Objective:

1. Make a proper decision tree for classifying the data to give status as “Alive” or “Dead”.
2. Fit the data to your decision tree model and show the validation loss and accuracy plot varying the epoch (iteration) value from 1 to 20.

Dataset:

This dataset of breast cancer patients was sourced from the SEER Program's November 2017 update, which offers population-based cancer statistics. It includes data on female patients diagnosed between 2006 and 2010 with infiltrating duct and lobular carcinoma of the breast (histology codes 8522/3 according to SEER primary site recode NOS).

Input variables with their descriptions:

1. **Age**
   Age of the patient.
2. **Race**
   Patient's race. Other categories: American Indian/AK Native, Asian/Pacific Islander.
3. **Marital Status**
   Marital status of the patient.
4. **T Stage**
   Adjusted AJCC 6th T stage classification for the tumor.
5. **N Stage**
   Adjusted AJCC 6th N stage classification for regional lymph nodes.
6. **6th Stage**
   Breast cancer adjusted AJCC 6th stage classification.
7. **Differentiate**
   The degree of differentiation of the tumor.
8. **Grade**
   The grade of the tumor, indicating its aggressiveness.
9. **A Stage**
   - Regional: Neoplasm that has extended.
   - Distant: Neoplasm that has spread to remote parts of the body either through direct extension or metastasis.
10. **Tumor Size**
    Exact size of the tumor, measured in millimeters.
11. **Estrogen Status**
    Status of estrogen receptor in the tumor.
12. **Progesterone Status**
    Status of progesterone receptor in the tumor.
13. **Regional Node Examined**
    Number of regional lymph nodes examined.
14. **Regional Node Positive**
    Number of regional lymph nodes positive for cancer.
15. **Survival Months**
    Number of months the patient survived after diagnosis.

Output variable: Status: Dead/Alive

Method:

1. For making the decision tree apply “gini impurity” approach.