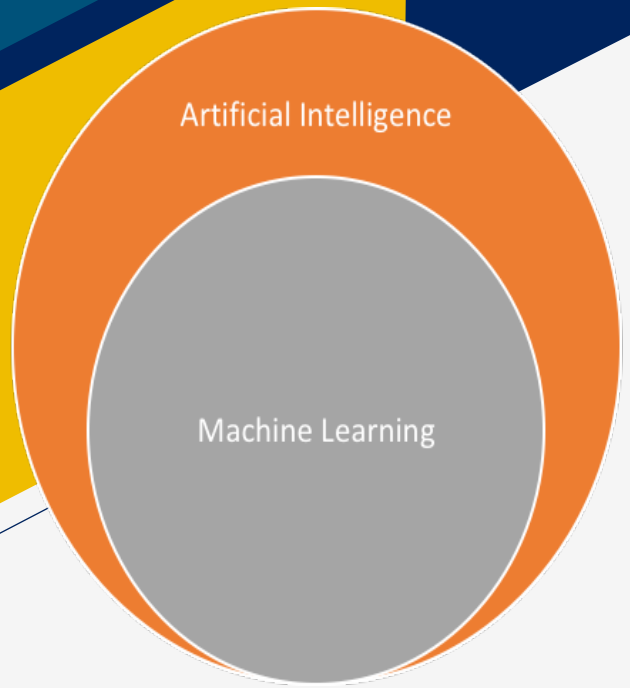
Disease Prediction Applying Machine Learning

By.

Robles Ontiveros Carlos Daniel

AI and Machine Learning

Technique that combines a set of algorithms with the purpose of training machines.



Among the machine learning algorithms it is possible to find:

- *Decision threes*
- *Naive Bayes*
- *Logistic regression*
- *K-means algorithm*
- *Linear discriminant analisis*
- *Support Vector Machines*

Last years

Machine learning has been applied with the objective of predicting scenarios in:

- *health*
- *banking*
- *Financial*
- *Educational*

FR

Our Research

Two Predictive Scenarios

- Breast Cancer
- Diabetes

We worked with two datasets each obtained from open internet sources

BREAST CANCER

- Data: Hospitals of the University of Wisconsin, Madison, of Dr. William H. Wolberg

