**NATIONAL UNIVERSITY OF MODERN LANGUAGES**

**ISLAMABAD**



Artificial Nural Network(Assignment 1)

**Assignment: 01**

**Submitted to**

Ms. Hina Ali

**Submitted By**

Adeel Naeem

(BSAI-146)

**Submission Date:** octuber 3rd,2024

**Dataset : Heart Failure Clinical Records**

**Basic Information:**

The Heart Failure Clinical Records dataset contains medical records of 299 patients with heart failure, capturing 13 key features such as age, sex, ejection fraction, serum creatinine, and blood pressure. These features are used to predict mortality during follow-up, making the dataset useful for survival analysis and medical decision-making. It is widely used for research on heart failure prediction and machine learning model development.

**Source:** [**https://archive.ics.uci.edu/datasets**](https://archive.ics.uci.edu/datasets)

**Information:**

The dataset consists of 299 patient records, each containing 13 columns of clinical features. These columns include continuous variables like age, creatinine levels, and platelets (in float64 format), and categorical/binary variables such as anaemia, diabetes, high blood pressure, and DEATH\_EVENT (in int64 format). All columns have complete data with no missing values. The dataset is primarily used for analyzing factors influencing heart failure outcomes.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*information of dataset\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 299 entries, 0 to 298

Data columns (total 13 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 age 299 non-null float64

1 anaemia 299 non-null int64

2 creatinine\_phosphokinase 299 non-null int64

3 diabetes 299 non-null int64

4 ejection\_fraction 299 non-null int64

5 high\_blood\_pressure 299 non-null int64

6 platelets 299 non-null float64

7 serum\_creatinine 299 non-null float64

8 serum\_sodium 299 non-null int64

9 sex 299 non-null int64

10 smoking 299 non-null int64

11 time 299 non-null int64

12 DEATH\_EVENT 299 non-null int64

dtypes: float64(3), int64(10)

memory usage: 30.5 KB

**Shape:**

The dataset has 299 rows and 13 columns, representing 299 patients with 13 clinical features each. These features include both continuous and categorical variables related to heart failure.

\*\*\*\*\*\*\*\*\*\*\*\*shape of dataset\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

(299, 13)

**Columns Names list:**

The dataset contains 13 features, including:

\*\*\*\*\*\*\*\*\*\*\*\*\*\* columns name list\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

['age',

'anaemia',

'creatinine\_phosphokinase',

'diabetes',

'ejection\_fraction',

'high\_blood\_pressure',

'platelets',

'serum\_creatinine',

'serum\_sodium',

'sex',

'smoking',

'time',

'DEATH\_EVENT']

**Data Type:**

The dataset includes health metrics such as \*\*age (float64)\*\*, indicators for conditions like \*\*anaemia (int64)\*\* and \*\*diabetes (int64)\*\*, enzyme levels (\*\*creatinine\_phosphokinase (int64)\*\*), heart function (\*\*ejection\_fraction (int64)\*\*), and outcomes like \*\*DEATH\_EVENT (int64)\*\*.

\*\*\*\*\*\*\*\*\*\*\*\*Data type\*\*\*\*\*\*\*\*\*\*\*\*\*\*

age float64

anaemia int64

creatinine\_phosphokinase int64

diabetes int64

ejection\_fraction int64

high\_blood\_pressure int64

platelets float64

serum\_creatinine float64

serum\_sodium int64

sex int64

smoking int64

time int64

DEATH\_EVENT int64

dtype: object

**Unique Columns(categorical):**

All the columns in the heart failure clinical records data set are numeric .

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* unique values for each catagorical column\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**4.List all necessary preprocessing techniques required to prepare the data for embedding into a neural network.**

 **Normalization/Standardization**: Scale features so that they are within a certain range (e.g., between 0 and 1) or have a mean of 0 and standard deviation of 1.

 **Handling Missing Data**: Impute missing values with mean, median, or a placeholder value.

 **One-hot Encoding**: Convert categorical variables into numerical form using one-hot encoding.

 **Feature Selection/Engineering**: Create new features or remove irrelevant ones.

 **Dimensionality Reduction**: Techniques like PCA or t-SNE can help in reducing the dimensions of the data if it's too high-dimensional.

**5. Identify which preprocessing techniques will be applied to your dataset and explain your reasoning.**

**1. Handling Missing Data:** Impute missing values with mean, median, etc.

**2. Normalization/Standardization:** Scale numerical features for uniformity.

**3. One-Hot Encoding:** Convert categorical variables into binary vectors.

**4. Feature Scaling:** Ensure all variables are on a similar range.

**5. Shuffling and Splitting:** Randomize and split data for training/validation/testing.

**6. Balancing the Dataset:** Address class imbalance to avoid model bias.