



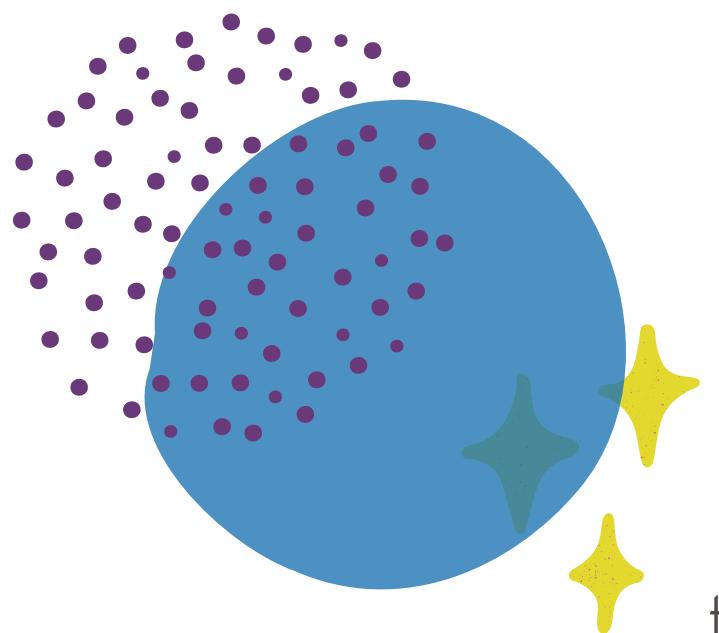
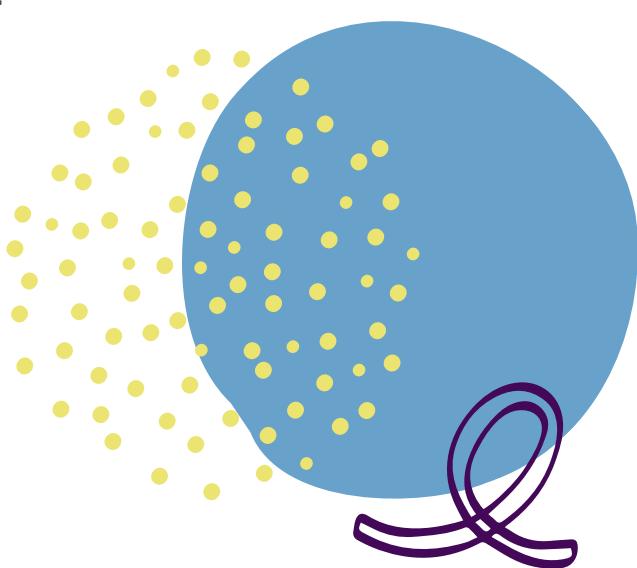
Clasificación multiclasse Datos dermatológicos

Proyecto Ciencia de Datos 2023
Mariel Palacio

Introducción

El diagnóstico diferencial de las 'enfermedades eritematoescamosas' es un verdadero **problema en dermatología**, ya que para su diagnóstico se tiene en cuenta **características clínicas** con diferencias mínimas. Por lo general, se necesita una biopsia para el diagnóstico, pero desafortunadamente, estas enfermedades también comparten muchas **características histopatológicas**.

Por esto, buscaremos determinar el tipo de 'Enfermedad Eritematoescamosa' en función de las **características clínicas e histopatológicas**.





Sobre las enfermedades

Las dermatosis eritematoescamosas corresponden a enfermedades inflamatorias de la piel caracterizadas por presentar placas erítomo descamativas.

Hay multiples tipos y sus nombres son: **psoriasis, dermatitis seborreica, liquen plano, pitiriasis rosada, dermatitis crónica, pitiriasis rubra pilaris.**





Objetivo y sobre el dataset



Objetivo

Clasificar las enfermedades :

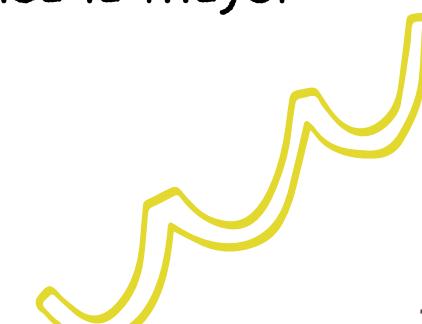
1. Psoriasis
2. Dermatitis seborreica
3. Líquen plano
4. Pitiriasis rosada
5. Dermatitis crónica
6. Pitiriasis rubra pilaris

En función de las características clínicas e histopatológicas

Dataset

El dominio esta construido con:

- La **características de historia familiar**: binario, si se ha observado estas enfermedades en la familia.
- La **características de edad**, representa la edad del paciente.
- Las **otras características clínicas e histopatológicas** recibió un valor en el **rango de 0 a 3**. El 0 indica que la característica no estaba presente, 3 indica la mayor cantidad posible y 1, 2 indican los valores intermedios relativos.

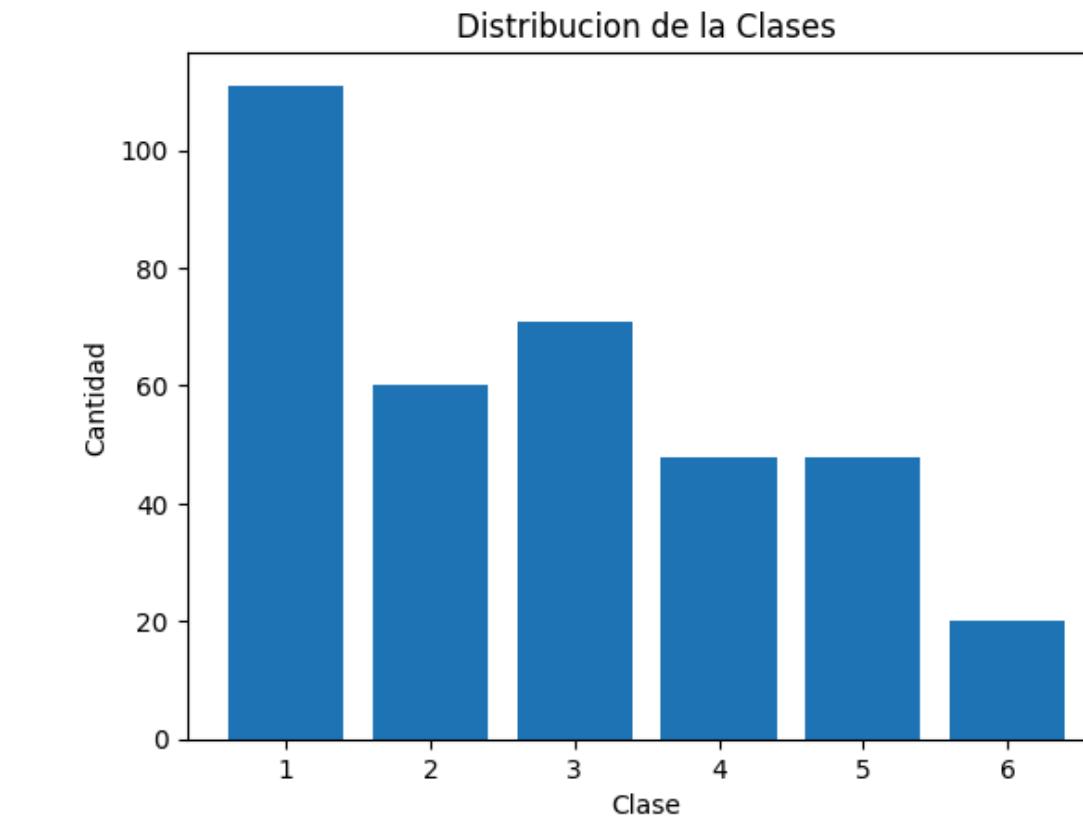
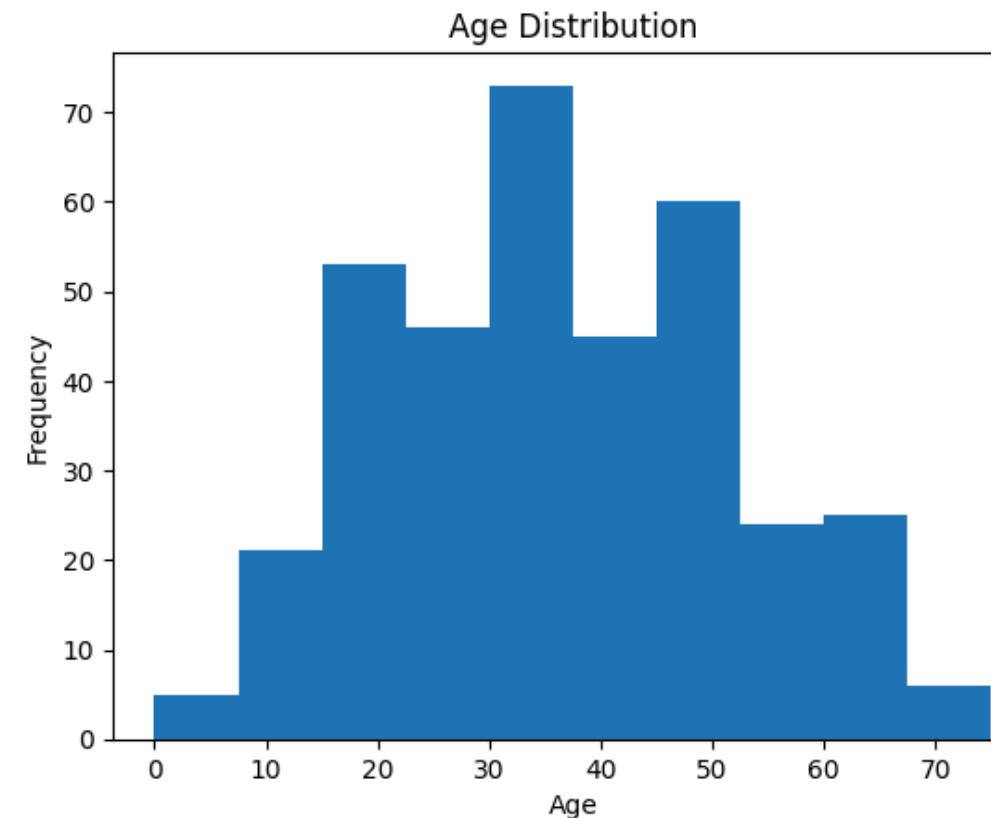


