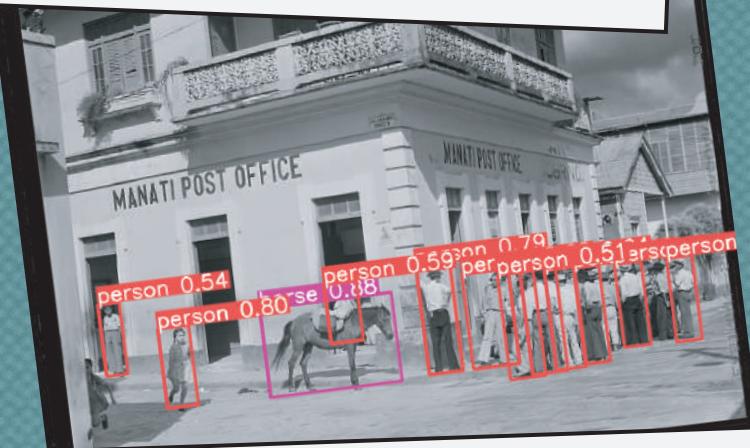




VAMOS A INTENTARLO CON ESTA IMAGEN..

UTILIZAREMOS UN MODELO DE APRENDIZAJE AUTOMÁTICO LLAMADO TRANSFORMADOR DE VISIÓN (VIT EN INGLÉS).

EL TRANSFORMADOR DE VISIÓN HACE DOS COSAS.



1.

INTENTA DETECTAR ELEMENTOS CONOCIDOS EN LA IMAGEN (COSAS COMO PERSONAS, AUTOS O FLORES), Y APRENDE EL TAMAÑO Y LA POSICIÓN DE ESAS COSAS.

2.

UTILIZA LO QUE HA APRENDIDO DE LAS IMÁGENES ETIQUETADAS, CON GENERAR UNA NUEVA LEYENDA

{
"generated_text":
"a man standing next to
a building with a horse"
}



A MENUDO SON PEORES QUE NO TENER NINGUNA LEYENDA.
ESTAS LEYENDAS GENERADAS POR MÁQUINAS ESTÁN LEJÍSIMOS DE LAS NORMAS DE EXACTITUD DE LA BIBLIOTECA.

{
"generated_text":
"a woman is standing in
a parking lot "
}

... Y LUGARES QUE NO ESTÁN EN SU VISIÓN DEL MUNDO, COMO LOS ASENTAMIENTOS IRREGULARES.



{
"generated_text":
"a man in a field with
a bunch of dead trees "
}

SIN EMBARGO, A MENUDO LOS MODELOS TIENEN DIFICULTADES CON OBJETOS QUE NO APARECEN MUCHO EN SU MATERIAL DE ENTRENAMIENTO (COMO LA CAÑA DE AZÚCAR).



{
"generated_text":
"people standing around a building "
}

A VECES, LAS DESCRIPCIONES DE LAS IMÁGENES QUE OBTENEMOS DE LAS FOTOS DE JACK DE PUERTO RICO SON PRECISAS... AUNQUE POCO IMAGINATIVAS.