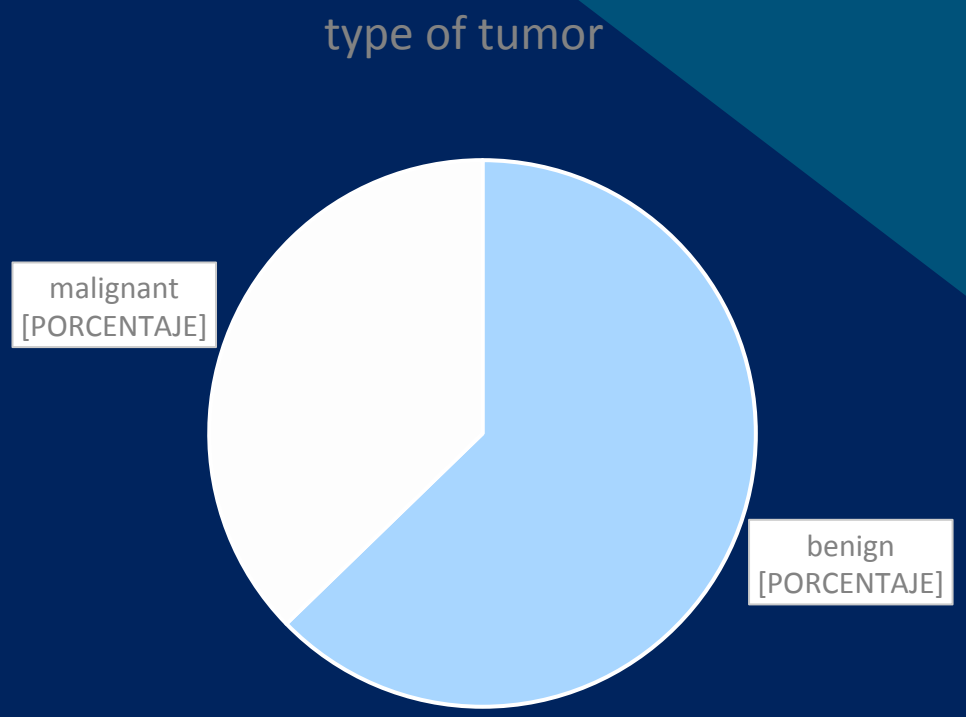
DIABETES

- Data: National Institute of Diabetes and Digestive and Kidney Diseases

Breast Cancer

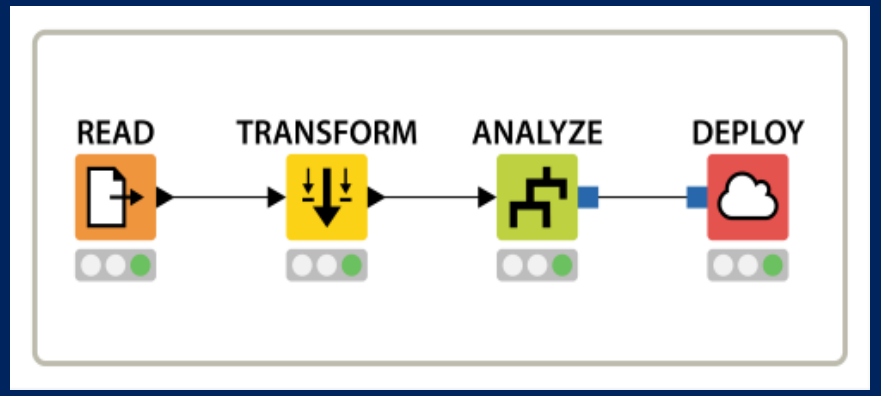
2nd cause of death in a woman!

596 records
of which 212 a malignant tumor was detected



Variables to take into account

- Tumor radius
- Size of the perimeter of the tumor.
- Size of the tumor area.
- Average variation in radio lengths.

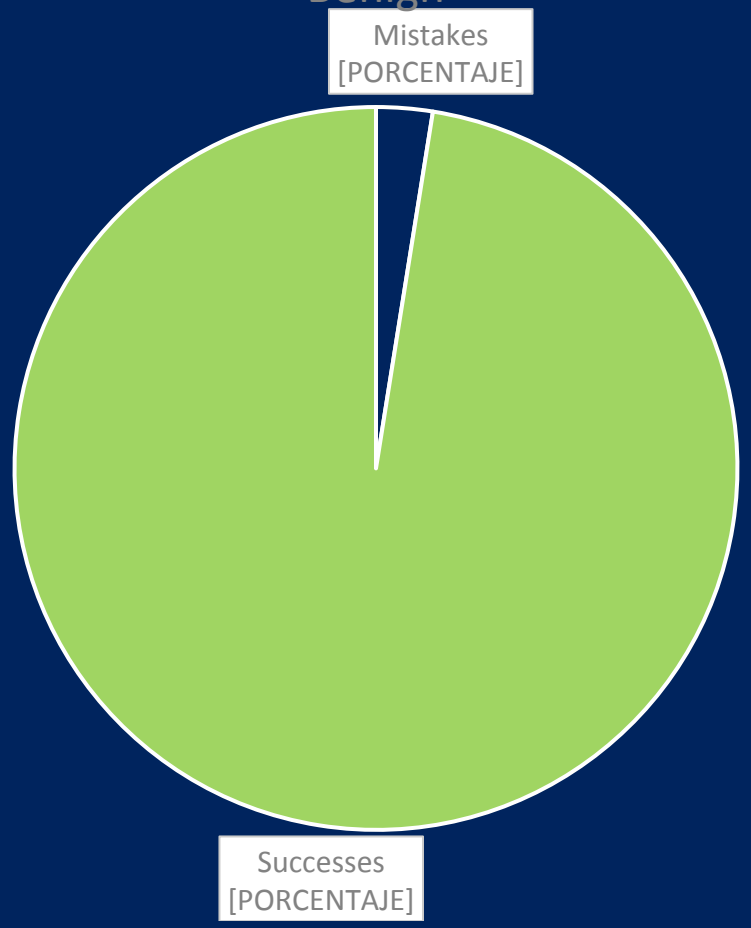


Results

Id	Radius (mm)	Average of the σ of the gray scale.(mm)	Perimeter (mm)	Área (mm^2)	Average radius length variation(mm)	Diagnostic prediction	Diagnostic
1	17.99	10.38	122.8	1001	0.1184	20%	Maligno
2	20.57	17.77	132.9	1326	0.08474	32%	Maligno
3	19.69	21.25	130	1203	0.1096	1%	Maligno
4	11.42	20.38	77.58	386.1	0.1425	38%	Maligno
5	20.29	14.34	135.1	1297	0.1003	20%	Maligno
20	13.54	14.36	87.46	566.3	0.09779	77%	Benigno
21	13.08	15.71	85.63	520	0.1075	67%	Benigno
22	9.504	12.44	60.34	273.9	0.1024	120%	Benigno
49	12.05	14.63	78.04	449.3	0.1031	86%	Benigno
50	13.49	22.3	86.91	561	0.08752	67%	Benigno

