

**UNIVERSIDAD POLITÉCNICA SALESIANA  
SEDE QUITO**

**CARRERA:  
INGENIERÍA DE SISTEMAS**

**Trabajo de titulación previo a la obtención del título de:  
Ingenieros de Sistemas**

**TEMA:  
CONSTRUCCIÓN DE UN CONJUNTO DE IMÁGENES FACIALES CON  
METADATOS BIOMÉTRICOS Y RASGOS ÉTNICOS DE ECUATORIANOS**

**AUTORES:  
JORGE HUMBERTO AVILÉS VALENCIA  
HENRY WENSESLAO TOAPANTA HERRERA**

**TUTORA:  
PAULINA ADRIANA MORILLO ALCÍVAR**

**Quito, julio de 2019**

## CESIÓN DE DERECHOS DE AUTOR

Nosotros, Jorge Humberto Avilés Valencia, con documento de identificación N° 1723083851, y Henry Wenseslao Toapanta Herrera, con documento de identificación N° 1723620033, manifestamos nuestra voluntad y cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del trabajo de titulación con el tema: “CONSTRUCCIÓN DE UN CONJUNTO DE IMÁGENES FACIALES CON METADATOS BIOMÉTRICOS Y RASGOS ÉTNICOS DE ECUATORIANOS.”, mismo que ha sido desarrollado para optar por el título de INGENIEROS DE SISTEMAS en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en nuestra condición de autores nos reservamos los derechos morales de la obra antes citada.

En concordancia, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

.....  
JORGE HUMBERTO  
AVILÉS VALENCIA

CI: 1723083851

.....  
HENRY WENSESLAO  
TOAPANTA HERRERA

CI: 1723620033

Quito, julio de 2019

## DECLARATORIA DE COAUTORÍA DE LA TUTORA

Yo declaro que bajo mi dirección y asesoría fue desarrollado el Artículo Académico, con el tema: CONSTRUCCIÓN DE UN CONJUNTO DE IMÁGENES FACIALES CON METADATOS BIOMÉTRICOS Y RASGOS ÉTNICOS DE ECUATORIANOS, realizado por Jorge Humberto Avilés Valencia y Henry Wenseslao Toapanta Herrera, obteniendo un producto que cumple con todos los requisitos por la Universidad Politécnica Salesiana, para ser considerados como trabajo final de titulación.

Quito, julio de 2019



.....  
PAULINA ADRIANA MORILLO ALCÍVAR  
CI: 1715646574

# CONSTRUCCIÓN DE UN CONJUNTO DE IMÁGENES FACIALES CON METADATOS BIOMÉTRICOS Y RASGOS ÉTNICOS DE ECUATORIANOS

## CONSTRUCTION OF A SET OF FACIAL IMAGES WITH BIOMETRIC METADATA AND ETHNIC CHARACTERISTICS OF ECUADORIANS

Jorge Avilés<sup>1</sup>, Henry Toapanta<sup>2</sup>, Paulina Morillo<sup>3</sup>, Diego Vallejo<sup>4</sup>

### Resumen

Actualmente, las técnicas de Machine Learning son cada vez más usadas en diversas áreas de la ciencia y la sociedad, su funcionamiento está basado en la recolección y análisis sistemático de datos que permiten crear modelos de predicción.

Hoy en día existen diversas fuentes para adquirir datos/imágenes, como los repositorios públicos de gobiernos, instituciones académicas, centros de investigación y datos generados por empresas, entre otros. A pesar de todas estas opciones no siempre se tienen los datos adecuados para la construcción de los modelos, más aún cuando se abordan problemas que requieren el uso de imágenes. Este artículo propone la construcción de un conjunto de imágenes faciales pertenecientes a las etnias más representativas del Ecuador: afro-ecuatorianos, mestizos, indígenas y blancos (europeo-descendientes).

### Abstract

Currently, Machine Learning techniques are increasingly used in various areas of science and society, its operation is based on the systematic collection and analysis of data that allow create prediction models.

Today, there are several sources to acquire data / images, such as public repositories of governments, academic institutions, research centers and data generated by companies, among others. Despite all these options, the adequate data for the construction of the models are not always available, especially when problems that require the use of images are addressed. This article proposes the construction of a set of facial images of Ecuadorians belonging to the most representative ethnic groups of Ecuador: afro-ecuatorians, mestizos, indigenous and european-descendants.

In addition, metadata referring to the facial features of each ethnic group obtained through the application of a

---

<sup>1</sup>Carrera de Ingeniería de Sistemas, Universidad Politécnica Salesiana, Ecuador, <https://orcid.org/0000-0002-7007-6997>; Autor para correspondencia: javilesv1@est.ups.edu.ec

<sup>2</sup>Carrera de Ingeniería de Sistemas, Universidad Politécnica Salesiana, Ecuador, <https://orcid.org/0000-0002-2053-577X>; Autor para correspondencia: htoapantahe@est.ups.edu.ec

<sup>3</sup>Grupo de investigación IDEIAGEOCA, Universidad Politécnica Salesiana, Ecuador,

<sup>4</sup>Grupo de investigación IDEIAGEOCA, Universidad Politécnica Salesiana, Ecuador, <https://orcid.org/0000-0002-2704-3858>

Además, se incluyen metadatos referentes a los rasgos faciales característicos de cada etnia obtenidos a través de la aplicación de una encuesta y las biométricas respectivas calculadas por medio de la ejecución de modelos *ssd inception v2 coco* y *faster rcnn inception v2 coco*, de la librería *TensorFlow*. El conjunto de imágenes resultante cuenta con 430 instancias, cada una con 24 atributos. Los resultados en la detección y adquisición de las biométricas del rostro fueron contrastados con mediciones físicas, consiguiendo errores inferiores al 10%.

