

**NAME:** VIJAYA DURGA BHAVANI THOTA

**NJIT UCID:** vt299

**Email Address:** [vt299@njit.edu](mailto:vt299@njit.edu)

**Professor:** Yasser Abdullah

**COURSE NAME:** CS634101- Data Mining

## **Final Project Report**

### **Abstract:**

This project focuses on predicting whether a breast tumor is cancerous (malignant) or not (benign) using data from the Breast Cancer Wisconsin dataset. The dataset contains measurements taken from images of breast tissue samples. The project includes exploring and understanding the data, selecting the most useful features, and preparing the data using techniques like standardization and normalization to improve accuracy. Machine learning models are then used to make predictions, aiming to support doctors in diagnosing breast cancer more effectively.

### **Introduction:**

One of the most prevalent and deadly illnesses impacting women globally is breast cancer. Improving treatment results and survival rates depends heavily on early and precise diagnosis. In order to develop a machine learning model that can determine if a breast tumor is benign (non-cancerous) or malignant (cancerous), this study makes use of the Breast Cancer Wisconsin dataset. The dataset contains a variety of attributes that were extracted from pictures of breast tissue fine needle aspirations (FNAs). The project intends to provide a dependable and effective tool to help medical practitioners make accurate diagnoses by utilizing data analysis, feature selection, and preprocessing approaches, followed by machine learning algorithms. The experiment demonstrates how machine learning can improve patient care and medical decision-making

## **Core concepts and principles:**

**Frequent dataset discovery:** A popular dataset in machine learning, especially for binary classification problems, is the Breast Cancer Wisconsin dataset. It includes measurements taken from digital pictures of breast mass fine needle aspirations (FNAs). The dataset contains 30 numerical variables that offer important information on tumor characteristics, including the radius, texture, area, perimeter, and smoothness of the cells. Whether the tumor is benign (non-cancerous) or malignant (cancerous) is indicated by the target variable.

Key features like mean radius, mean texture, and mean perimeter that show a substantial association with the objective variable are frequently found in this dataset. Visualization of exploratory data analysis (EDA), particularly in feature distributions and scatter plots, frequently shows substantial differences between benign and malignant tumors

## **Project Setup:**

1. Clone the github repository from link

<https://github.com/vijaya9490/Breastcancer>

2. Go to project directory

3. Install python libraries by executing pip install

4. Python-3.9 version( For tensorflow)

## **Conclusion:**

In this project, we successfully applied machine learning techniques to the Breast Cancer Wisconsin dataset to predict whether a tumor is malignant or benign. By performing data preprocessing, including standardization and normalization, and selecting key features, we were able to improve the model's performance. Through various machine learning algorithms, we developed a reliable prediction model that can assist healthcare professionals in making more accurate diagnoses of breast cancer. The project demonstrates the potential of machine learning to enhance early cancer detection, contributing to better patient outcomes and more efficient medical practices. Moving forward, further optimization and the use of more advanced models could help increase the accuracy and robustness of predictions.

THANKYOU.....