



UNIVERSIDAD DE ANTIOQUIA

Facultad de Ingeniería

Alumno

Albarrán Carlillo José Francisco

Profesor

Dr. Raúl Ramos Pollán

Proyecto

Contexto

- Como parte del proyecto se decidió abordar una competencia del año 2021 esta competencia es; “Plant Pathology 2021-FGVC8”, esta competencia consiste en desarrollar un modelo para la clasificación de imágenes de hojas de arboles de manzana, para poder detectar enfermedades en los mismos, estas enfermedades pueden afectar a la calidad y producción de las manzanas, la detección manual de estas enfermedades es costosa y demora mucho tiempo. Se han desarrollado modelos de visión artificial para poder identificar las enfermedades de manera más eficiente el problema es detectar de qué enfermedad se trata, es por esto que fue creada esta competición.

Objetivo de machine Learning.

- Dado un conjunto de entrenamiento de imágenes poder identificar las diferentes enfermedades que tiene la hoja del árbol.

Dataset.

- 18637 archivos, de tamaño 16.1 de tipo jpg, csv
- Para el entrenamiento se dan las imágenes de entrenamiento y las etiquetas, las imágenes que contengan muchas varias enfermedades tendrán una “clase compleja”, que tendrán a su vez un “sub-set” de enfermedades identificadas.
- Para los datos de entrenamiento se tendrá que hacer una entrega a la competencia del código para tener disponibles el set completo de imágenes de entrenamiento.

Métricas de desempeño

- Dentro de la competencia se tiene el denominado "mean f1-score" este es una métrica de evaluación utilizada comúnmente en competencias de aprendizaje automático de Kaggle para evaluar la calidad de un modelo de clasificación. Esta métrica se calcula como el promedio del valor F1 de cada clase en un conjunto de datos.
- Debido a que la competencia ya ha terminado podremos subir nuestro código para verificar nuestro puntaje aunque solo será para fines académicos, otra forma de saber que tan bien es nuestro modelo podemos basarnos en nuestra función de pérdida y en la exactitud.
- En lo que respecta al desempeño del negocio; La competencia fue desarrollada para ayudar a los agricultores que se dedican a la cosecha de manzanas, esto puede ayudar a que sus pérdidas debido a las enfermedades en los árboles disminuyan significativamente, ya que el conocimiento de qué enfermedad tiene un árbol puede ayudar a su tratamiento y control.

Referencias y resultados previos

- Esta competencia fue realizada en 2021 por lo cual ya se tienen resultados favorables el primero con lugar con un puntaje de 0.88336, según la medida "mean f1-score". Estos resultados y códigos pueden ser consultados en el siguiente link: <https://www.kaggle.com/competitions/plant-pathology-2021-fgvc8/leaderboard>.
- Muchos de los participantes de la competencia comparten sus avances y códigos, esto es muy útil para poder hacer modificaciones y poder ampliar nuestro conocimiento.
- Esta competencia fue basada en un data set de un artículo el cual es el siguiente; The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. Escrito por Thapa, Ranjita; Zhang, Kai; Snively, Noah; Belongie, Serge; Khan, Awais

Un resumen del artículo es el siguiente

Los huertos de manzanas en los Estados Unidos son amenazados la mayor parte del tiempo por una gran cantidad de patógenos e insectos. La gestión adecuada y oportuna de las enfermedades depende de la detección temprana de las mismas, se han capturado manualmente 3651 imágenes de síntomas de alta calidad de la vida real de múltiples enfermedades foliares en manzanas, se hizo disponible un conjunto de imágenes anotadas por expertos para crear un conjunto de datos piloto para el desafío de Patología Vegetal en la comunidad de Kaggle, como parte de taller de (FGVC), se pidió a los participantes que utilizaran el conjunto de datos de imágenes para entrenar un modelo de aprendizaje automático para clasificar las categorías de enfermedades y desarrollar un algoritmo para la cuantificación de la gravedad de la enfermedad. Se entrenó una red neuronal convolucional predefinida en estos datos para la clasificación de enfermedades y logramos una precisión del 97% en un conjunto de prueba reservado.

Como podemos observar la competencia de Kaggle sirvió para poder utilizarlo en la publicación del artículo con lo cual se ayudó en esta problemática.