***Palabras Clave:*** Visión Artificial, Redes Neuronales, TensorFlow, Antropometría, Dataset

survey and the respective biometrics, calculated through the execution of *ssd inception v2 coco* and *faster rcnn inception v2 coco*, models of the TensorFlow library are included. The resulting image dataset has 430 instances, each with 24 attributes. The results in the detection and acquisition of the face biometrics were contrasted with physical measurements, achieving errors less than 10%.

***Keywords:*** Artificial Vision, Neural Networks, TensorFlow, Anthropometry, Dataset

## 1. Introducción

El desarrollo de sistemas de *Machine Learning* (ML) y *Artificial Vision* (AV) depende en gran parte de la cantidad y calidad de datos que utilizan para ser precisos [1]. Cuando no existen datos suficientes o los *datasets* presentan inconsistencias, los sistemas basados en ML producirán resultados erróneos poco extrapolables a la realidad [2].

En la actualidad, existen muchos sistemas de ML y AV orientados al reconocimiento de etnias, los cuales utilizan *datasets* específicos para dicho propósito [3] [4] [5] [6].

Ecuador, a pesar de ser un país multiétnico y pluricultural, no cuenta con un *dataset* o conjunto de imágenes faciales propias de ecuatorianos pertenecientes a las etnias más representativas: afro-ecuatorianos, mestizos, indígenas y europeo-descendientes.

En este artículo se propone la construcción de un conjunto de imágenes faciales y sus respectivos metadatos biométricos propios de ecuatorianos. Además, se propone una metodología específica para la recolección de datos, mediante una encuesta, la misma que será aplicada en regiones con mayor densidad demográfica de cada uno de estos grupos étnicos.

Este artículo está organizado de la siguiente manera: en la Sección II, se presenta el estado del arte de trabajos relacionados con la identificación étnica, incluyendo la respectiva descripción del conjunto de imágenes utilizado para dicho propósito. La Sección III, describe la metodología empleada para la recolección de datos e imágenes mediante una encuesta. Además, se describe el procedimiento que se llevó a cabo para el

entrenamiento de una Red Neuronal (RN), a través de la herramienta *TensorFlow* [7], así como la toma y estructuración de los metadatos biométricos respectivos. Por otro lado, en la Sección IV se incluyen los experimentos realizados y sus respectivos resultados. Finalmente se presentan las conclusiones y perspectivas del trabajo.

## 2. Trabajos relacionados

Hoy en día, existen varias técnicas para el reconocimiento étnico. Algunas de ellas abordan esta tarea mediante el estudio del patrón y forma del iris de cada individuo [8] [9].

Otras investigaciones lo resuelven a través de la extracción de las características geométricas del rostro [10] [11] [12]. Ambos enfoques requieren el uso de *datasets*, que pueden ser tomados desde repositorios de datos públicos [13] o pueden construirse a través de la aplicación de entrevistas o encuestas tal como se propone en esta investigación.

En [14], se utilizaron imágenes de la base de datos *Iris de la Chinese Academy of Sciences Institute of Automation* (CASIA v2), conformada por 2400 fotografías de ojos asiáticos, con el objetivo de detectar patrones de textura del iris que permitan determinar las características étnicas y género de cada individuo. Para esto se realizó un análisis comparativo entre diferentes técnicas y algoritmos cuyo rendimiento fue del 84.95% de acierto.

En la investigación [15], construyeron un *dataset* específico conformado por más de 220 imágenes de rostros para el reconocimiento de las principales etnias en China. Para la extracción de rasgos étnicos, se analizaron las características algebraicas y geométrico-faciales más representativas.

Posteriormente, se aplicaron diferentes algoritmos de ML que obtuvieron un porcentaje de acierto del 79% al 90.95%.

De la misma manera en [16], realizaron la construcción de un *dataset* mediante la captura de imágenes faciales y la recolección de metadatos como: sexo, etnia, edad y expresión facial, en total este dataset se conformó por 1991 imágenes. Esta investigación usó una *Support Vector Machine* (SVM) para la clasificación étnica de los individuos, el rendimiento del clasificador fue superior al 90%, bajo ciertas condiciones de luz.

En [17], utilizaron un total de 675 imágenes faciales tomadas desde un plano lateral conformadas por 135 sujetos multiétnicos con el fin de llevar a cabo una clasificación étnica binaria mediante la extracción de la geometría facial utilizando alrededor de 10 puntos faciales antropométricos. El reconocimiento y clasificación se lo realizó mediante una SVM, obteniendo una precisión del 76.06%.

La mayoría de los trabajos relacionados con el reconocimiento étnico requieren la construcción de un *dataset* con imágenes de rostros con características particulares a las de su etnia, ergo, este trabajo propone la creación de un *dataset* mediante una metodología específica para recolección de datos, toma de imágenes y extracción de los respectivos metadatos biométricos de: cejas, ojos, nariz y boca.

