

### 5.4 Exercise do it yourself:

*Download Source Data Set from GitHub link* : <https://github.com/Smartbrain2024/Mastering_AI_2.git>

**Chapters/Chp\_05/5.4/DIY\_Exercise/breast-cancer-data.csv**

###### Cancer Prediction

Context:

This breast cancer dataset was provided by M. Zwitter & M. Soklic, physicians at Institute of Oncology, University Medical Center, Ljubljana, former Yugoslavia.

In the dataset, each entry represents a patient’s medical record. Each patient is classiﬁed into no-recurrence or recurrence of cancer.

Problem Statement:

The objective of this case study is to build a classiﬁcation model that can predict the likelihood of recurrence of breast cancer given some input features like age, menopause status, tumor scan attributes, etc. and identify factors that affect the likelihood of recurrence of breast cancer and help in early diagnosis.

Approaching the Problem :

1. Understand the problem statement and attributes of the dataset.
2. Gather insights from the data by performing Univariate & Bivariate analysis.
3. Preprocess the data and make it ﬁt for model building.
4. Build a Logistic Regression model to predict the probability of recurrence of breast cancer in a patient.
5. Choose the appropriate evaluation metric(s) and observe the performance of the model on Train Data and Test Data.
6. Tune the thresholds and observe the change in model performance.