Vision of the Department

To be a well-known centre for pursuing computer education through innovative pedagogy, value-based education and industry collaboration.

Mission of the Department

To establish learning ambience for ushering in computer engineering professionals in core and multidisciplinary area by developing Problem-solving skillsthrough emerging technologies.

Session 2025-2026

Vision: To help businesses uncover crucial insights	**Mission:** To be a good data scientist

Program Educational Objectives of the program (PEO): (broad statements that describe the professional and career accomplishments)

PEO1	**Preparation**	**P: Preparation**	**Pep-CL abbreviation pronounce as Pep-si-lL easy to recall**
PEO2	**Core Competence**	**E: Environment (Learning Environment)**	
PEO3	**Breadth**	**P: Professionalism**	
PEO4	**Professionalism**	**C: Core Competence**	
PEO5	**Learning Environment**	**L: Breadth (Learning in diverse areas)**	

Program Outcomes (PO): Implement and analyze various deep learning algorithms such as linear regression, neural networks and CNN using modern tools and frameworks

Keywords of POs:
Engineering knowledge,Problem analysis, Design/development of solutions,Conduct Investigations of Complex Problems, Engineering Tool Usage, The Engineer and The World,Ethics, Individual and Collaborative Team work,Communication, Project Management and Finance, Life-Long Learning

PSO Keywords: Cutting edge technologies, Research

"I am an engineer, and I know how to apply engineering knowledge to investigate, analyse and design solutions to complex problems using tools for entire world following all ethics in a collaborative way with proper management skills throughout my life." *to contribute to the development of cutting-edge technologies and Research*.

Integrity: I will adhere to the Laboratory Code of Conduct and ethics in its entirety.

Name and Signature of Student and Date
Soham pimpalgaonkar – 06/11/2025

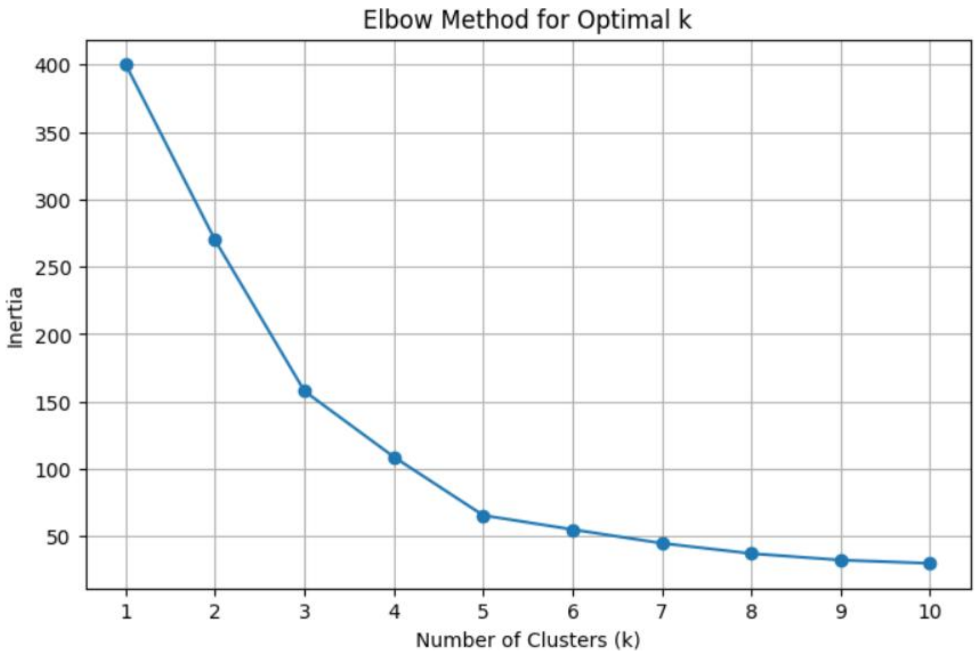
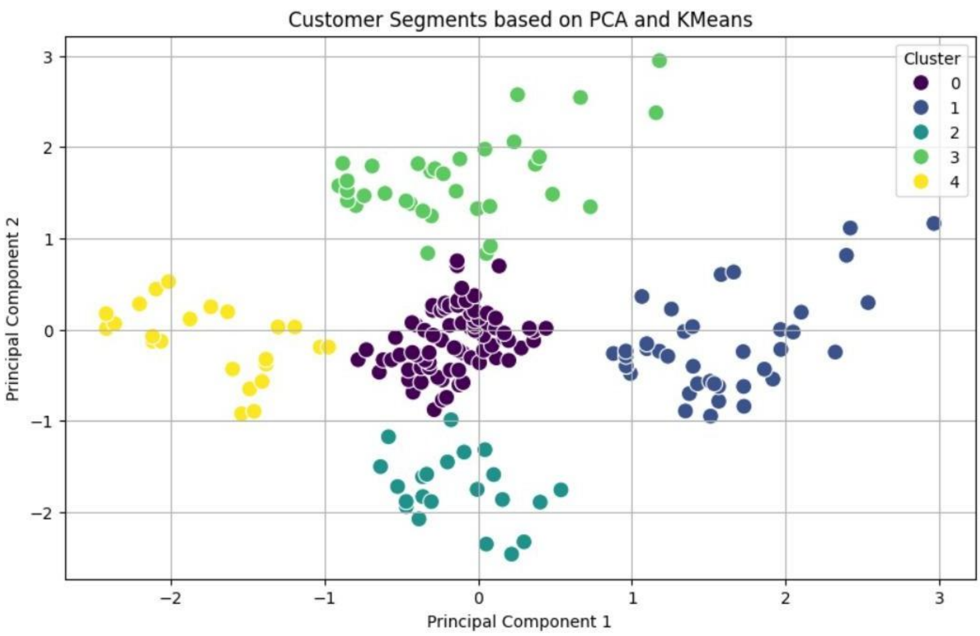
Session	**2024-25 (ODD)**	**Course Name**	**Deep Learning Lab**
Semester	**7**	**Course Code**	**22ADS702**
Roll No	62	**Name of Student**	Soham pimpalgaonkar