Esta base de datos contiene 34 atributos y el atributo de clase es la última columna.

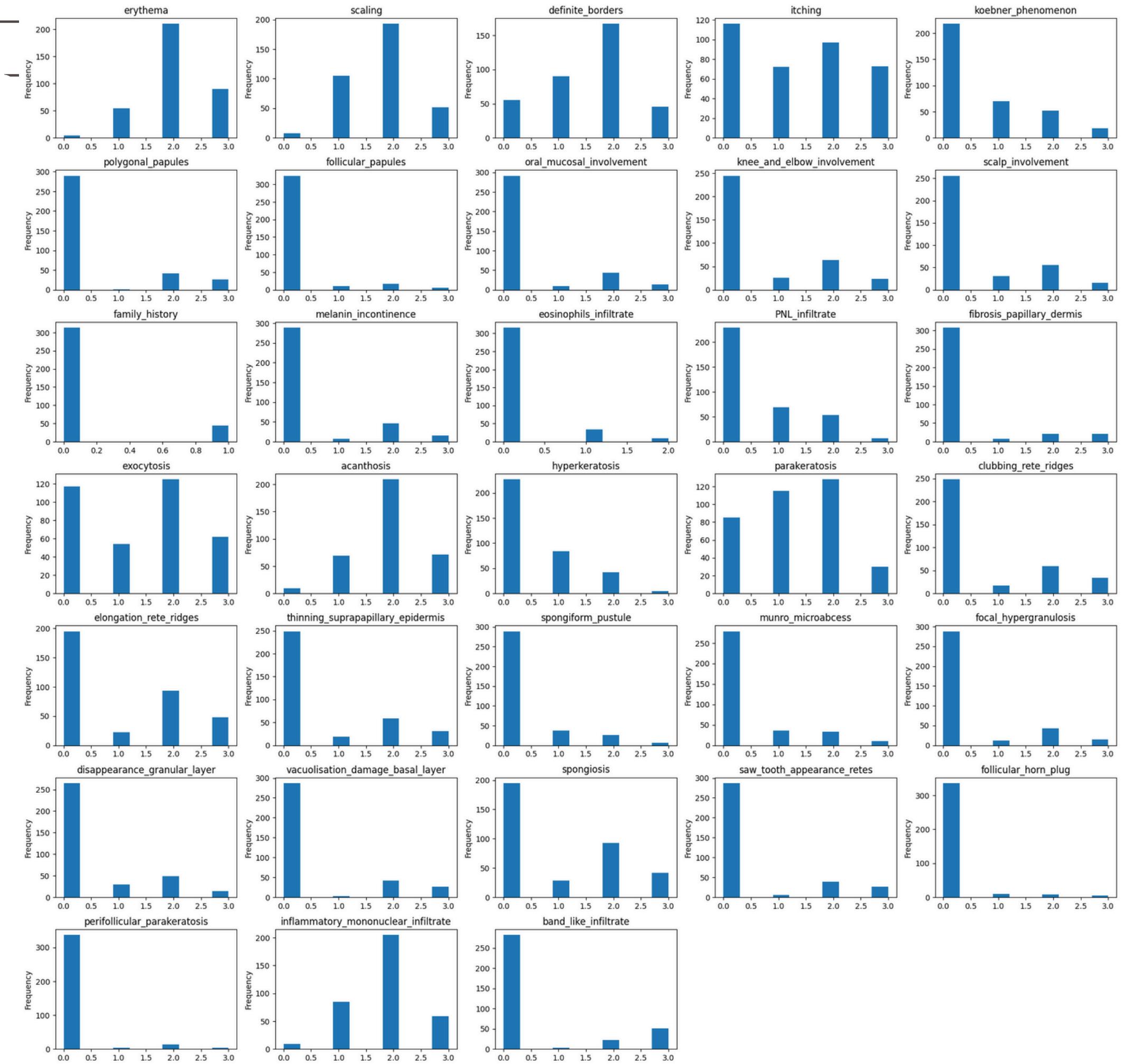
Ademas presenta 366 instancias: 366 con 8 valores de atributos faltantes (en el atributo Edad) distinguido con '?'.

