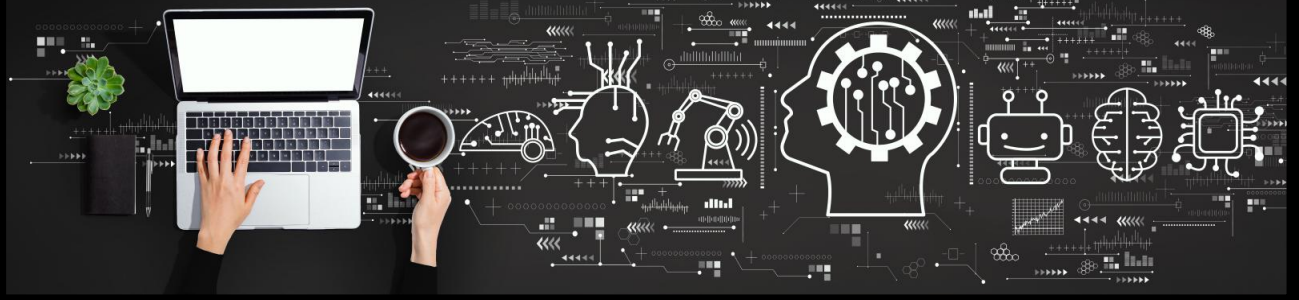
BREAST CANCER MACHINE LEARNING CAPSTONE PROJECT REPORT

BY: UTPAL MISHRA



INTRODUCTION/ BUSINESS PROBLEM

BACKGROUND

Breast cancer starts its activity in the breast cells of both men and women. Globally, it ranked as the second most common type of cancer after lung cancer and the fifth most common cause of death due to cancer. The **National Breast Cancer Foundation** has valued around 2,00,000 new breast cancer cases and 40,000 deaths annually in women. In men, these stats are 1700 and 450, correspondingly.

Breast cancer is highly major in today's world. Cancer begins in cells and spreads to other parts of the human being. Additional cell growth develops a bulk of tissue named as a lump. Therefore, early detection of cancer is more important. Mammography is an exclusive screening test for breast cancer detection. According to the National Cancer Institute, an estimated 2,07,090 new cases and 39,840 deaths from breast cancer in only women, are predicted to occur in the United States, despite the recent breakthroughs in treatment. Given such conditions, early diagnosis of breast cancer is considered vivacious because statistics show a five-year survival rate of 96% for those who were diagnosed with cancer at early stages.

In the women's world, according to **Cancer Prevention and Control** (CPC), breast cancer is the second largest cause of death next to lung cancer, but if it is diagnosed at an early stage, it's also one of the curable cancers. This is a tedious and confusing task, thus causing oversight errors that fail to detect cancer.

Among these, cancer is a term for a class of diseases identified by abnormal cells that grow and plunder healthy cells in the body. Breast cancer starts in the cells of the breast as a group of cancer cells that attack the surrounding tissues or spread to other areas of the body. Next to lung cancer, breast cancer is the second leading cause of cancer death among women. It occurs in both genders, though male breast cancer is rare. Recent statistics show that breast cancer is a serious disease with a high incidence rate and one of the leading causes of the early death of women.

PROBLEM

To build an efficient Machine Learning Classification model in order to analyse and detect Breast Cancer on the basis of lump dimensions and distribution available from the WDBC Data.