### **3. Metodología**

La metodología utilizada en la investigación consta de cuatro fases: diseño y aplicación de la encuesta, adquisición y corrección de imperfecciones de la imagen, entrenamiento y evaluación de

*TensorFlow* para la identificación de las partes del rostro, extracción y almacenamiento de los metadatos biométricos del rostro a partir de la imagen y los resultados de la encuesta.

#### **3.1. Diseño y aplicación de la encuesta**

La encuesta se utilizó para la recolección de información personal, características faciales y ancestralidad. Para su desarrollo, se utilizaron dos tipos de métodos: el método objetivo utilizado para obtener información como: apellidos, edad, género, ciudad y provincia de nacimiento. Por otra parte, el método subjetivo permitió determinar características faciales y ancestralidad [18].

##### **3.1.1. Diseño de la encuesta**

La elaboración de la encuesta consideró las características faciales más representativas de los ecuatorianos, de acuerdo a la Coordinación Nacional de Criminalística, Medicina Legal y Ciencias Forenses de la Policía Nacional del Ecuador [19] [20] [21], se conforma de tres tercios faciales. Tercio Superior: cabello y color de piel, Tercio Medio: cejas, ojos, nariz y orejas. Tercio Inferior: boca y mentón [22].

Las preguntas de la encuesta abarcaron cada uno de los tercios faciales, sin embargo, debido al objetivo de esta investigación, para la Sección IV solo se tomaron en cuenta cuatro partes del rostro, correspondientes al tercio medio e inferior: cejas, ojos, nariz y boca.

La encuesta diseñada para la obtención de las características etno-faciales, tomó en cuenta tres dimensiones: el lenguaje, la redacción y los tipos de preguntas.

En general, el lenguaje utilizado para la elaboración de las preguntas, fue de uso

coloquial con palabras fáciles de comprender, evitando ambigüedad y relectura de las preguntas [23], la redacción fue desarrollada de tal forma que las preguntas son claras y especifican, exactamente lo que se desea conocer.

La encuesta se conformó a través de dos tipos de preguntas: abiertas y cerradas de alternativa múltiple. A través de las preguntas abiertas se obtuvo información personal, mientras con las preguntas cerradas de alternativa múltiple se obtuvo información sobre las características faciales y ancestralidad de cada encuestado [24].

El método, tipo de pregunta y variables, que conforman la encuesta se describen en la Tabla 1.

**Tabla 1.** Variables en la encuesta.

Método	T. Pregunta	Variable
Objetivo	Abierta	apellido paterno, apellido materno, edad, ciudad de nacimiento
	Cerrada	provincia de nacimiento
Subjetivo	Abierta/Cerrada	género
	Cerrada	tipo de cabello, color de cabello, forma de labios, color de piel, color de ojos, forma de rostro, forma de orejas, forma de mentón, forma de nariz, tipo de ojos, forma de cejas
	Cerrada	ancestralidad, porcentaje afro-americano, porcentaje europeo, porcentaje indígena

### 3.1.2. Aplicación de la encuesta

El último Censo de Población y Vivienda realizado en el Ecuador en 2010 obtuvo

resultados de la autoidentificación étnica, mediante la consulta de costumbres y tradiciones de los ciudadanos.

El Censo determinó que los habitantes del Ecuador se identificaban de acuerdo con los datos mostrados en la Tabla 2 [25].

**Tabla 2.** Distribución étnica en el Ecuador.

Etnia	Porcentaje
Mestizos	71.9%
Montubios	7.4%
Afro-ecuatorianos	7.2%
Indígenas	7%
Blancos	6.1%
Otros	0.3%

Para la aplicación de la encuesta, se tomó en cuenta la distribución étnica de la población ecuatoriana del Censo 2010, de este modo, se seleccionó el método de muestreo aleatorio por conglomerados, para dividir la población en las zonas geográficas [26], con mayor concentración demográfica de cada una de las etnias más representativas del Ecuador: afro-ecuatorianos, mestizos, indígenas y blancos (europeo-descendientes).

El tamaño de la muestra se calculó en base a (1) [27], con tamaño de población conocido:

$$n = \frac{Z^2 p q N}{e^2(N-1) + Z^2 p q} \quad (1)$$

Donde:

n = Tamaño de la muestra.

N = Tamaño de la población.

p = Probabilidad a favor.

q = Probabilidad en contra.

Z = Nivel de confianza.

e = Error de la estimación.

Con el objetivo de asegurar que la encuesta este redactada correctamente y sea entendida por los grupos de estudio, se



realizó un *pretest* [28]. El *pretest* fue aplicado en la ciudad de Quito- Ecuador, a una población con características demográficas similares a las de la muestra.

El número total de personas encuestadas fue de 69 y los resultados arrojaron pequeñas inconsistencias de contenido como ancestralidad que fueron corregidos a-posteriori.

### 3.2. Adquisición y corrección de imperfecciones de la imagen.

La imagen del rostro de cada uno de los participantes de este proyecto, fueron adquiridas por medio de una cámara Canon EOS Rebel T3i, bajo condiciones similares de: luz, fondo y posición de la cámara. La configuración externa para la captura de la imagen se detalla en la Figura 1.