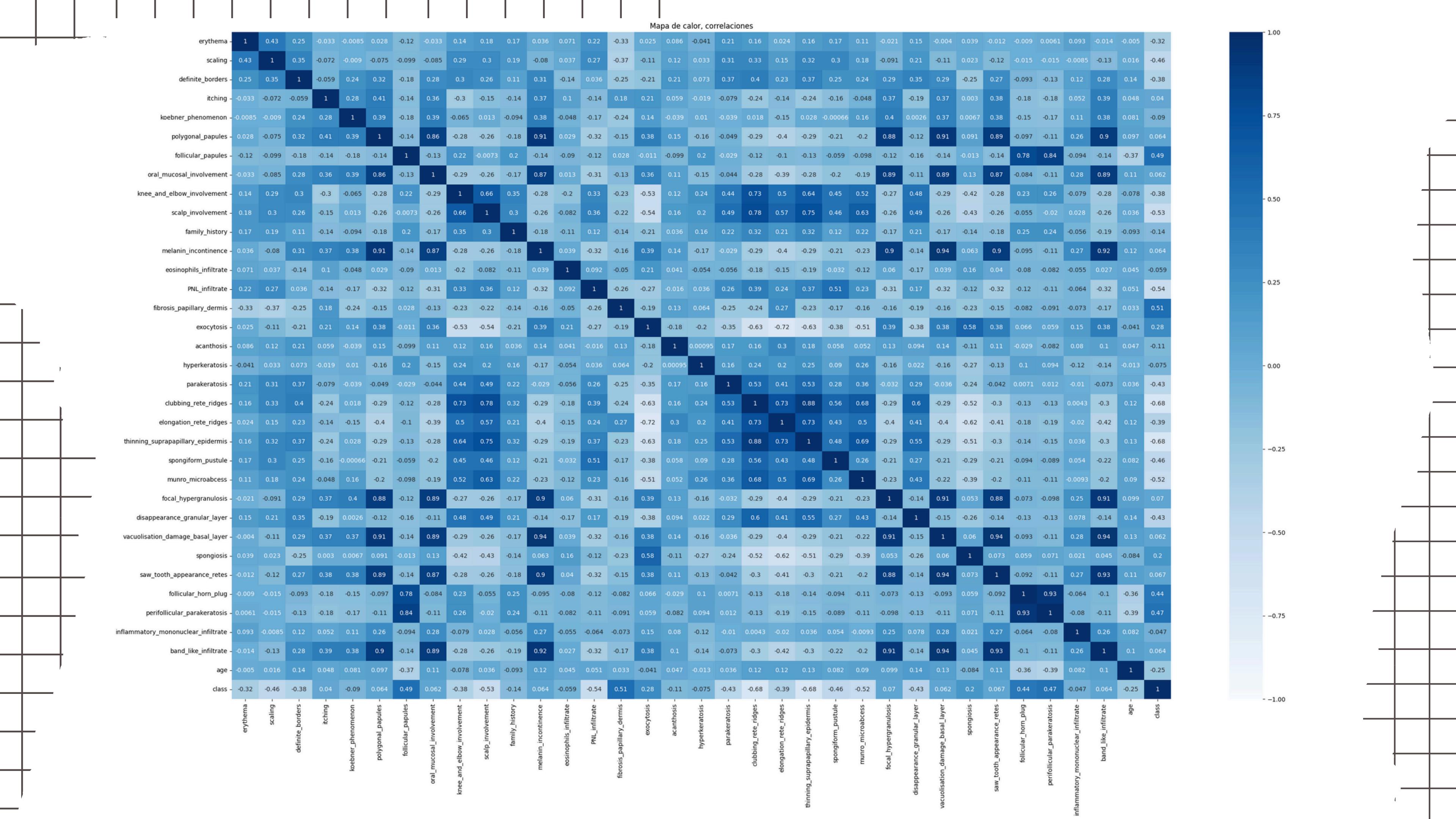
Analisis Exploratorio

Veamos un poco los datos



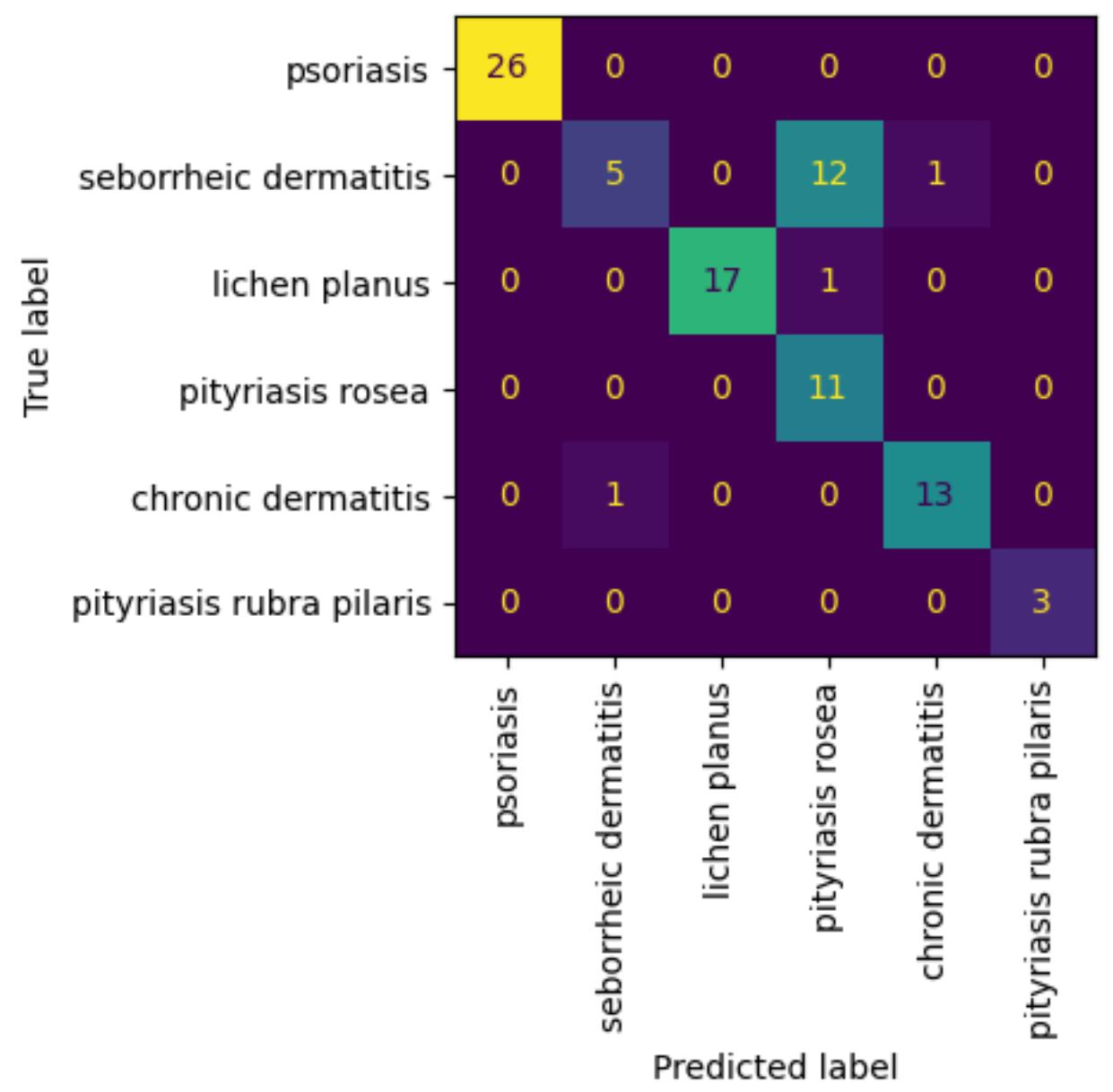
Histogramas de las Características vs Gravedad de la Característica