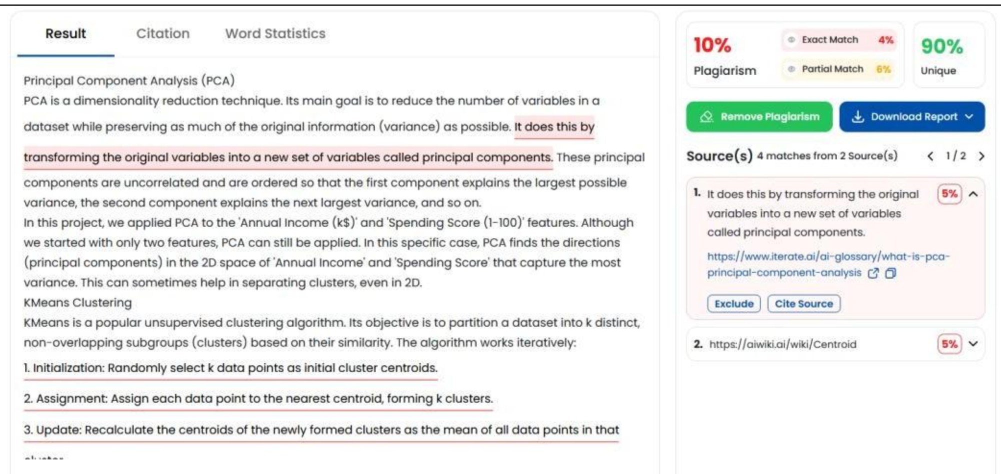
Practical Number	9
Course Outcome	**CO1:-** Implement and analyze various deep learning algorithms such as linear regression, neural networks and CNN using modern tools and frameworks **CO2:-** Develop and evaluate deep learning models for sequence data, images processing, and unsupervised learning using appropriate optimization techniques and validation strategies
Aim	Customer Segmentation using PCA + Clustering
Problem Definition	Customer Segmentation using PCA + Clustering
Theory (100 words)	**Principal Component Analysis (PCA)** PCA is a dimensionality reduction technique. Its main goal is to reduce the number of variables in a dataset while preserving as much of the original information (variance) as possible. It does this by transforming the original variables into a new set of variables called principal components. These principal components are uncorrelated and are ordered so that the first component explains the largest possible variance, the second component explains the next largest variance, and so on. In this project, we applied PCA to the 'Annual Income (k$)' and 'Spending Score (1-100)' features. Although we started with only two features, PCA can still be applied. In this specific case, PCA finds the directions (principal components) in the 2D space of 'Annual Income' and 'Spending Score' that capture the most variance. This can sometimes help in separating clusters, even in 2D. **KMeans Clustering** KMeans is a popular unsupervised clustering algorithm. Its objective is to partition a dataset into *k* distinct, non-overlapping subgroups (clusters) based on their similarity. The algorithm works iteratively: 1. **Initialization**: Randomly select *k* data points as initial cluster centroids. 2. **Assignment**: Assign each data point to the nearest centroid, forming *k* clusters.

