

INTELLIGENT SYSTEMS AD2022

BREAST CANCER PREDICTION

ANDREA MELISSA RINCON TREJO A01365736

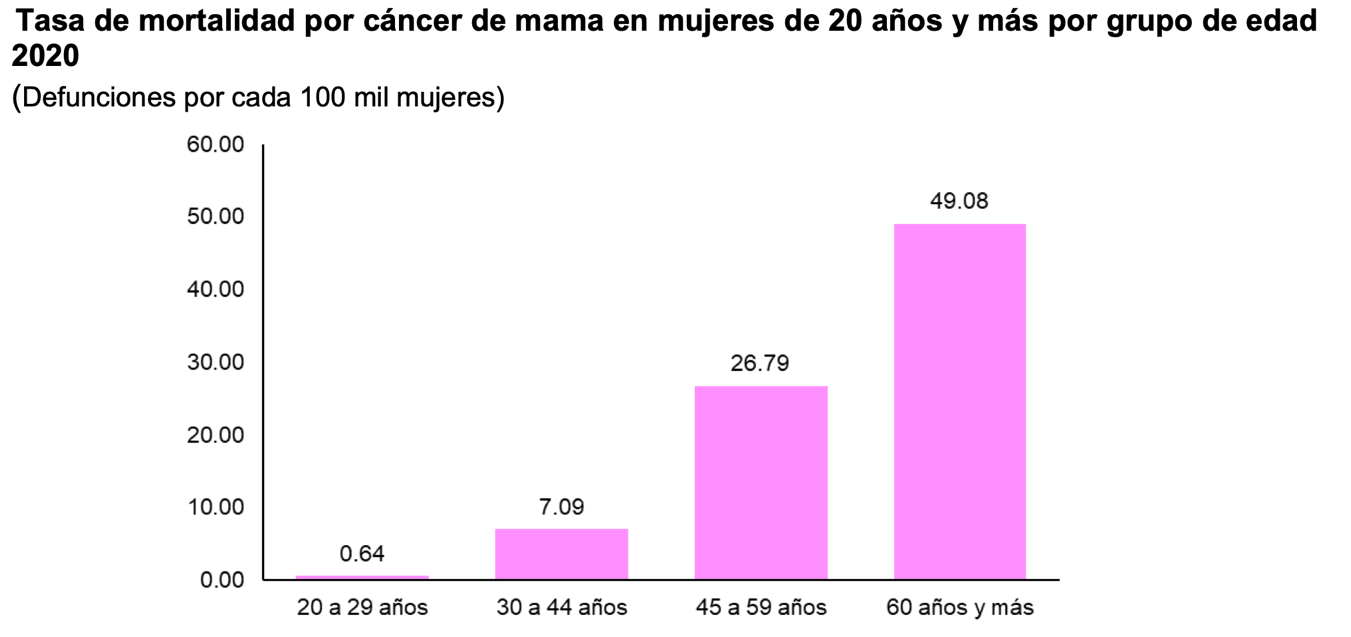
**PURPOSE**.

The main objective of this program is the classification of breast tumors based on several characteristics. The classification will be made to distinguish the tumors between benign and malignant and using a decision tree algorithm.

**PROBLEM**.

MEXICO 2020.

* 97,323 people died from malignant tumours. 7,880 of the deaths were due to malignant breast tumors in specific, which is equivalent to 8% of this total.
* The highest rate of deaths was among women of 60 years or older (49. 08 per 100,000 women in this age group).



WORLD.

Breast cancer originates from the uncontrolled and independent growth of cells that form a malignant tumor that can invade surrounding tissues and lead to metastases.

The World Health Organization WHO recognizes it as the most common type of cancer in the world. [1]

**DATA SET.**

Title: Wisconsin Breast Cancer Database (January 8, 1991) This breast cancer databases was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg.

From: UCI Machine Learning Repository.

The characteristics we are taking as features are.

1. Clump Thickness 1 - 10
2. Uniformity of Cell Size 1 - 10
3. Uniformity of Cell Shape 1 - 10
4. Marginal Adhesion 1 - 10
5. Single Epithelial Cell Size 1 - 10
6. Bare Nuclei 1 - 10
7. Bland Chromatin 1 - 10
8. Normal Nucleoli 1 - 10
9. Mitoses 1 - 10
10. Class: (2 for benign, 4 for malignant)

Number of Instances: 699 (as of 15 July 1992)

Number of Attributes: 10 plus the class attribute

Algorithm: decision trees This is a classification algorithm.

**Decision Trees (DTs)** are a non-parametric supervised learning method used for [classification](https://scikit-learn.org/stable/modules/tree.html#tree-classification) and [regression](https://scikit-learn.org/stable/modules/tree.html#tree-regression). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation. [2]

HOW TO RUN.

You can make a copy of the repository using the command

**git clone**

then you can move to the folder you can run it with the command

**python tree\_cancer.py** (it depends on the version).

In the framework you will be asked to introduce 9 different numbers, one for each characteristic to consider in the calculation.

The calculation is going to be the number of the class **+ --> +** if the sample results **malignant or benign.**

**References.**

<https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2021/EAP_LUCHACANCER2021.pdf> [1]

<https://scikit-learn.org/stable/modules/tree.html> [2]