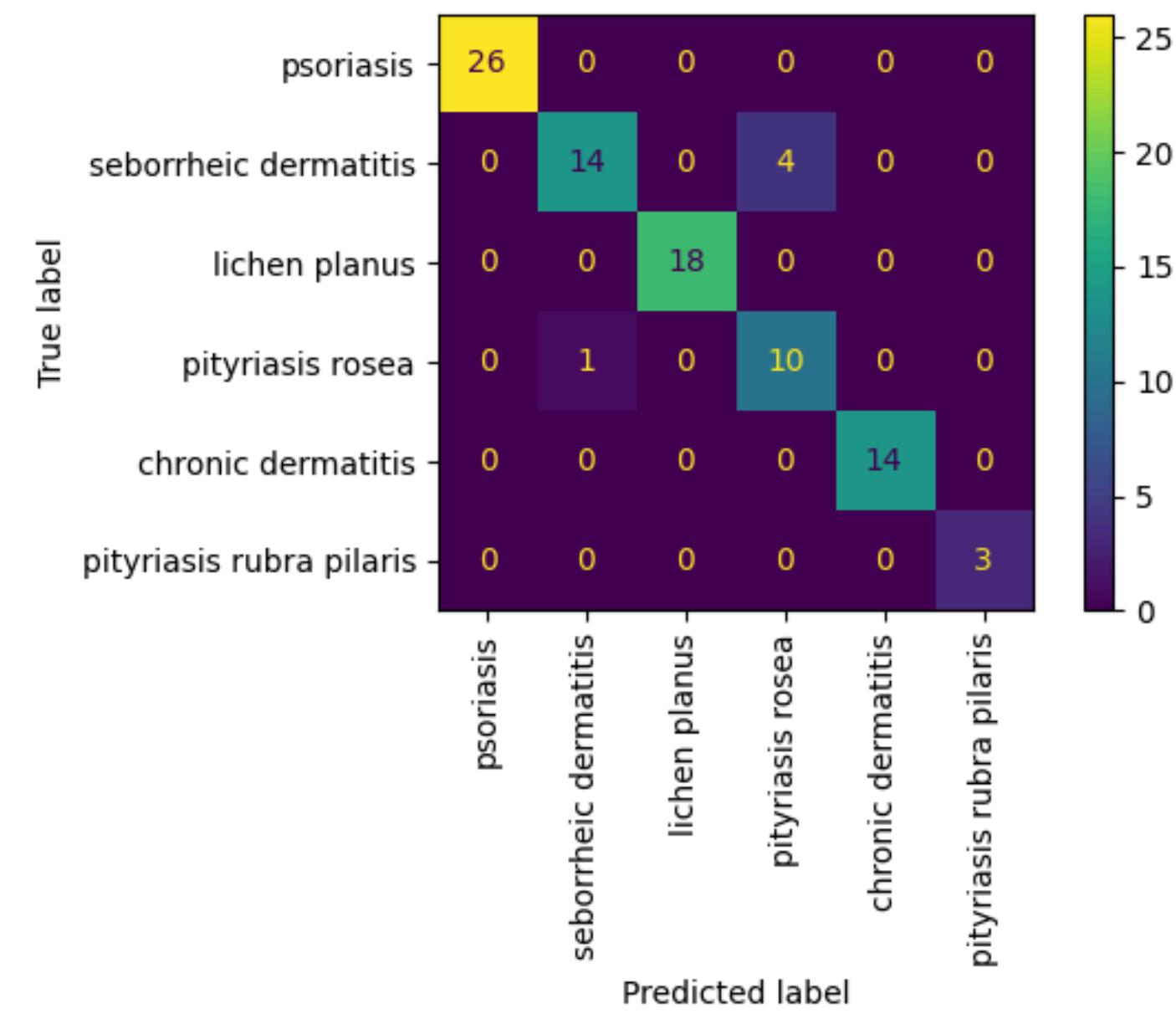


Modelos elegidos

Naive Bayes

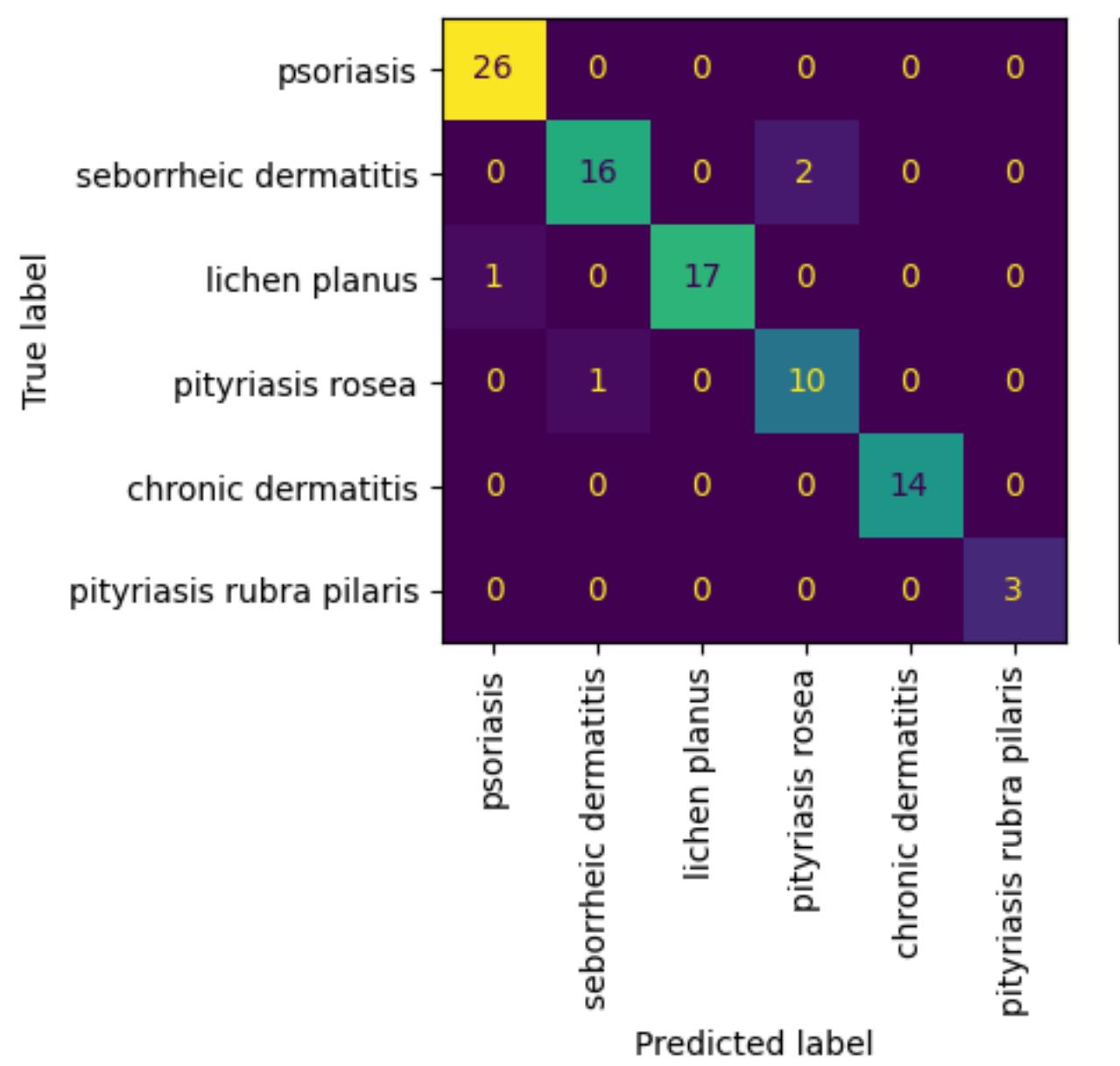


Principal Component Analysis

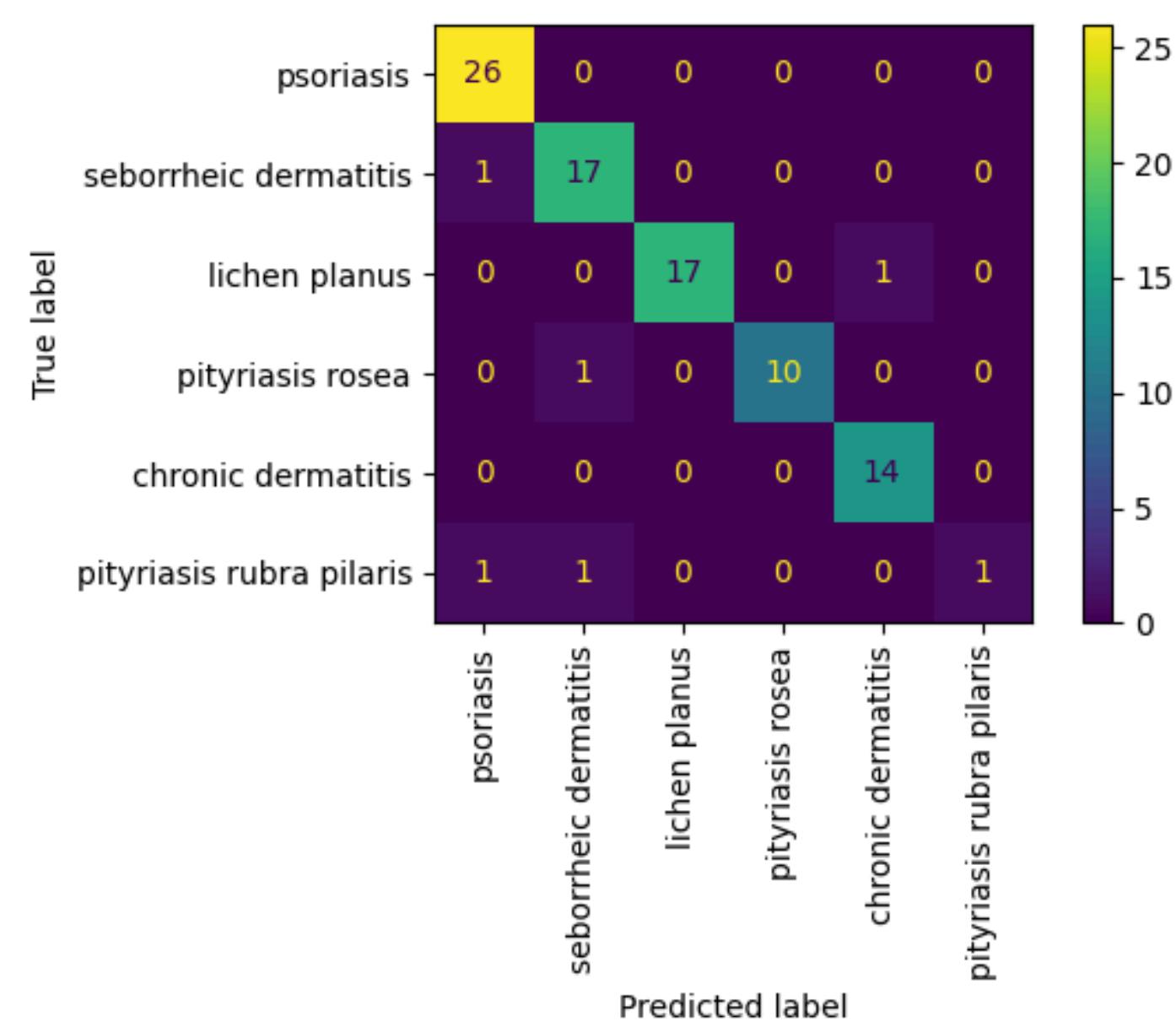


Modelos elegidos

Support Vector Machines

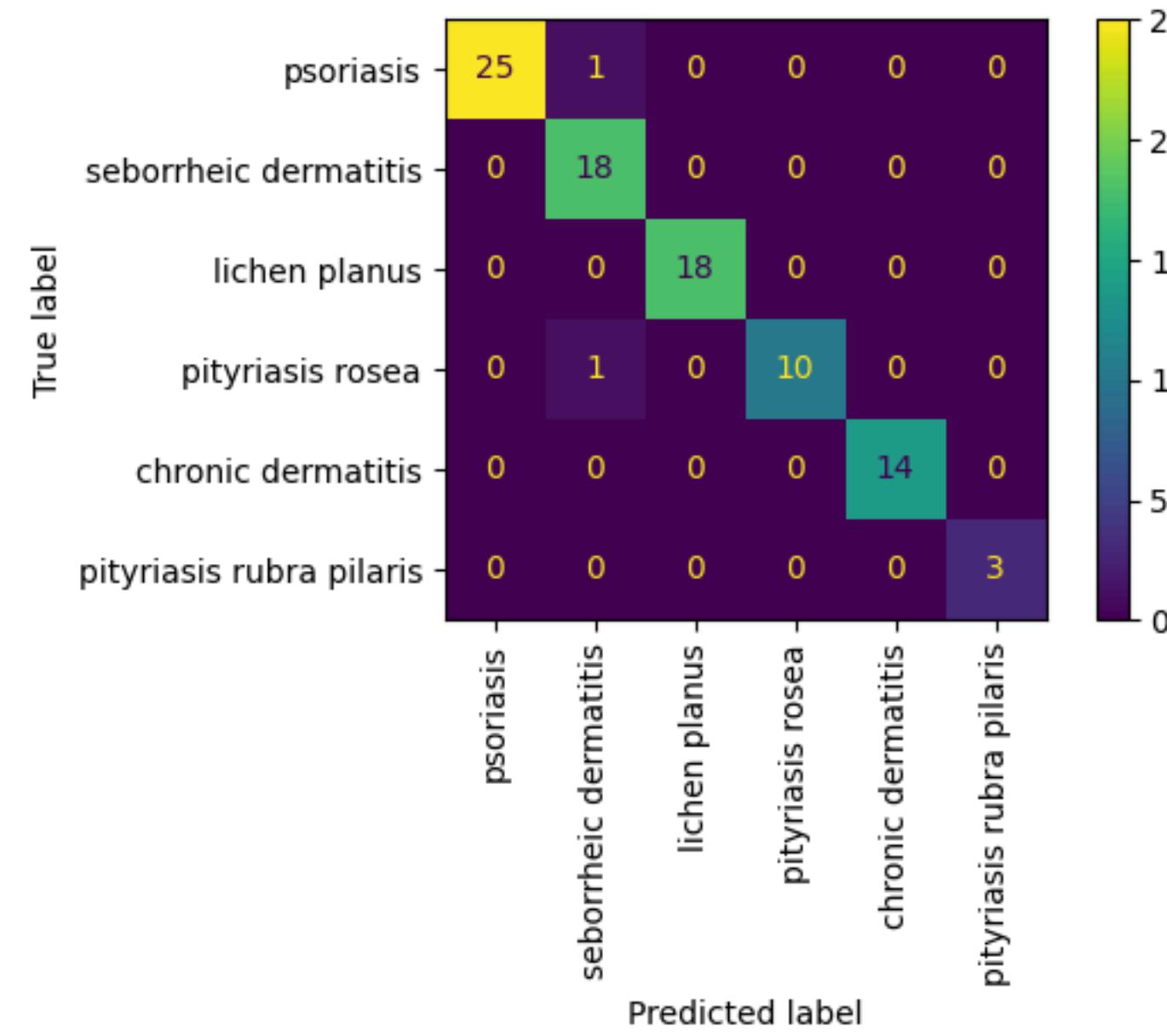


Decision Tree



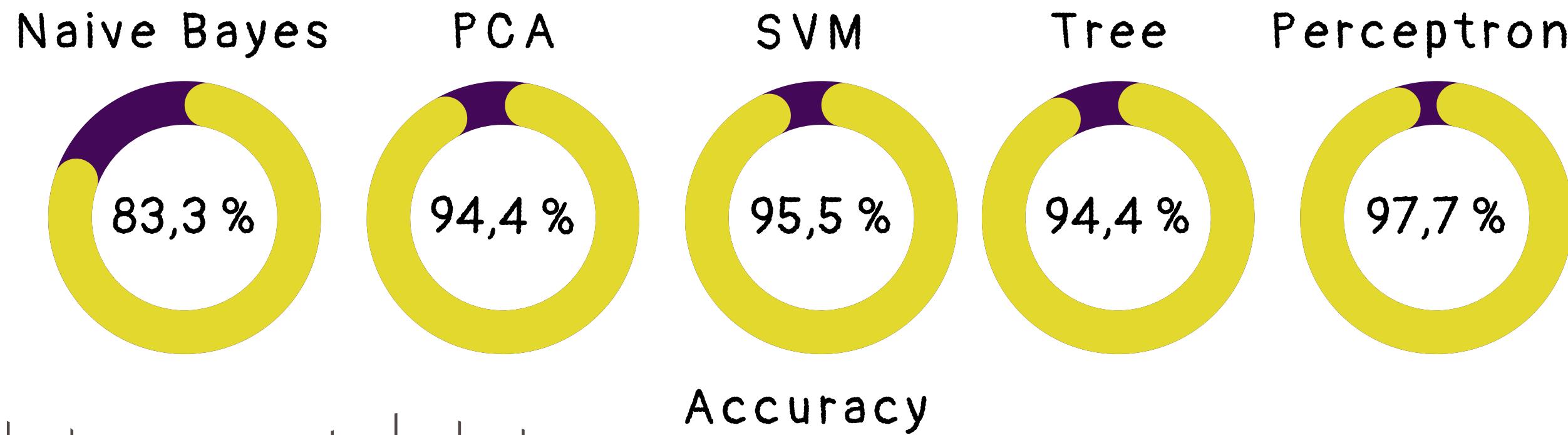
Modelos elegidos

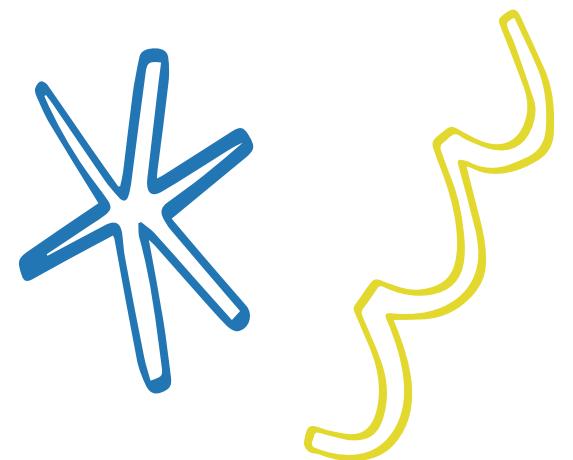
