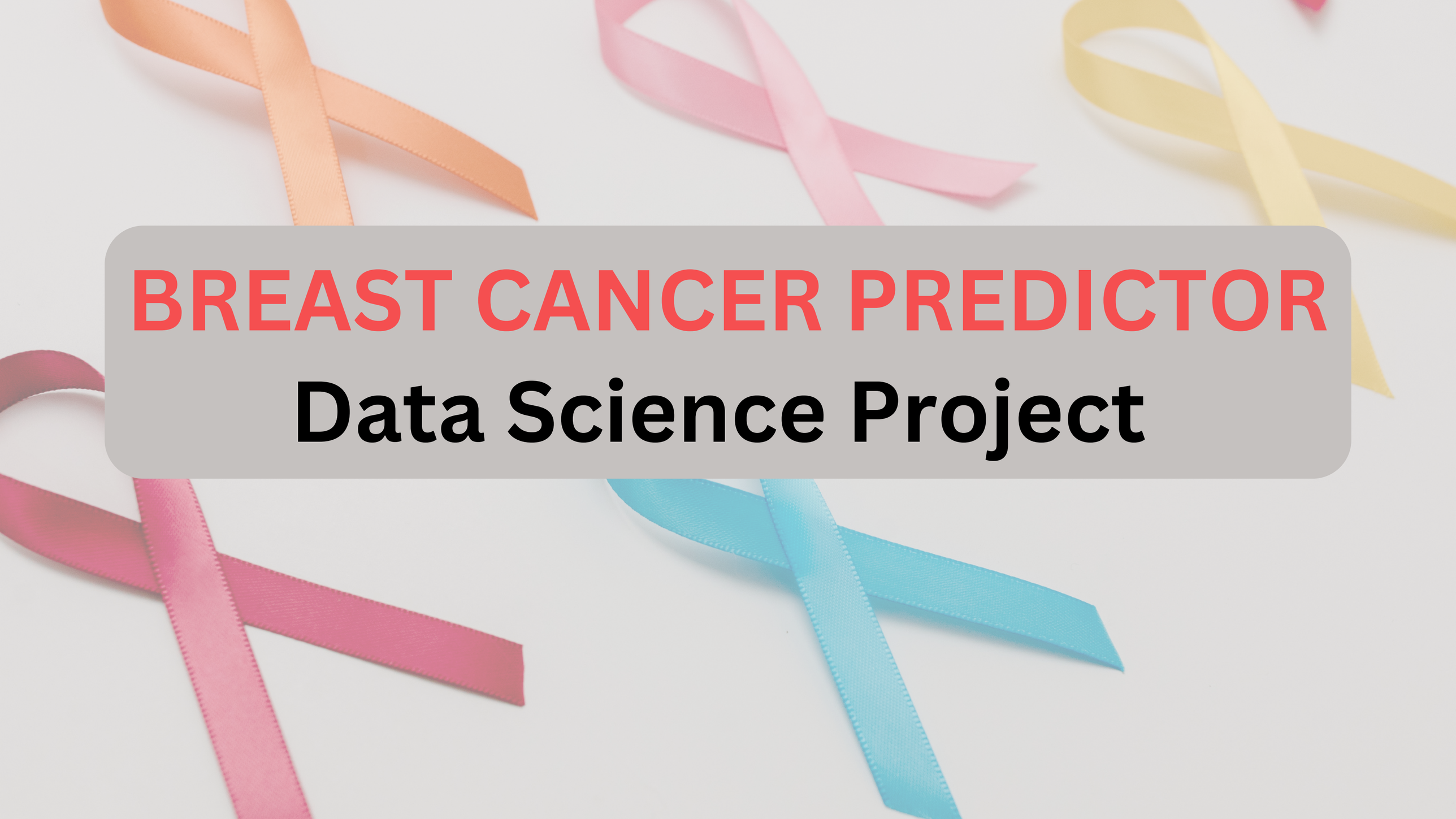