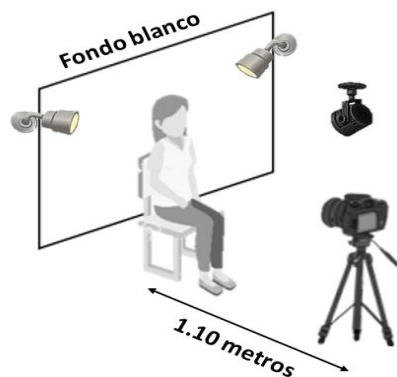


Figura 1. Escenario fotográfico.

La posición horizontal entre el sujeto y la cámara se definió en 1.10 metros, así se logró que el rostro aparezca bien definido y centrado en la escena. La posición vertical de la cámara varía de acuerdo con la altura del sujeto, sin embargo, se consideró que el lente de la cámara se fije a la altura de los ojos del individuo.

El esquema de iluminación empleado fue de tipo artificial con tonalidad blanca, aplicando técnicas de tipo frontal y lateral para evitar la aparición de sombras indeseables. Debido a que la fotografía fue

adquirida en diferentes locaciones, fue necesario la utilización de un fondo color blanco con el fin de minimizar en lo posible los cambios en la escena y evitar elementos distractores.

Por otro lado, para guardar una apariencia uniforme entre las imágenes se operaron las mismas condiciones de postura y vestimenta. Las cuales consideran una postura erguida, con posición natural de cabeza, hombros relajados, y una expresión facial neutra. Además, los sujetos utilizaron un mandil blanco, con el rostro y cuello descubiertos, cabello apartado de la cara, sin lentes y ningún elemento que obstaculice el rostro.

La configuración interna para la captura de la imagen, a través de la cámara, contempla una distancia focal de 24 milímetros, con un tiempo de exposición de 60 segundos, sobre el individuo.

Finalmente, tomando en cuenta la relación calidad/peso de la imagen se seleccionó la resolución y tamaño de esta a 2.5 Mpx y 1920 x 1280, respectivamente. El peso promedio de cada imagen resultante fue de 1.3 MB.

En la etapa de captura de la imagen se tomaron tres fotografías por cada individuo, de las cuales se seleccionó la de mejor calidad en cuanto a nitidez e iluminación, como resultado de esta actividad se obtiene una imagen principal, que será utilizada en etapas posteriores.

Adicionalmente, en el caso de que la imagen principal presente imperfecciones en el fondo, se realizan las correcciones utilizando el programa Gimp v2.10. A continuación, se recorta la imagen, enmarcando el rostro de cada encuestado.

No obstante, hubo fotografías deficientes, ajenas al escenario propuesto,

por lo cual fue necesario definir ciertos parámetros mínimos que ha de cumplir la fotografía para no ser descartada. Una fotografía es deficiente, porque a pesar de haber sido capturada conforme al escenario propuesto todavía presenta: imperfecciones en el fondo, vestimenta que obstaculiza en su totalidad el cuello del encuestado, ojos cerrados o postura de cabeza inclinada.

### 3.3. Entrenamiento y evaluación de TensorFlow

Este apartado describe el procedimiento que se llevó a cabo para el entrenamiento y evaluación de los modelos usados para la identificación de las partes del rostro *ssd inception v2 coco* (SSD) y *faster rcnn inception v2 coco* (FASTER).

#### 3.3.1. Entrenamiento

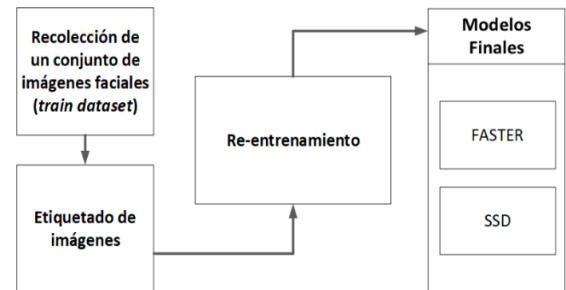
El hardware y software requerido para el entrenamiento de *TensorFlow* se muestran en la Tabla 3.

**Tabla 3.** Características de hardware y software.

Características	Especificaciones
Ram	16 GB , 2400 MHz
Cpu	4 núcleos, 4 hilos, 3.0 MHz, 6MB en caché
Gpu	6GB , GDDR5, 1280 núcleos cuda, 1506 MHz
Disco SSD	120 GB, Lectura 500 MB/s, Escritura 450MB/s
SO	Windows 10 – 1809 x64
Tensorflow	v1.12

Para lograr la compatibilidad de *TensorFlow* con la *Graphics Processing Unit* (GPU), se utilizaron versiones específicas de: Anaconda v1.9.6, Python v3.7, Protobuf v3.4.0 y OpenCV v3.4.2. De la misma forma, se utilizaron las versiones de *Compute Unified Device Architecture* (CUDA) y CUDNN v9.0 [29].

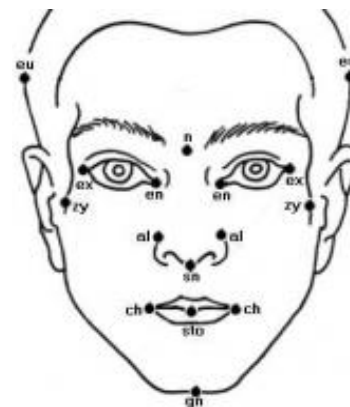
Después de cumplir los requerimientos de hardware y software, se siguió el diagrama de la Figura 2.