Perceptron



Resultados

	Accuracy	Precision	Recall	F1-Score	Training Time
Gaussian Naive Bayes	0.833333	0.889352	0.833333	0.821111	0.029322
Principal Component Analysis (PCA)	0.944444	0.951746	0.944444	0.945253	0.014424
Support Vector Machines (SVM)	0.955556	0.957165	0.955556	0.955750	0.021719
Decision Tree	0.944444	0.947942	0.944444	0.939519	0.005354
Perceptron	0.977778	0.980000	0.977778	0.977989	0.020897





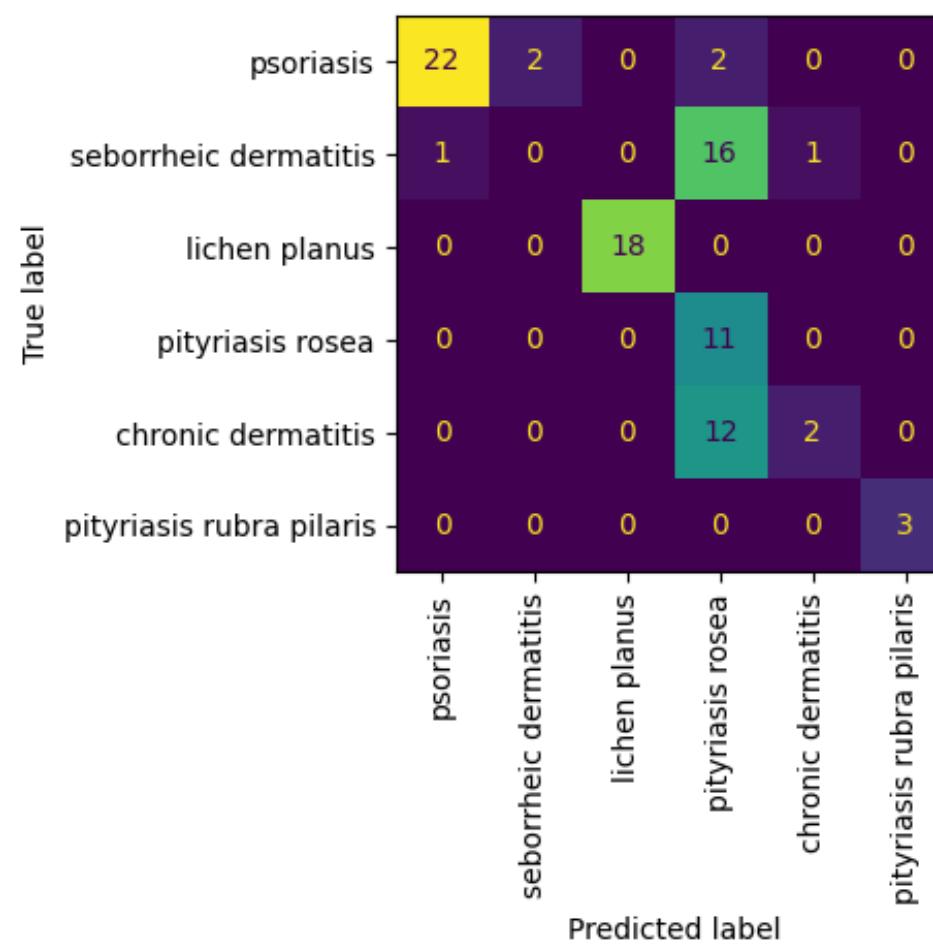
¿Sin biopsia, se puede?

Supongamos que tenemos el mismo problema de clasificación pero queremos usar solo las características clínicas, evitando realizar una biopsia para el diagnóstico de la enfermedad.

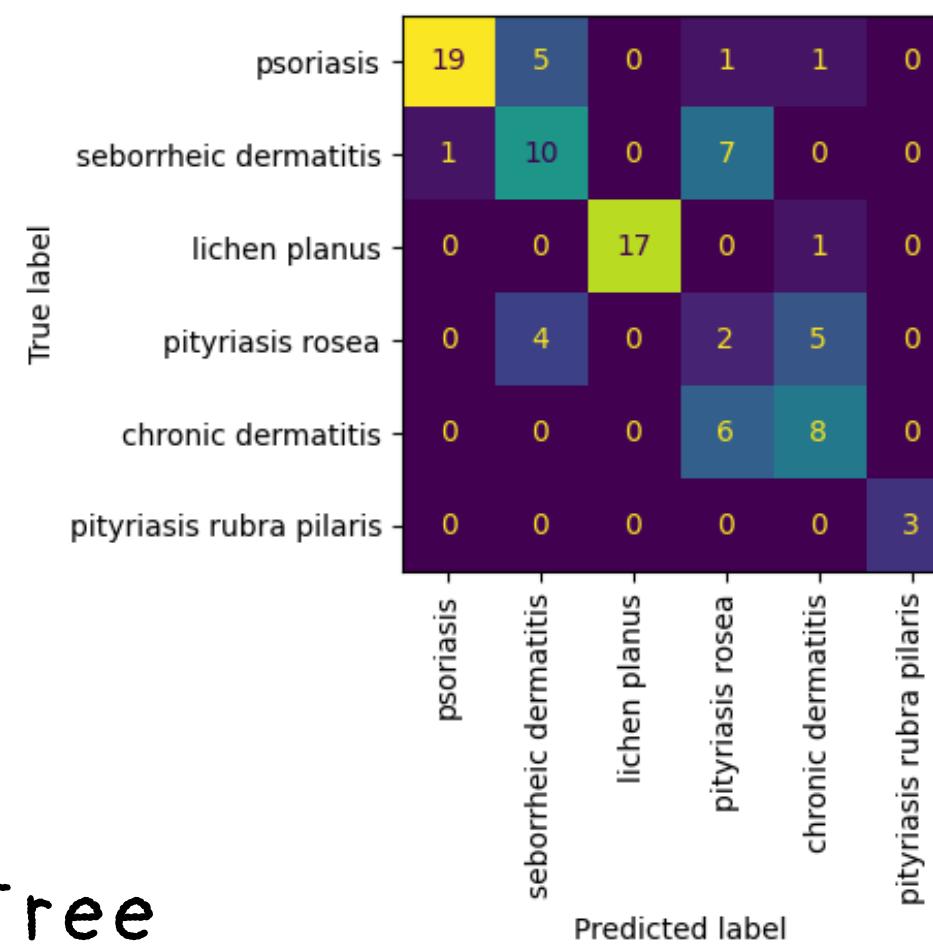
Si usamos los mismos clasificadores seguirán funcionando igual, serán mejores o los empeorara?