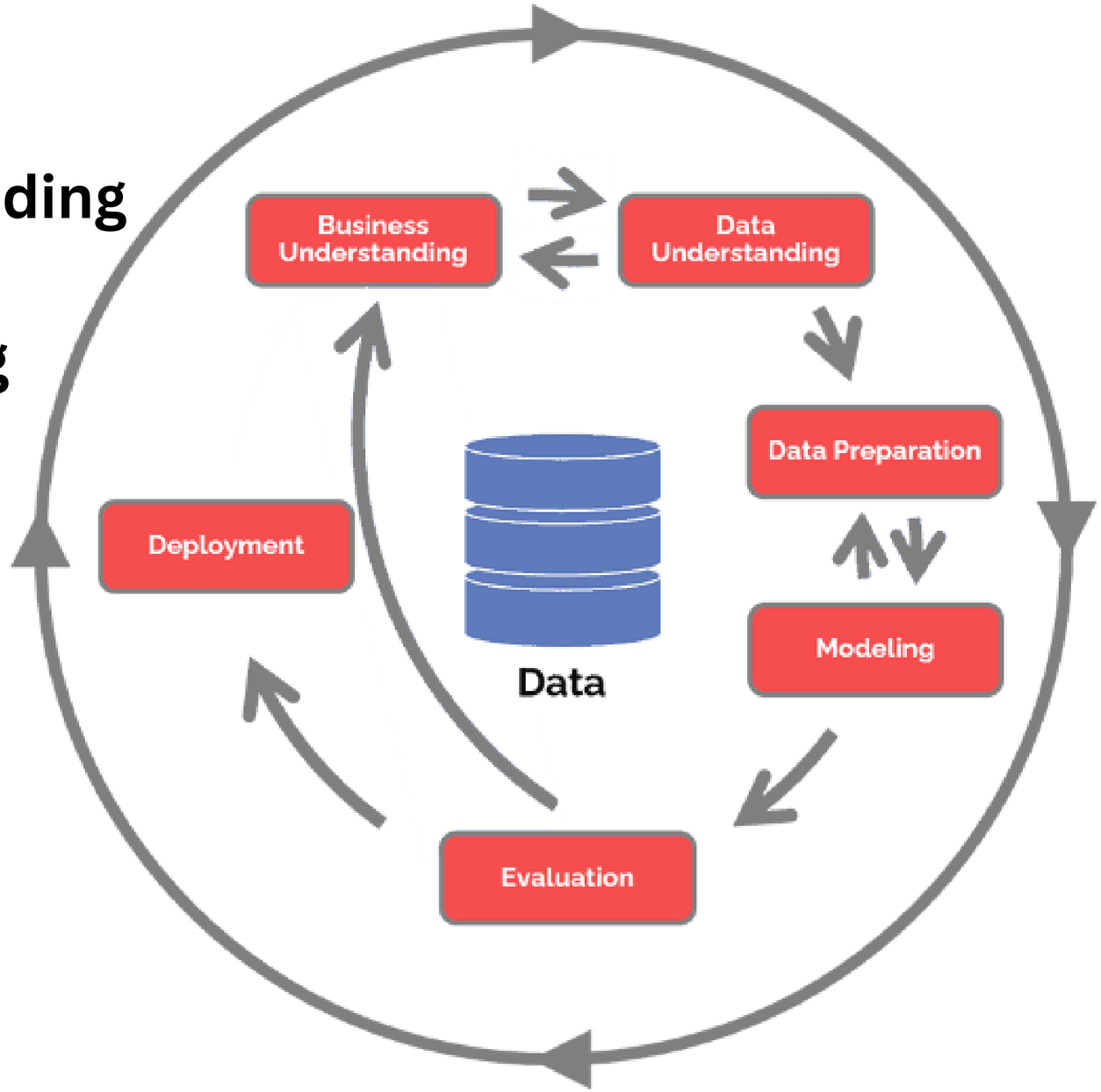