**Figura 2.** Etapas del entrenamiento.

La recolección de imágenes para el entrenamiento (*train dataset*) se construyó utilizando 1050 imágenes de rostros con diferentes rasgos étnicos tomadas de los *datasets* Helen [30] y Figshare [31] respectivamente. La postura de los individuos en las imágenes es erguida, sin objetos que obstaculicen el rostro, con expresiones faciales que varían de neutro a alegre. La resolución de las imágenes fue de 1920 x 1280, en formato JPG de 24 bits de profundidad (16.7 millones de combinaciones RGB) [32] y con un peso promedio de 1.3 MB.

Las imágenes del *train dataset* pasaron por el etiquetado, que consiste en marcar los puntos antropométricos y etiquetarlos de acuerdo con cuatro clases: cejas, ojos, nariz y boca mostrados en la Figura 3. Este proceso se llevó a cabo de forma manual y tomando en cuenta la antropometría médica facial.



**Figura 3.** Puntos antropométricos faciales [33].

Las descripciones de los puntos antropométricos definidos anteriormente se muestran en la Tabla 4.

**Tabla 4.** Medidas antropométricas faciales [33].

Región	Nombre de la medida
Ojos	Longitud de las fisuras palpebrales derecha e izquierda (ex - en)
Nariz	Anchura de la nariz (al - al), altura de la nariz (n - sn)
Labios y boca	Altura del labio superior (ls - sto), altura del labio inferior (sto - li), anchura de la boca (ch - ch)

Los modelos seleccionados para la identificación de las partes del rostro fueron: SSD y FASTER. Estos dos modelos son usados con frecuencia en la detección e identificación de objetos en una imagen. Además, son modelos pre-entrenados basados en redes neuronales convolucionales [34].

Según el repositorio oficial de *TensorFlow*, ambos modelos fueron entrenados sobre el conjunto de imágenes *COCO* [35]. Las características más relevantes de los modelos empleados se muestran en la Tabla 5 [36].

**Tabla 5.** Características de los modelos.

Característica	SSD	FASTER
Dataset	COCO	COCO
Red	CNN	CNN
Velocidad x Iteración	Depende del hardware	Depende del hardware
Salida	Cajas / Máscaras	Cajas / Máscaras
TotalLoss Recomendado	(1<0<2) %	(0≤0.05) %

Los modelos seleccionados fueron re-entrenados usando el *train dataset* propio

hasta lograr que el *TotalLoss*, se mantenga dentro del rango recomendado [29].

### 3.3.2. Evaluación

Para medir el rendimiento de los modelos implementados, se seleccionó una submuestra (*test dataset*) de los individuos encuestados, con el fin de tomar sus medidas antropométricas faciales como ancho y alto de: cejas, ojos, nariz y boca. Estas medidas se contrastaron con la información entregada por los modelos.

La submuestra se calculó a través de la Ecuación (1), con  $N$  igual a 430,  $p$  y  $q$  igual a 0.5,  $Z$  igual a 90 y  $e$  igual a 10%. El tamaño resultante de la submuestra fue de 60 individuos distribuidos de la siguiente manera: 10 afro-ecuatorianos, 30 mestizos, 10 indígenas y 10 blancos (europeo-descendientes).

La herramienta para obtener las medidas en cm fue un calibrador de vernier o pie de rey, de precisión 0.05 mm.

Para validar la biometría obtenida a través de la aplicación de los modelos, con los medidos, se utilizó la desviación estándar, para verificar la dispersión de los datos con relación a la media [37]. Además, se verificó que la relación de aspecto no afecte la determinación de las medidas, por lo cual se aplicó un factor de corrección ( $fc$ ) (2) a cada coordenada ( $x, y$ ) de cada clase. El factor de corrección se calculó usando el método de mínimos cuadrados [37].

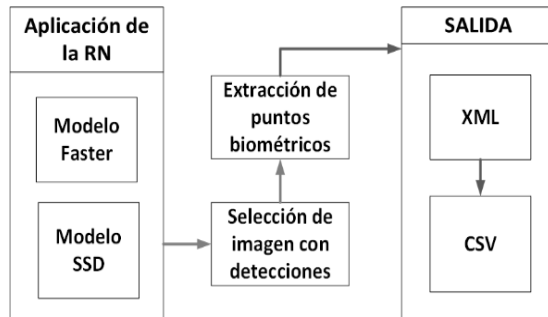
$$(x_i, y_i) = (x_i, y_i) \pm fc \quad (2)$$

### 3.4. Extracción y almacenamiento de los metadatos biométricos del rostro.

La última etapa de la metodología consiste en la extracción y almacenamiento de los

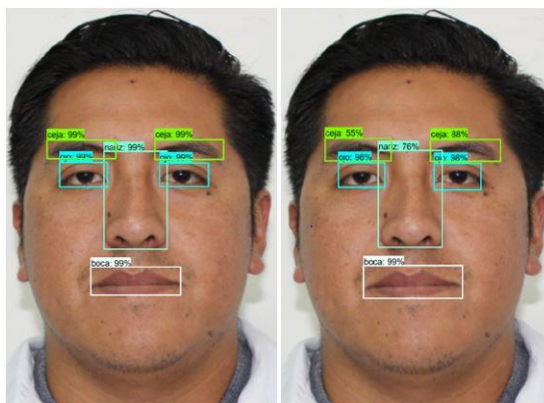
metadatos de cada una de las imágenes capturadas. Los metadatos se extrajeron de las respuestas de la encuesta y las biométricas se calcularon a partir de la información entregados por la RN.