Graphics

Benign



Malignant

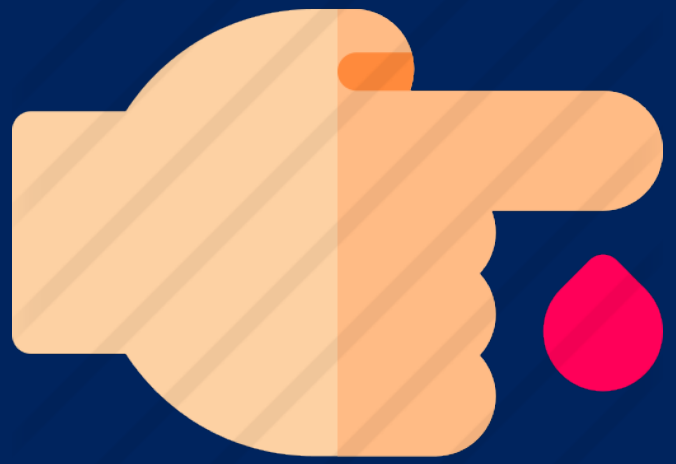
Diabetes

FR

Study population:

768 people: This study was conducted only for women of the original Pima group of at least 21 years.

Variables to take into account



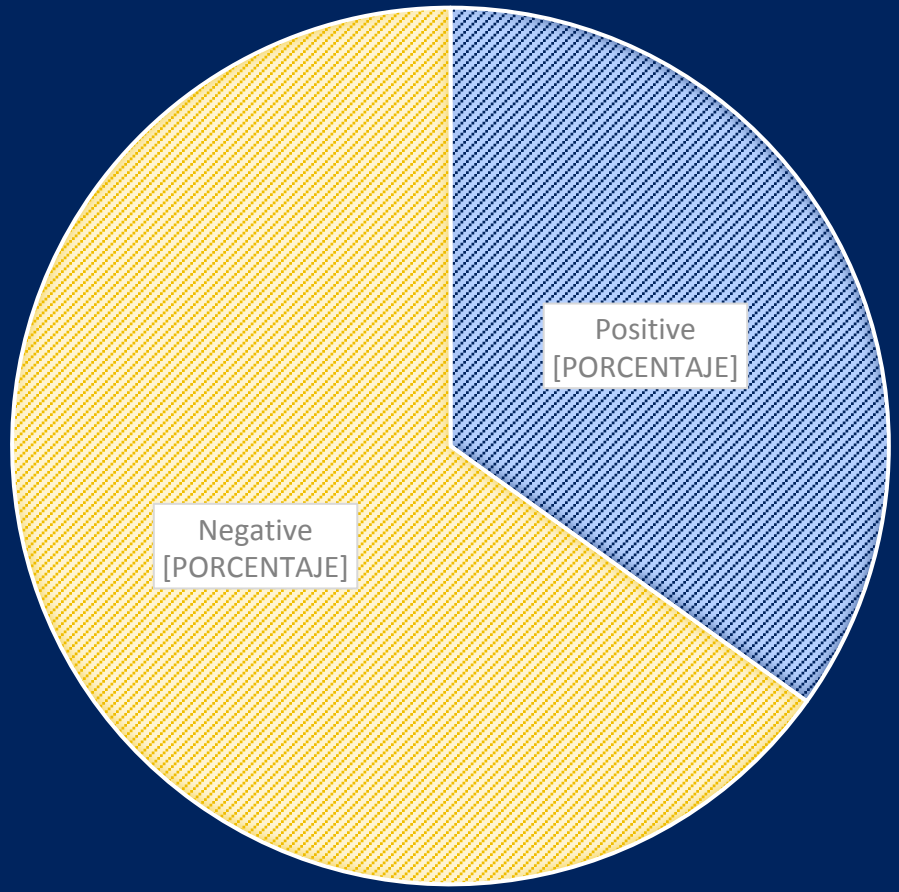
- Number of times pregnant
- Plasma glucose concentration at 2 hours in an oral glucose tolerance test
- Diastolic blood pressure (mm Hg)
- Thickness of the triceps skin fold (mm)
- 2-hour serum insulin (mu U / ml)
- Body mass index
- Pedigree function of diabetes (genetics)
- Age (years)

Results

FR

ID	Pregnancies	Plasma	Blood pressure	Skin thickness	Insulin	Body mass		Age	Prediction	Class
1	6	148	72	35	0	33.6	0.627	50	0.65175729	1
2	1	85	66	29	0	26.6	0.351	31	0.00573265	0
3	8	183	64	0	0	23.3	0.672	32	0.73642449	1
4	1	89	66	23	94	28.1	0.167	21	-0.0219232	0
5	0	137	40	35	168	43.1	2.288	33	0.83318937	1
6	5	116	74	0	0	25.6	0.201	30	0.21054145	0
7	3	78	50	32	88	31	0.248	26	0.05736238	1
8	10	115	0	0	0	35.3	0.134	29	0.59612039	0
9	2	197	70	45	543	30.5	0.158	53	0.66541521	1

Test prediction



Thank you very much!

Robles Ontiveros Carlos Daniel

55 1289 0378

carlosdanielroblesontiveros@Gmail.com

Instituto Politécnico Nacional