



AAM IPL Week 7

K-Means Clustering – Student’s Choice Project

B.Tech – CSE(AIML)

V Semester - ML and AIUP, Sept-Nov 2024

Department of Computer Science Engineering – AI and ML (CSM)

G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP

Algorithm of Application

K-Means Clustering

Project Title

Clustering Analysis with K-Means

Project Objective

Implement and analyze K-means clustering on a dataset of your choice to explore patterns and group data points based on similarity.

Dataset

- Description
 - Students may choose any dataset relevant to clustering, such as customer segmentation, image compression, or geographical clustering
- Dataset Details:
 - Ensure the dataset has multiple features that can be visualized meaningfully in 2D space
 - Suggested examples include the Iris dataset, Mall Customer Segmentation, or any real-world dataset that allows for clustering
- Features:
 - List the features of your chosen dataset and their descriptions
- Usage in Machine Learning:
 - Describe the purpose of clustering within the context of your dataset
- Data Source and Published By:
 - Specify the source or URL of the dataset for reproducibility
- Data Download Link:
 - Provide the link to download the dataset





Implementation Steps

1. Data Loading and Preprocessing
 - a. Import the chosen dataset, clean, and preprocess it as necessary.
2. Plotting Original Dataset
 - a. Create a scatter plot of the original data to visualize its structure.
3. PCA Transformation and Plotting
 - a. Apply PCA to reduce dimensionality and visualize the data in 2D if the dataset has more than two features.
4. Elbow Method for Optimal K
 - a. Implement the elbow method by plotting inertia against the number of clusters and identify the optimal number of clusters.
5. Silhouette Score Calculation
 - a. Calculate and plot the silhouette score for each K value to evaluate clustering quality.
6. K-Means Clustering and Visualization
 - a. Apply K-means with the optimal K value, visualize the resulting clusters, and annotate centroids.
7. Final Analysis
 - a. Describe observations about cluster distributions and how they relate to your dataset.

Project Files Provided

- Project shell code file - **AAM-IPL-Wk-7-K-Means-Clustering-Student_Choice-Proj-Shell-Code.ipynb**
- Training Data – **Not Provided**
- Watermark image for plots - **AAM-IPL-Watermark-for-Plots.png**

Project Overview, Implementation and Submission Timeline

02-11-2024 – Saturday 10:30 AM – Next Week Project Details Announcement – Topic, Data Set, Shell Code etc. Announcement Channels – Google Class, Industry Projects WhatsApp Group			
12	02-11-2024 - Saturday Duration: 1.5 Hrs	K Means Clustering – Model Overview/Recap, Project Description, and Interactive Q&A	Online – Google Class
13	02-11-2024 - Sunday Duration: 1.5 Hrs	K Means Clustering – Model Building, Output Demonstration, Q&A	Online – Google Class
07-11-2024 – Thursday 11:59 PM - Deadline to upload the project code submission by all students in Google Class.			

Guest Lecture Timings:

Saturdays: 10:30 AM IST – 12:00 Noon IST

Sundays: 10:30 AM IST – 12:00 Noon IST

Mondays: 6:30 PM IST – 8:00 PM IST

Development Environment

- Computing Language – Python
- IDE – Visual Studio Code with Jupyter Notebook

Instructor

Instructor

Venkateswar Reddy Melachervu
(alumnus of GPREC, ECE Class of '92)
CTO, Brillium Technologies, Bengaluru
Email: venkat.reddy.gf@gprec.ac.in
Profile: [LinkedIn](#)

Visiting Faculty

Coordination

All the activities of this programme – lecture venues, weekly projects details announcements, general announcements, changes in lecture timings, etc. will be coordinated by CSM faculty member Sri V.Suresh.

Channels of Communication and Announcements

- Google Classroom
- Whatsapp group - **Applied AI & ML Industry Projects Lab**
- Emails (Strictly GPREC email addresses only)

Programme Coordinator

Prof. V.Suresh
Email: vsuresh.ecs@gprec.ac.in

Faculty Member, CSM

Reference Books

- Pattern Recognition and Machine Learning by Chris Bishop, 2006 – [PDF Link](#)
- Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar, Wiley 2019 – [PDF Link](#)

Policies

- **Attendance:** All sessions are expected to be attended by all the enrolled students. In case of inability to attend, prior information is expected to be provided by the student to the coordinator with a copy to the visiting faculty
- **Project Submissions:** Duly completed projects (Jupyter NB file and a PDF of the Jupyter NB file) are expected to be submitted through google class prior to the deadline. In case of inability to complete due various unforeseen circumstance, students are expected to seek extension for the submission deadline.
- **Academic Integrity:** Students are expected to uphold the highest standards of academic integrity in all assignments for the Applied AI & ML Industry Projects Lab. Each assignment must be the student's own work, and all sources and collaborators must be properly acknowledged. By submitting their completed project source code, students confirm that they have adhered to this integrity policy and completed their work in an honest and ethical manner.

Additional Information

For students interested in engaging with special projects in the field of Gen AI, please reach out to the visiting faculty at venkat@brillium.in for further details and opportunities.

Contact Information

For any questions or concerns or further details on this programme, please contact **Program Coordinator** during office hours or via email.

----- End of the Document -----