BREAST CANCER PREDICTOR
Data Science Project

Agenda

- Data Science Lifecycle
- Project Overview
- Data
- Analysis
- Modeling
- Model Evaluation
- Summary

Data Science Lifecycle

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment.



Project Overview

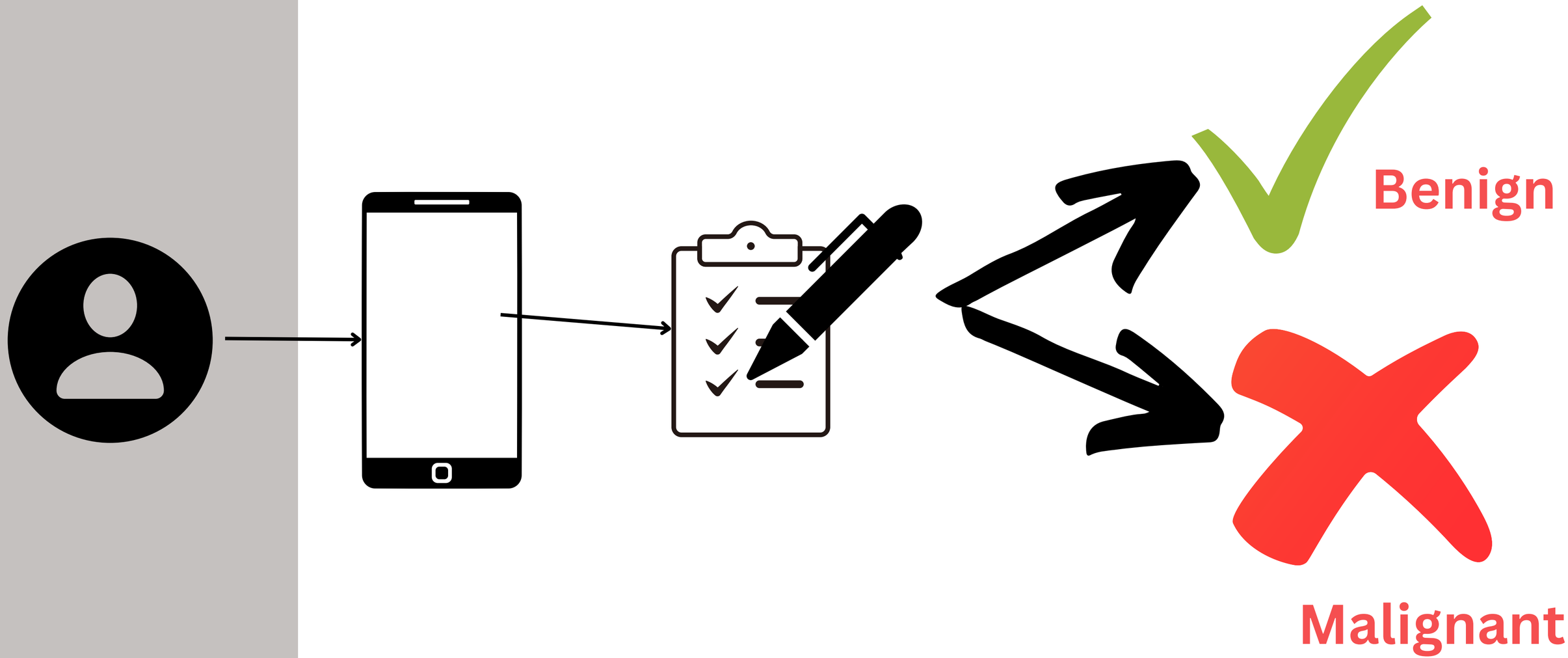
Business Objective:

Make a model to predict either the person have a Breast cancer or not immideately after filling a details.

Hypothesis:

we will use the historical data of the **Cancer Data** and make a model to predict the Breast Cancer.

