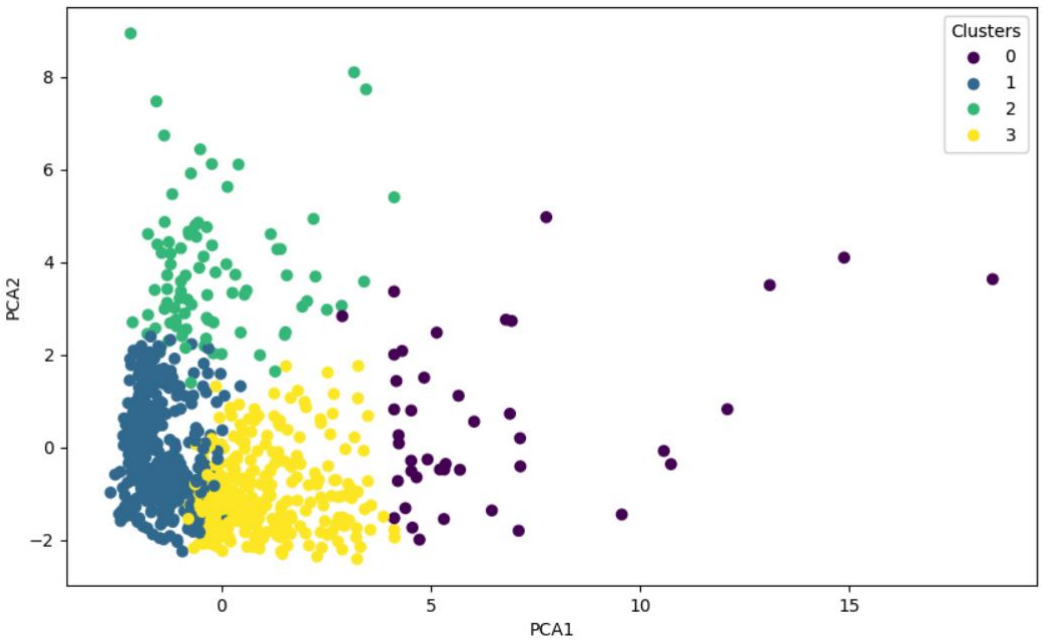
	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREC
0	40.900749	0.818182	95.40	0.00	95.4	0.000000	0.166667
1	3202.467416	0.909091	0.00	0.00	0.0	6442.945483	0.000000
2	2495.148862	1.000000	773.17	773.17	0.0	0.000000	1.000000
3	1666.670542	0.636364	1499.00	1499.00	0.0	205.788017	0.083333
4	817.714335	1.000000	16.00	16.00	0.0	0.000000	0.083333

Cluster Centers

	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FRE
0	6005.736594	0.997835	12892.686190	8984.216667	3908.469524	861.774377	0.960317
1	1792.111691	0.825026	362.986114	271.581522	91.404592	892.121526	0.178018
2	6280.947121	0.967620	900.802738	560.622500	340.180238	6950.007955	0.393254
3	1823.871667	0.966900	2098.375738	1185.251705	913.570721	314.523976	0.909647

Shape of Dataset: (799, 17)

Silhouette Score for 4 Clusters: 0.2278520228572157 9


Clusters
0
1
2
3
PCA2
8
6
4
2
0
−2
0
5
10
15
PCA1