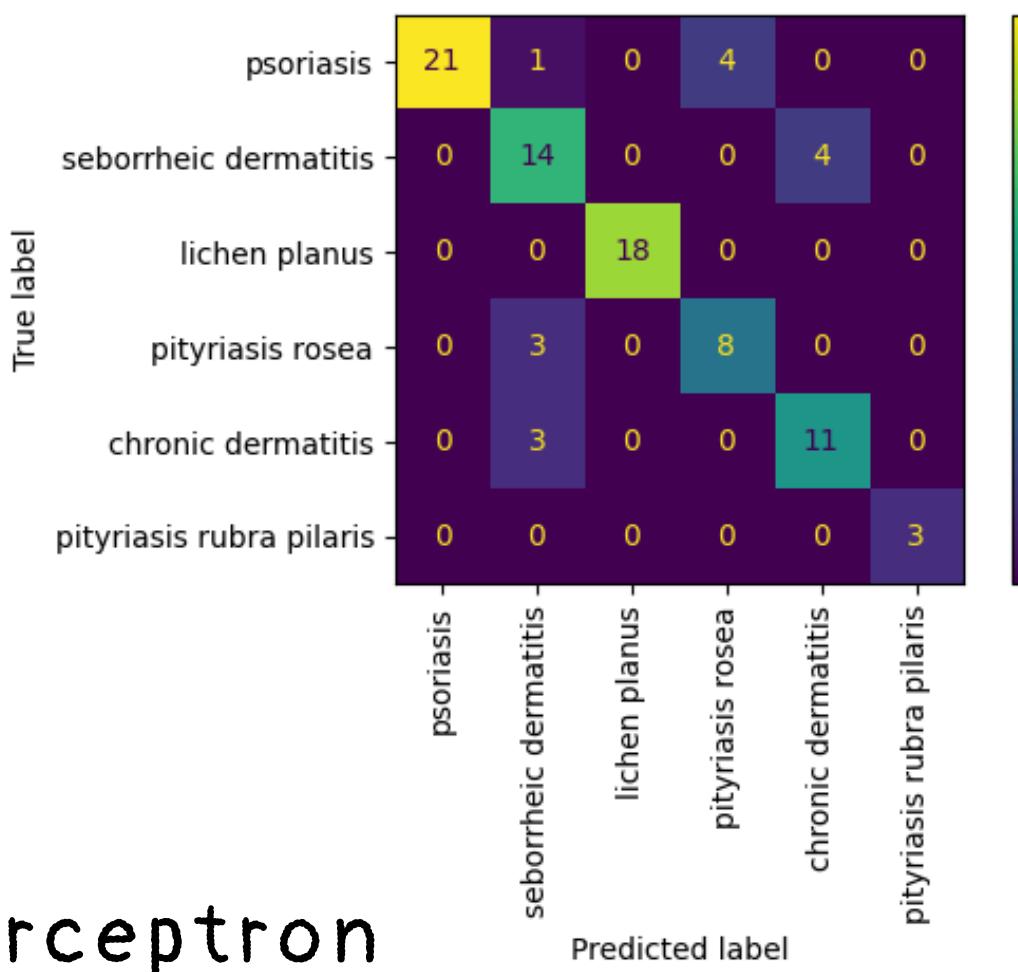
Naive Bayes



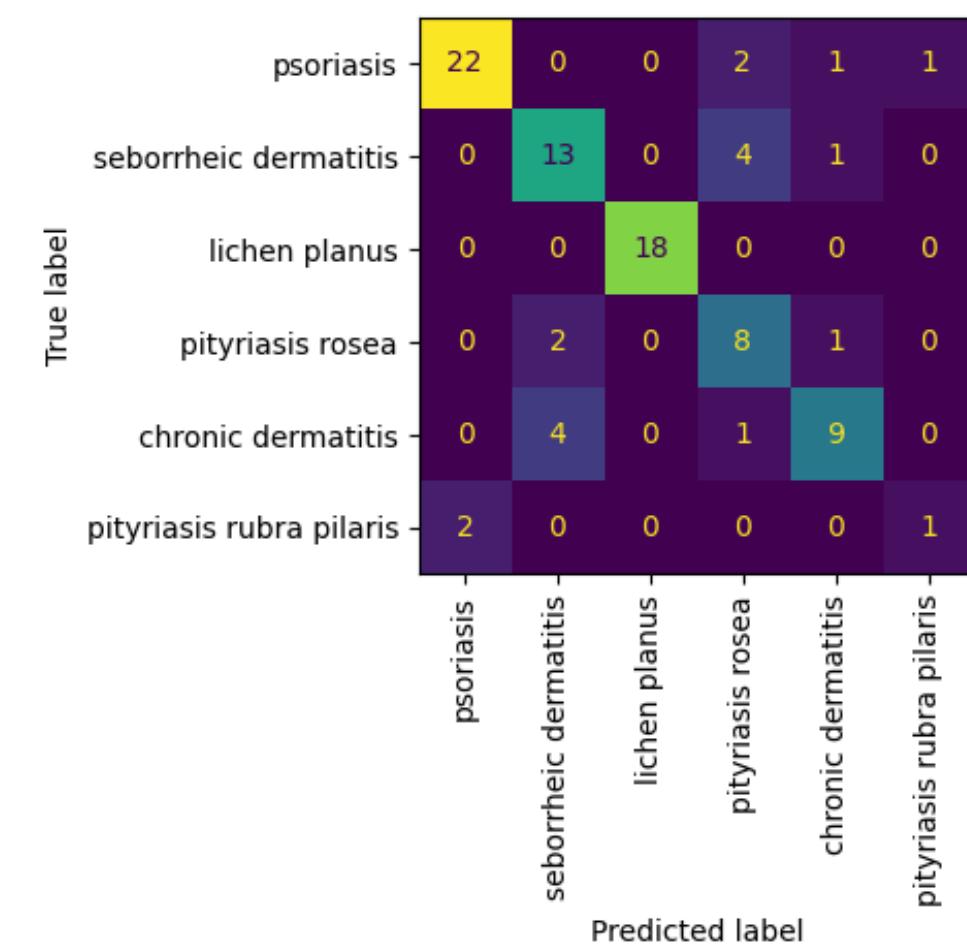
PCA



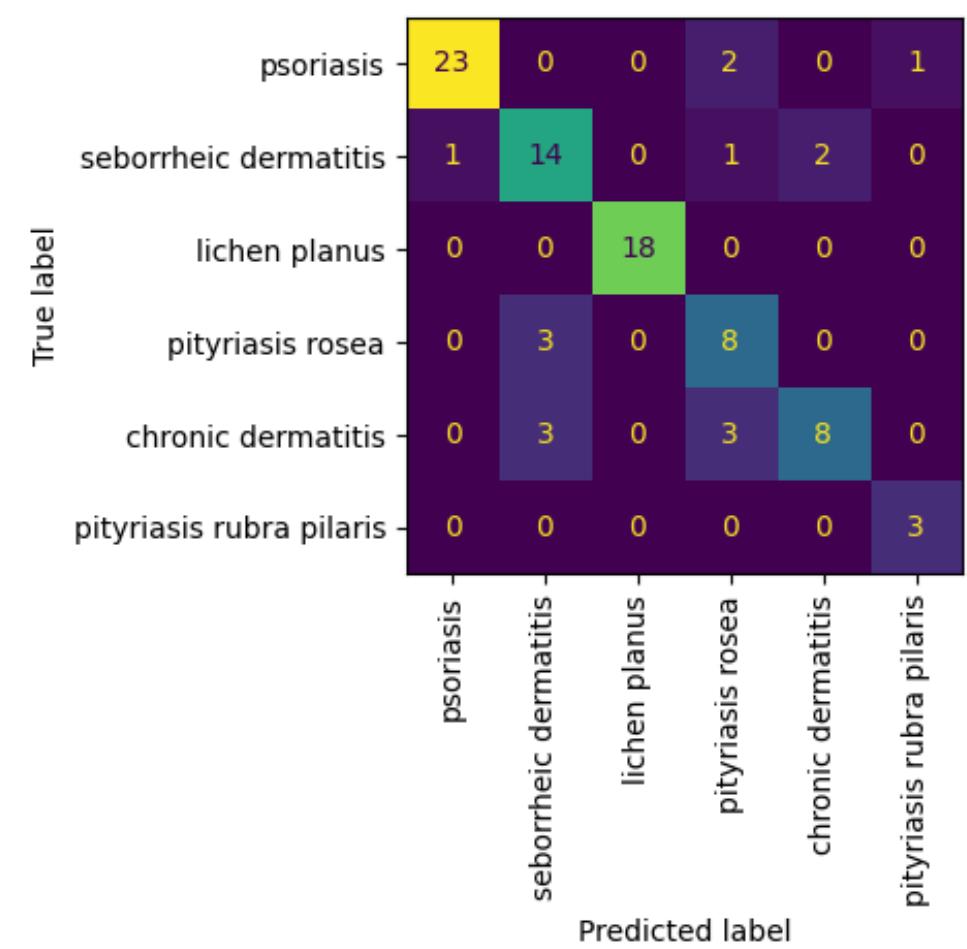
SVM



Tree

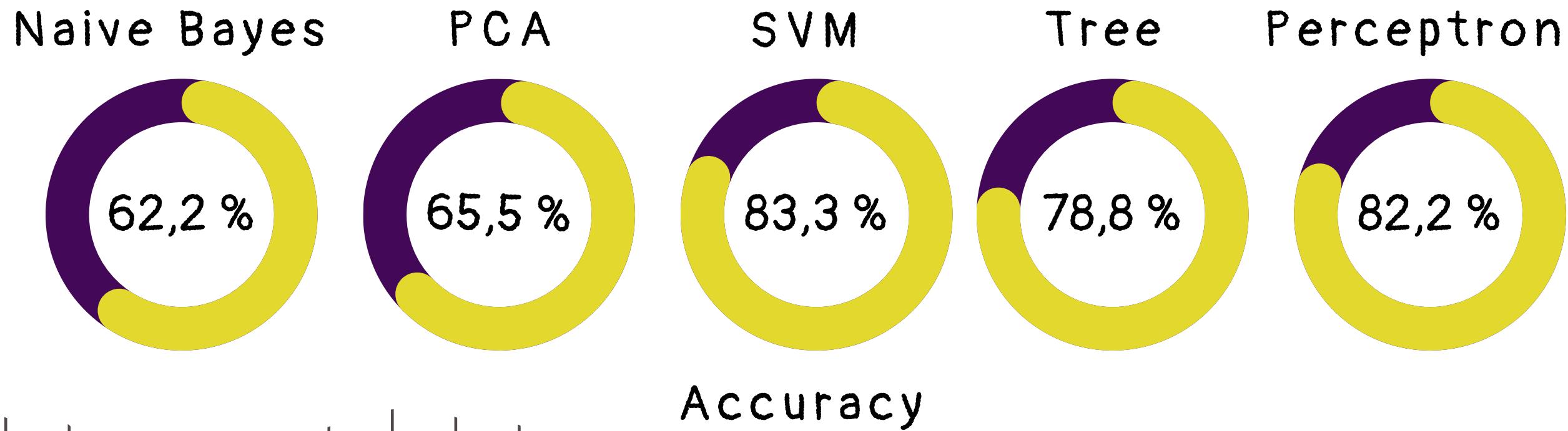


Perceptron



Resultados-Sin biopsia

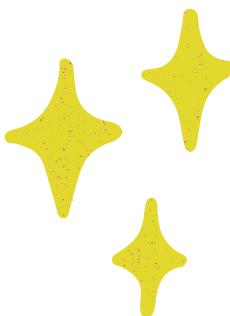
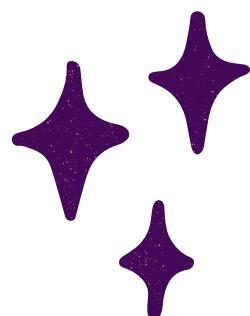
	Accuracy	Precision	Recall	F1-Score	Training Time
Gaussian Naive Bayes	0.622222	0.646157	0.622222	0.581054	0.004388
Principal Component Analysis (PCA)	0.655556	0.711282	0.655556	0.678305	0.003284
Support Vector Machines (SVM)	0.833333	0.851111	0.833333	0.838111	0.010729
Decision Tree	0.788889	0.800175	0.788889	0.791002	0.004035
Perceptron	0.822222	0.836138	0.822222	0.823644	0.012822



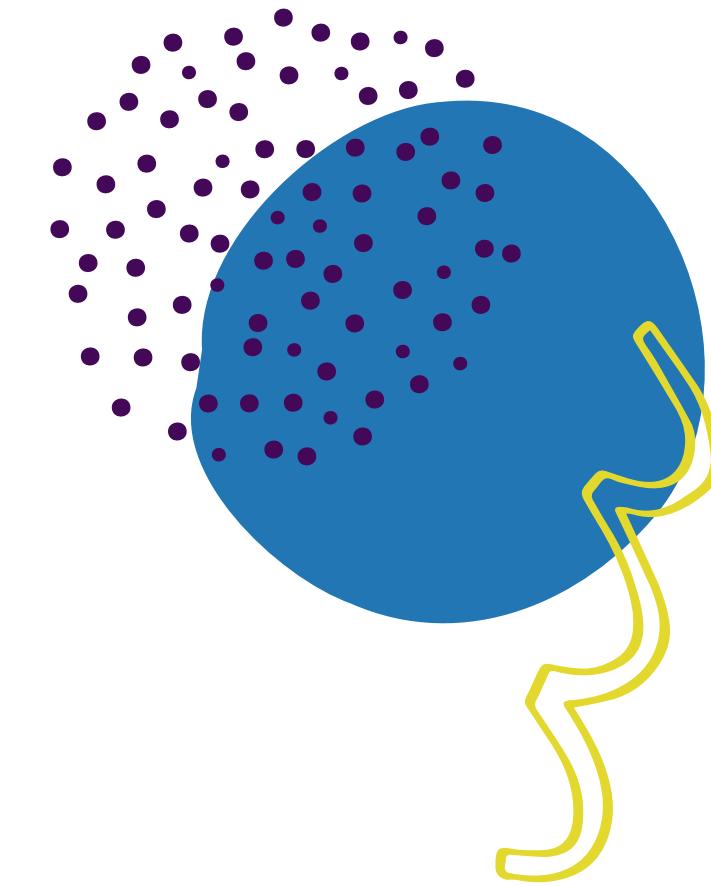
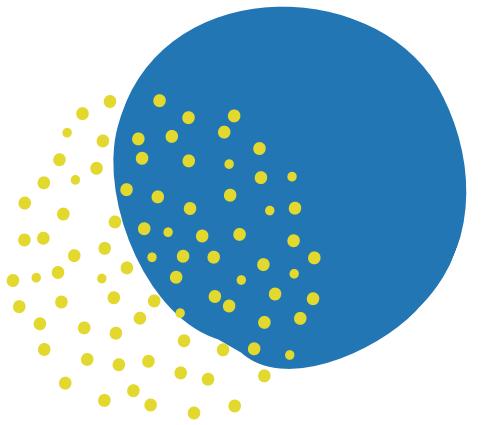
Conclusiones y Próximos pasos