Process Overview / Solution

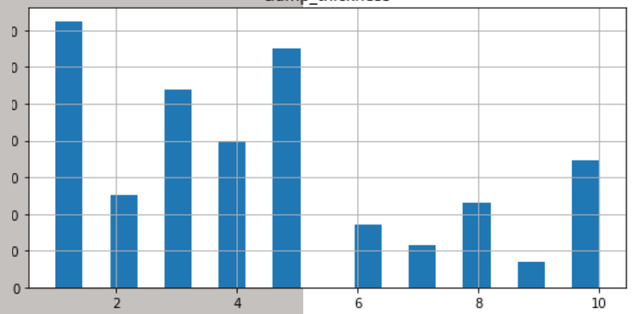


Data

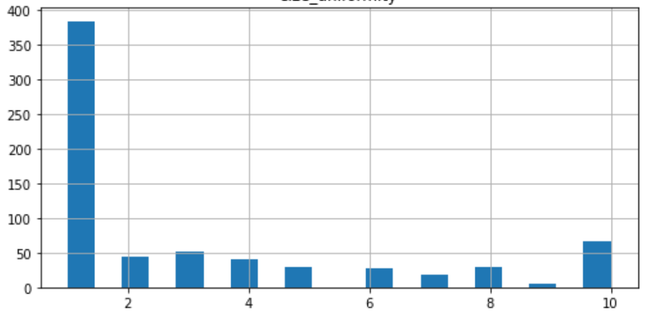
- The number of records – **699**
- The number of columns – **10**
- The number of numerical columns – **10**
- Target/Loan Status – **Y (458)** vs **N (241)**