Para la extracción de las medidas antropométricas del rostro se siguió el diagrama de la Figura 4.



**Figura 4.** Detección de las clases.

Como se puede observar en la Figura 4, el proceso comienza con la ejecución de los modelos, los cuales devuelven las imágenes con la detección de cada parte del rostro identificada, tal y como se muestra en la Figura 5.

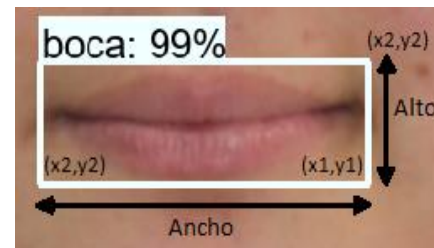


**Figura 5.** Detección de clases FASTER vs SSD.

La ejecución de las modelos genera dos imágenes, una por cada una, de las cuales se seleccionó aquella que cuente con todas las partes del rostro detectadas y con el mayor porcentaje de identificación de cada clase.

Para extraer las biométricas del rostro se realizó un procedimiento manual de re-etiquetado de imágenes (Figura 6), en el que se marcan las coordenadas  $(x_1, y_1)$  y  $(x_2, y_2)$ , para calcular el ancho y alto de cada clase detectada.

Las coordenadas en pixeles de cada detección fueron almacenadas en un archivo con formato *XML* (*Extensible Markup Language*) y este a su vez se transformó en formato *CSV* (*Comma-Separated Values*), para un mejor tratamiento de los datos.



**Figura 6.** Representación de la biometría.

Para el cálculo del ancho y alto, se utilizó (3) [38].

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

Por último, se transformó, tanto el ancho y el largo en pixeles a centímetros (4) [39]. Es importante mencionar que los modelos generan imágenes de 100 ppp (pixeles por pulgada) este valor se debe considerar en la fórmula de transformación.

$$cm = \frac{pixel}{100} \times 2.54 \quad (4)$$

### 3.5. Resultados y Discusiones

Este apartado describe los resultados a partir de la ejecución de la encuesta y la extracción de las biométricas del rostro.

La muestra del *dataset* para una población de 17.273.683 (estimación de la población ecuatoriana al 2019 [40]), tuvo

un tamaño de 450 individuos. La Tabla 6 resume la distribución étnica ponderada de la muestra.

**Tabla 6.** Muestra étnica ponderada.

Etnia	Muestra ponderada
Mestizos	300
Afro-ecuatorianos	50
Indígenas	50
Blancos	50
<b>TOTAL</b>	<b>450</b>

Se realizó un total de 520 encuestas, de las cuales se llevó a cabo, un filtrado de estas, para eliminar aquellas, que no cumplieron con los parámetros mínimos establecidos en la Sección B. De forma que el *dataset* final [41], se estructuró, agrupando las imágenes en 4 subconjuntos (etnias) de acuerdo a la auto-identificación de cada individuo en la encuesta. Los metadatos fueron compilados en un archivo CSV.

De esta forma, el *dataset* final está compuesto de un total de 430 imágenes, con sus respectivos metadatos, además cuenta con 24 atributos, uno por cada pregunta realizada al encuestado, en la Figura 7, se presenta un resumen de los atributos étnicos de las imágenes, que componen el *dataset* final.

Atributo	Tipo
Afro-Americano_porcentaje	:1 Categórica
Ancestros	:1 Categórica,Numérica
Apellido_Materno	:1 Categórica,Texto Libre:
Apellido_Paterno	:1 Numérica
Autorización	:1 Texto Libre
Ciudad_de_Nacimiento	:1
Color_de_Cabello	:1
Color_de_Ojos	:1
Color_de_Piel	:1
Edad	:1
Europeo_porcentaje	:1
Forma_de_Cejas	:1
Forma_de_Labios	:1
Forma_de_Mentón	:1
Forma_de_Nariz	:1
Forma_de_Orejas	:1
Forma_de_Rostro	:1
Género	:1
ID	:1
Marca_Temporal	:1
Nativo_Americano(Indígena)_porcentaje:	
Provincia_de_Nacimiento	:1
Tipo_de_Cabello	:1
Tipo_de_Ojos	:1

