|  | **PES University, Bengaluru**  (Established under Karnataka Act No. 16 of 2013) | | **UE20CS932** |
| --- | --- | --- | --- |
| **October 2024: END SEMESTER ASSESSMENT (ESA)**  **M TECH DATA SCIENCE AND MACHINE LEARNING\_ SEMESTER II**  **UE20CS932 - MACHINE LEARNING - III** | | | |
| Time: 3 Hrs | | Answer All Questions | Max Marks: 100 |
| **Instructions**   1. Answer all the questions. 2. Section A should be handwritten in the answer script provided. 3. Section B and C are coding questions which have to be answered in the system and uploaded. 4. Smartly use GridSearchCV as it might affect the system performance. | | | |

| **Section A (20 marks)** | | | |
| --- | --- | --- | --- |
| 1 | a) | What is the density-based clustering algorithm? Explain its key concepts. | 5 |
| b) | Explain Principal Component Analysis (PCA). How does PCA work, and what are its key applications? | 5 |
| c) | Describe a content-based recommendation system. How does it work, and what are its main advantages and limitations? | 5 |
| d) | What is Market Basket Analysis, how is it used to uncover relationships between items in transactional data? Explain the key concepts and techniques involved. | 5 |
|  | | | |
| **Section B (40 Marks)** | | | |
| 2 |  | Note:   * Use **penguins.csv** for all the clustering and dimensionality reduction questions. * Use **store\_data.csv** for Association rule mining questions. * Use **Hotel\_Recommender.csv** for recommendation system questions. |  |
| (a) | Perform EDA and pre-processing techniques required for PCA and clustering. (10 marks) Print the top 5 Eigenvalues and Eigenvectors. (4 marks) | 14 |
| (b) | Calculate the optimal number of clusters for K-means, using the principal components that represent 90% of the explained variance.(use SilhouetteVisualizer). | 6 |
| (c) | Create dendrograms (top 100 cluster) using five different linkage methods and compare their performance, utilizing the principal components that explain 90% of the variance and calculate the cophenetic correlation coefficient for each linkage, displaying the results. | 6 |
| (d) | Group the data into the optimal number of clusters and rank the clusters based on their quality, using the within-cluster sum of squares (WCSS) for each cluster. | 6 |
|  | (e) | Cluster the data using both K-Means and Agglomerative methods without applying PCA, and analyze the differences in their performance. | 8 |
| **Section C (40 marks)** | | | |
| 3 | (a) | Read the 'store\_data.csv' file, perform required pre-processing, Build the Apriori ML model with minimum support 10% and print values of Lift greater than 2. | 20 |
| (b) | Build a collaborative recommendation engine using SVD to recommend a top hotel to the specific user. Measure the model quality in terms of RMSE. Use the dataset: Hotel\_Recommender.csv | 20 |