- 01 Cumplimos exitosamente el objetivo
- 02 Otras publicaciones
 - Lograron una accuracy de 98.61% con Random Forest con Optimizacion Bayesiana

- 03 Mejorar y analizar el problema sin biopsia
- 04 Pasar a un problema de analisis de imagenes



Fuentes

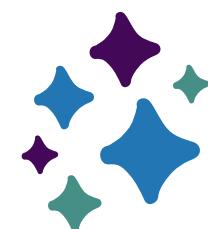


Fuente original: (1998) <https://archive.ics.uci.edu/dataset/33/dermatology>

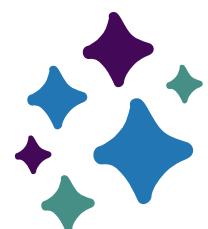
Fuente utilizada: <https://www.kaggle.com/datasets/olcaybolat1/dermatology-dataset-classification>

Link al colab: <https://colab.research.google.com/drive/1tKERc5IYJ-kPlxBEhIIAMDP6SzqGyJos?usp=sharing>

Otros colaboradores: <https://www.kaggle.com/datasets/olcaybolat1/dermatology-dataset-classification/code>



Muchas gracias



**Cuando te dicen que nunca
han ido al dermatólogo**



La decepción, la traición

Preguntas?