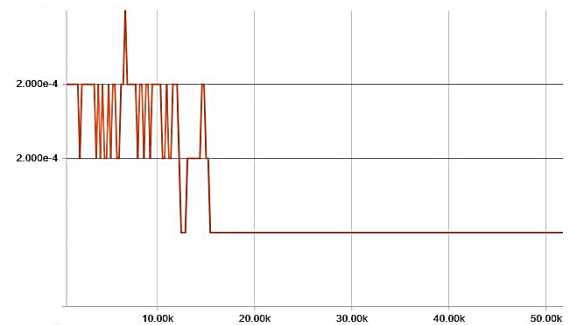
  

opciones	Tipo
0-100%	:7
1,2,3,4	:3
Afro-Americana,Europea,Nativo Americana (Indígena)	:1
Alomado,Cóncavo,Convexo,En "S",Recto,Sinuso	:1
Arqueadas,Rectilíneas,Sinuosas,Oblicuas Internas,oblicuas Externas,Asimétricas	:1
Delgados,Medianos,Gruesos y Abultados	:1
Diamante,Triangular,circular,óvalo,Cuadrada,Rectangular	:1
Hundidos,Saltones,Alargados,oblicuos Internos,oblicuos Externos	:1
Marrón Oscuro,Verde,Ambar,Azul,Avellana	:1
Masculino,Femenino,other	:1
Negro,castaño,Rubio,Pelirrojo	:1
Ondulado,Rizado,Lacio,Afro	:1
Ovoídes,Triangular,Rectangular,circular	:1
Partido,Agudo,Cuadrado,Plano,Arqueado,Fugitivo,Saliente	:1
Sí,No	:1

**Figura 7.** Resumen de atributos de encuesta.

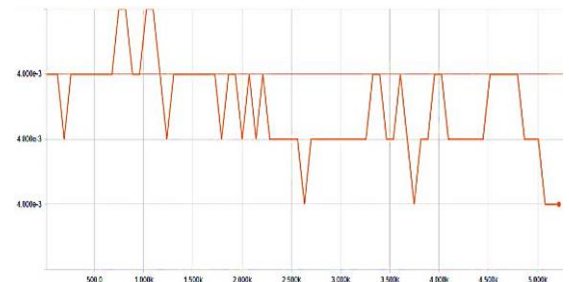
El hiperparámetro *learning rate* (*lr*) controla directamente el tiempo total de entrenamiento de cada modelo, es decir, un *lr* demasiado bajo, aumenta el tiempo en que el modelo se adapta a un problema mientras que un *lr* demasiado alto puede hacer que el modelo se adapte más rápido, pero genera soluciones no viables [42]. En el caso de los modelos al ser pre-entrenados este hiperparámetro estuvo configurado por defecto.

En el modelo FASTER se usó un *lr* diferente a medida que aumentan las iteraciones, es decir, a partir de la iteración cero el modelo utilizó un *lr* de 0.00019 y al llegar a la iteración 90k usó un *lr* de  $2.000e^{-3} \sim 0.0366$  (Figura 8), esto implicó que el modelo se adaptó al problema en un periodo más corto de tiempo [43], obteniendo un *TotalLoss* de 0.02647% dentro del rango aconsejable (Figura 10).

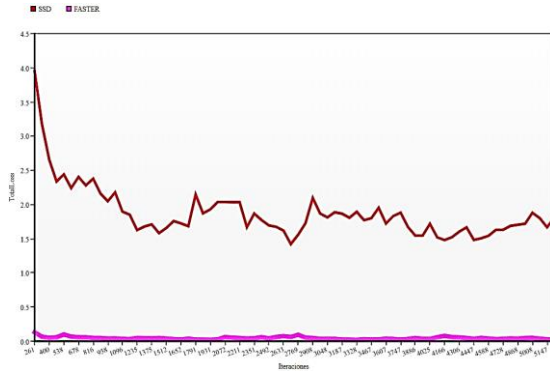


**Figura 8.** Learning rate FASTER.

A diferencia del modelo FASTER, el modelo SSD utilizó un único valor de *lr* de 0.00400 para todas las iteraciones (Figura 9), obteniendo un Total Loss de 1.808% dentro del rango recomendado (Figura 10).



**Figura 9.** Learning rate SSD.



**Figura 10.** TotalLoss de modelos utilizados.

A través del re-entrenamiento llevado a cabo, los modelos se ajustaron a la detección de las clases objetivo (cejas, ojos, nariz, boca).

Por otra parte, el *TotalLoss* alcanzado, cantidad de iteraciones y tiempo total del re-entrenamiento se describen en la Tabla 7.

**Tabla 7.** Re-entrenamiento de modelos.

Indicador	FASTER	SSD
Iteraciones	100333	105642
Tiempo total	1d 11h 2m 30s	3d 14h 4m 16s
Total Loss alcanzado	0.02647%	1.808%

Con el cumplimiento de los parámetros descritos en la etapa de entrenamiento se garantiza que las clases sean detectadas correctamente, logrando un porcentaje alto de detección e identificación superior al 95% para FASTER y 70% SSD.

Al contrastar la información extraída del encuestado, con los datos obtenidos a través de la extracción de la biometría facial, arrojó, dispersión de estos, tal como se muestra en la Figura 11, cabe mencionar que la imagen mostrada hace referencia solo a una clase en particular (ceja), debido a que los resultados obtenidos para las otras clases fueron similares.

El cálculo de la biometría se realizó en base a tres coordenadas, siendo  $(x1, y1)$

común para el alto y ancho (Figura 6) por lo cual, el factor de corrección fue aplicado a cada una de las coordenadas de las seis detecciones, es decir, en el caso específico de las cejas se cuenta con dos detecciones: Detección Ceja 1 (DC1) y Detección Ceja 2 (DC2), de igual manera, en el caso de los ojos: Detección Ojo 1 (DO1) y Detección Ojo 2 (DO2), en general los factores de corrección aplicados a cada clase se describen en la Tabla 8.

