



EL ENTRENAMIENTO DE ESTE TIPO DE MODELOS DE APRENDIZAJE AUTOMÁTICO SUELEN BASARSE EN UN CONJUNTO DE DATOS DE 2012 LLAMADO IMAGENET.

IMAGENET CONTIENE 14.197.122 IMÁGENES EXTRAÍDAS DE INTERNET.

ESTAS IMÁGENES FUERON ETIQUETADAS INDIVIDUALMENTE POR TRABAJADORES DE CONTENIDOS DE TODO EL MUNDO (QUE A MENUDO GANABAN MUCHO MENOS QUE EL SALARIO MÍNIMO).

LOS TRABAJADORES PODÍAN ELEGIR ETIQUETAR LAS IMÁGENES A PARTIR DE 22.000 'CLASES', TOMADAS DE UNA BASE DE DATOS LÉXICA LLAMADA WORDNET.

LAS PROPIAS CLASES PROCEDEN DE ESTUDIANTES DE POSGRADO QUE, EN LOS AÑOS 90, ETIQUETARON A MANO EL CONTENIDO ESCRITO DEL WALL STREET JOURNAL.

UN MODELO ENTRENADO EN IMAGENET, ENTONCES, SÓLO SABE LO QUE HA APRENDIDO DE INTERNET (INCLUYENDO MEMES Y EL RESTO), LEÍDO A TRAVÉS DE LA LENTE DE UN PERIÓDICO DE NUEVA YORK, ESCRITO PARA OPERADORES DE BOLSA DE MEDIANA EDAD.

ES UN MODELO QUE SABE MUCHO SOBRE GATOS...

Y MUY POCO SOBRE PUERTO RICO EN LA DÉCADA DEL 40..