	3. **Update**: Recalculate the centroids of the newly formed clusters as the mean of all data points in that cluster. 4. **Repeat**: Repeat steps 2 and 3 until the centroids no longer move significantly or a maximum number of iterations is reached. In this project, we applied KMeans clustering to the data after it was transformed by PCA. The elbow method was used to help determine a suitable number of clusters (*k*). The algorithm then grouped the customers into *k* clusters based on their values in the principal components space. By combining PCA and KMeans, we first reduced the potential noise and correlation in the features (though less critical with only two features) and then grouped similar customers together based on their income and spending patterns, as represented by the principal components. This allows us to identify distinct customer segments.
Procedure and Execution (100 Words)	Steps for Implementation: 1. Load the dataset 2. Select the features 3. Scale the data 4. Apply Principal Component Analysis 5. Determine optimal number of clusters 6. Apply KMeans clustering 7. Visualise the clusters

	Code:
	import pandas as pd
	
	df = pd.read_csv("/content/Mall_Customers.csv")
	display(df.head())
	
	from sklearn.preprocessing import StandardScaler
	from sklearn.decomposition import PCA
	from sklearn.cluster import KMeans
	import matplotlib.pyplot as plt
	import seaborn as sns
	
	# Select features
	features = ['Annual Income (k$)', 'Spending Score (1-100)']
	X = df[features]
	
	# Preprocess data
	scaler = StandardScaler()

```
X_scaled = scaler.fit_transform(X)

# Apply PCA
pca = PCA(n_components=2) # Reduce to 2 dimensions for visualization
X_pca = pca.fit_transform(X_scaled)

# Determine optimal number of clusters using the elbow method
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(X_pca)
    inertia.append(kmeans.inertia_)

# Plot the elbow method graph
plt.figure(figsize=(8, 5))
plt.plot(range(1, 11), inertia, marker='o')
plt.title('Elbow Method for Optimal k')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Inertia')
plt.xticks(range(1, 11))
plt.grid(True)
plt.show()

# Choose the optimal number of clusters based on the elbow method plot.
Let's assume 5 for now.
optimal_k = 5

# Apply KMeans clustering with the optimal number of clusters
kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init=10)
clusters = kmeans.fit_predict(X_pca)

# Add the cluster labels to the original dataframe
df['Cluster'] = clusters

# Visualize the clusters in the PCA-transformed space
plt.figure(figsize=(10, 6))
sns.scatterplot(x=X_pca[:, 0], y=X_pca[:, 1], hue=clusters, palette='viridis',
s=100)
plt.title('Customer Segments based on PCA and KMeans')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.legend(title='Cluster')
plt.grid(True)
plt.show()
```

	Output:  
Output Analysis	In this case, the original features were 'Annual Income (k$)' and 'Spending Score (1-100)'. PCA transformed these two features into two new principal components. These components are linear combinations of the original features. The first principal component (Principal Component 1) captures the

	largest amount of variance in the data, and the second principal component (Principal Component 2) captures the remaining variance. By reducing the dimensions to two principal components, we were able to visualize the customer segments in a 2D scatter plot, as shown in the last plot. The scatter plot uses these two principal components as the x and y axes.
Link of student Github profile where lab assignment has been uploaded	https://github.com/Sohampimpalgaonkar/DL
Conclusion	We have successfully segmented the customers, and the specific conclusions about what those segments represent require further analysis of the cluster characteristics.
Plag Report (Similarity index < 12%)	
Date	06/11/2025