

INTELIGENCIA ARTIFICIAL

David Gonzalo Ejea Carbonell

PROYECTO FINAL



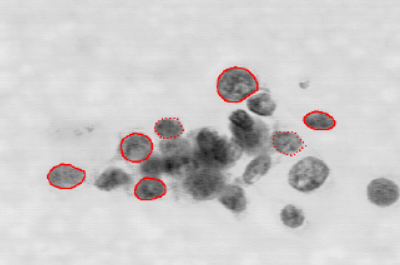
David Gonzalo Ejea Carbonell

RESPONSABLE DEL DESARROLLO DE LA APLICACIÓN

**Fecha:** 14, 02, 2021

**Introducción**

Los datos proceden de un estudio sobre diagnóstico del cáncer de mama por imagen. Mediante una punción con aguja fina se extrae una muestra del tejido sospechoso de la paciente. La muestra se tiñe para resaltar los núcleos de las células y se determinan los límites exactos de los núcleos. Las variables consideradas corresponden a los valores medios de distintos aspectos de la forma de los núcleos de cada muestra

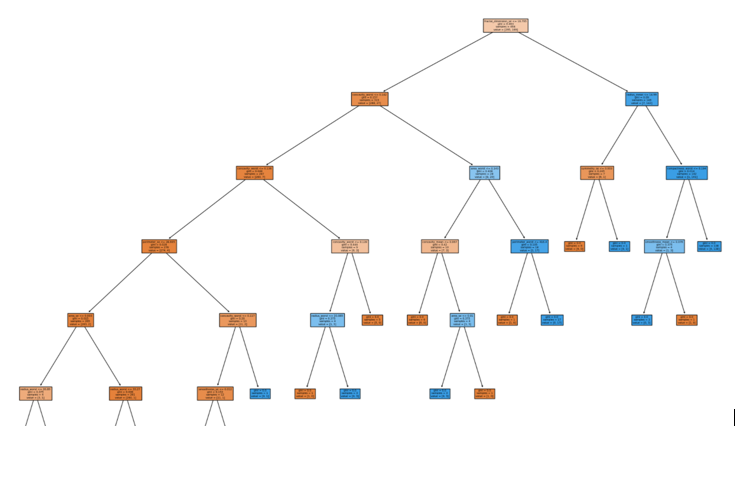


* Las 10 variables explicativas medidas en pacientes cuyos tumores fueron diagnosticados posteriormente.
* La variable tipo que contiene el tipo de tumor (benigno o maligno).

**Objetivo:**

* El objetivo de este trabajo será ajustar un modelo para ver si predecimos el cancer de mama

**Gráfica:**

. 

Lo tenéis mejor en el programa de python que os paso.

**Conclusión:**

El mejor modelo que se puede utilizar para diagnosticar el cáncer de mama, tal como se encuentra en este análisis, es el modelo Random Forest o Decision tree . Con los 5 predictores principales,'concave points\_mean', 'area\_mean', 'radius\_mean', 'perimeter\_mean', 'concavity\_mean'. Proporciona una precisión de predicción del 100%, lo que me preocupa de esto es que se de el overfitting, es decir que el modelo solo sirva para estos datos y que no se pueda aplicar a otro conjunto, sería bueno tener otro conjunto para probarlo.

Nuestro proyecto seria de Machine Learning utilizo algoritmos para analizar datos, aprender de esos datos y tomar decisiones basadas en lo aprendido.

En definitiva, he sacado unas predicciones adecuadas, aunque podría ser mejor si tuviera un conjunto diferente para comprobar.