Analysis

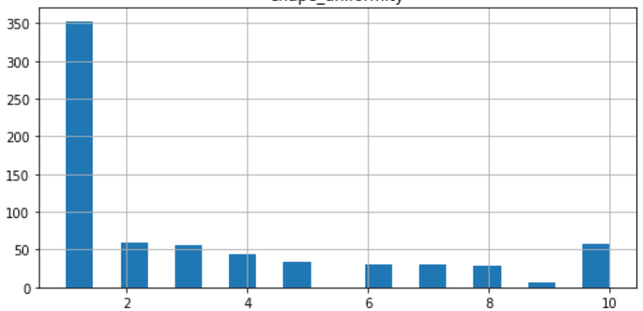
clump_thickness



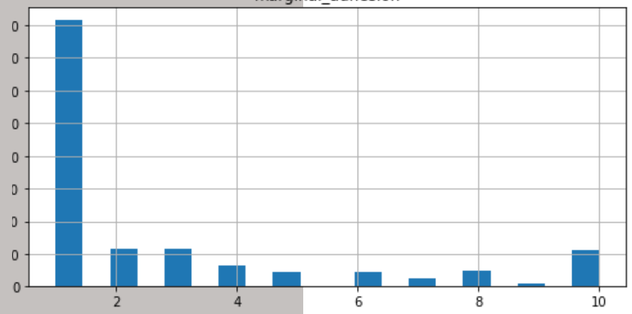
size_uniformity



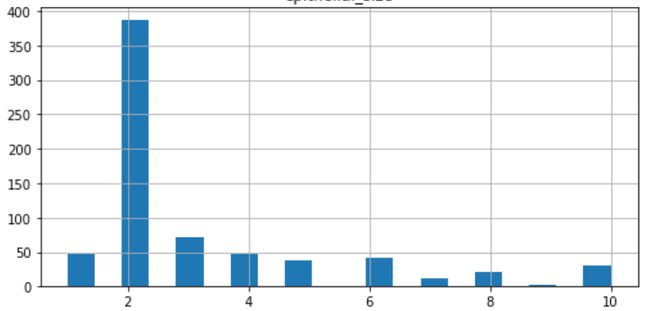
shape_uniformity



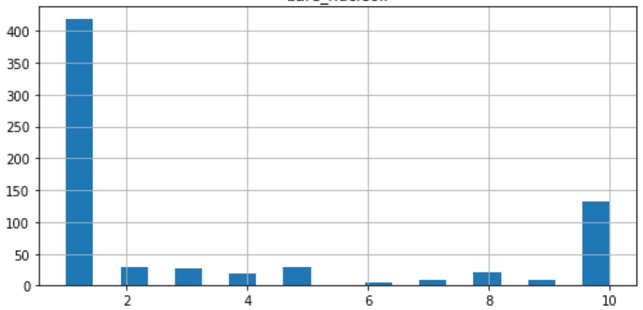
marginal_adhesion



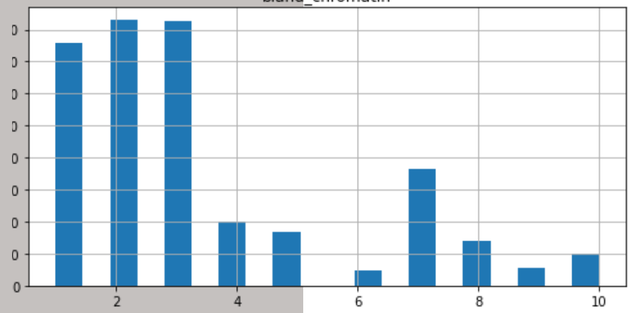
epithelial_size



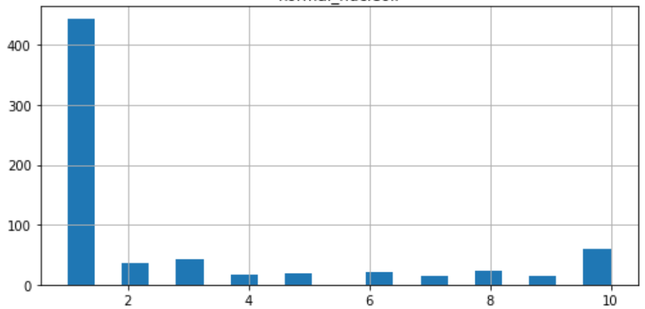
bare_nucleoli



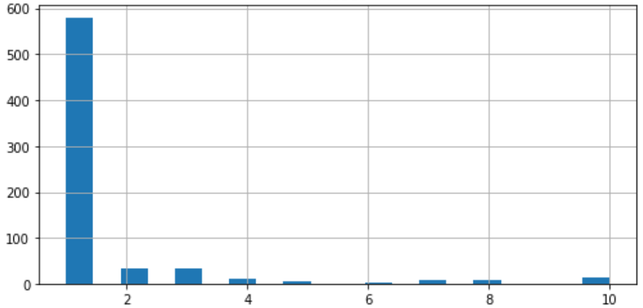
bland_chromatin



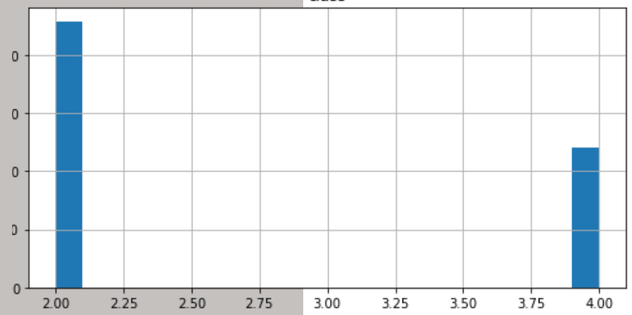
normal_nucleoli



mitoses



class



Modeling

Here we can try to make a model having high accuracy.

So,

We trained Machine learning Model through **SVC** and second time **KNN CLASSIFICATION** is used.

Model Evaluation

SVC classification report

---

	precision	recall
2	0.99	0.99
4	0.97	0.97

98%

knn classification report

---

	precision	recall
2	0.97	0.99
4	0.99	0.95

97%

Summary

- Based on this broad analysis of SVC and KNN classifiers, both works good on the test data with almost same accuracy.
- More test data will be required to validate the robustness of the model.