CSE 432
Machine Learning Lab
Lab Final

Rules and guidelines:
1. You are allowed to use internet, but are not allowed to share your code or communicate during the test. Sharing codes will result in a zero score.
2. Name your colab notebook with your id. For example, if your id is 10023, then the name of the colab file will be 10023.ipynb.
3. Share notebook with permission or upload the file in the assignment titled Final Lab Test in your classroom.

Dataset
In the shared folder, there is another folder named datasets. There are 10 datasets in csv format inside the folder. Choose the dataset based on the last digit of your id. For example, if your id is 10023, then you must work on the dataset titled dataset_03.csv.

Tasks:

Task 1: Data Visualization

(1.1)   Load the dataset and display the first five rows.
(1.2)   Visualize the distribution of the age column using a histogram.
(1.3)   Create a bar plot showing the count of different class_label values.
(1.4)   Generate a heatmap of the correlation matrix for numerical features.

Task 2: Data Preprocessing

(2.1)   Handle missing values in the dataset using appropriate imputation techniques.
(2.3)   Normalize the numerical features (age, recovery_days, etc.) for consistent scaling.
(2.4)   Create a categorical feature from age named age_bins by dividing it into 5 bins.
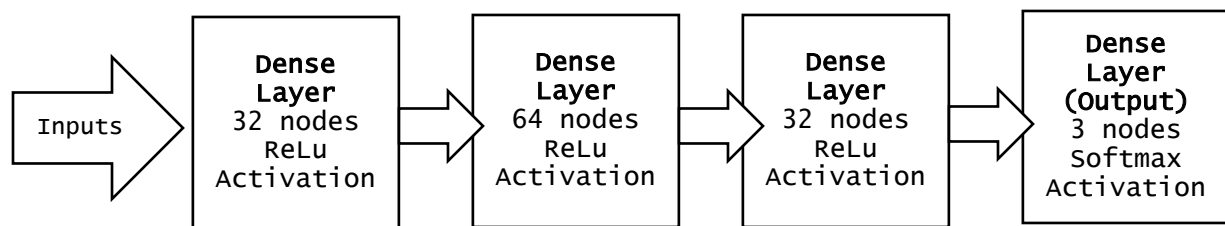
Task 3: Classification

(3.1)   Create a feature vector by taking any 7 features.
(3.2)   Split the dataset into training and test sets (80% training, 20% testing).
(3.3)   Use any 5 classifiers to predict the class_label. Evaluate the model using accuracy, precision, recall.
(3.4)   Create a bar plot showing the accuracy achieved from each model.

Task 4: Clustering

(4.1)   Perform clustering on the dataset (excluding class_label) using k-means clustering.

(4.2)   Visualize the clusters using a scatter plot of two selected numerical features.

Task 5: Neural Network

(5.1) Perform one hot encoding on each of the categorical attributes and the <u>class_label</u>.

(5.2) Create the feature vector by taking the numerical columns and the one hot encoded columns.

(5.2) Build a neural network model based on the following figure to classify the class_label.

```
             ┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
             │    Dense     │   │    Dense     │   │    Dense     │   │    Dense     │
   Inputs    │    Layer     │   │    Layer     │   │    Layer     │   │    Layer     │
  ═════════▷ │   32 nodes   │▷  │   64 nodes   │▷  │   32 nodes   │▷  │   (Output)   │
             │    ReLu      │   │    ReLu      │   │    ReLu      │   │    3 nodes   │
             │  Activation  │   │  Activation  │   │  Activation  │   │   Softmax    │
             │              │   │              │   │              │   │  Activation  │
             └──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘
```

(5.3) Train the neural network on the training set and evaluate its performance on the test set using accuracy and loss metrics.