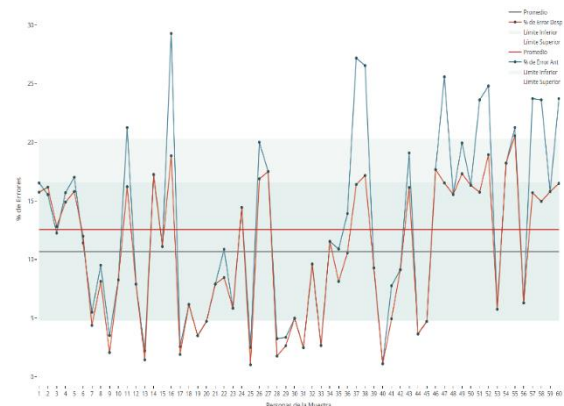
**Tabla 8.** Factores de corrección (cm).

Objeto		Alto		Ancho
		$(x1, y1)$	$(x2, y2)$	$(x2, y2)$
Cejas	DC1	0.00523	0.00798	0.00088
	DC2	0.00057	0.00077	0.00035
Ojos	DO1	0.00602	0.00107	0.00156
	DO2	0.00080	0.00150	0.00055
Nariz		0.00123	0.00656	0.00057
Boca		0.00931	0.01182	0.00316

Los errores tras la aplicación de los modelos y posterior a la aplicación del factor de corrección para la clase ceja, se muestran en la Tabla 9.

**Tabla 9.** Error cuadrático medio.

	Porcentaje de error por clase	
	Alto	Ancho
<b>Cejas</b>	3.67%	7.60%
<b>Ojos</b>	3.73%	4.62%
<b>Nariz</b>	9.89%	7.27%
<b>Boca</b>	3.87%	7.46%



**Figura 11.** Errores de clase ceja, antes y después del factor de corrección.

### 3.6. Conclusiones y Prospectivas

Este trabajo logró la construcción de un conjunto de imágenes faciales con metadatos biométricos y rasgos étnicos de ecuatorianos, a través de la aplicación de una encuesta y el uso de algoritmos de ML.

La efectividad de la biometría de las partes del rostro dependió en gran parte de la calidad de la imagen, que, a su vez, estuvo sujeta a la distancia focal y el tiempo de exposición usados en la captura de la fotografía (1/60), una distancia focal muy grande o baja producirá el alargamiento o ensanchamiento del rostro afectando la obtención de las biométricas faciales. Por otro lado, una distancia focal de 24 mm guarda relación con la perspectiva y el ángulo de visión observado por el ojo humano.

Si las condiciones de la imagen no son las adecuadas, es decir, la calidad de la imagen es pobre, con mala iluminación, arrugas en el fondo y baja nitidez, a pesar de tener una RN correctamente entrenada y robusta, las detecciones serán bajas e imprecisas. Por tanto, es menester contar con una fase de corrección de imperfecciones sobre la imagen, antes de pasar a la fase de evaluación del modelo.

Los modelos FASTER y SSD seleccionadas en este trabajo son dos de los mejores modelos utilizados para la detección y clasificación de objetos, debido a que presentan un mayor rendimiento/velocidad en comparación con otros modelos. En general la efectividad de los modelos empleados en la detección de las partes del rostro fue superior al 70%. Siendo el modelo FASTER superior a SSD con niveles de detección superiores al 90%.

Las clases detectadas con mayor precisión fueron: los ojos y la boca,

mientras que la que clase que presentó mayores inconvenientes en la detección fueron las cejas, debido principalmente a su estructura, tamaño, color y forma, lo que provocó diferencias significativas entre las medidas de ancho y alto tomadas por el calibrador. Para mejorar la detección de esta parte del rostro se recomienda tomar imágenes en una postura semi-lateral.

Otro aspecto relevante que se debe considerar en el cálculo de las biométricas faciales es el factor de corrección, ya que, como se muestra en los resultados, sin la aplicación del factor de corrección los errores alcanzados estuvieron cercanos al 18%. Al aplicar el factor de corrección el error promedio disminuyó significativamente alcanzado un promedio menor al 10%.

La clase que presentó mayor precisión en el cálculo de sus medidas de alto y ancho corresponde a la de los ojos, alcanzado un error cuadrático medio menor al 5%. Por el contrario, la clase que presenta menor precisión fue la perteneciente a la nariz, específicamente en la medida de la altura, con un error cuadrático medio cercano al 10%. Esto podría deberse principalmente a la relación de aspecto de la imagen, por lo que sería recomendable recalcular el factor de corrección para las medidas verticales del rostro.

Generar una nueva versión del dataset, tomando en cuenta, todas las etnias correspondientes al Ecuador.

Llevar a cabo un procedimiento para el tratamiento y mejora de la imagen mediante la utilización de herramientas de procesamiento digital de las mismas.

## Referencias

- [1] G. Webb, M. Pazzani y D. Billsus, «Machine Learning for User Modeling,» *User Modeling and User-Adapted Interaction*, vol. 11, n° 1, pp. 19-29, 2001.
- [2] E. Ikonomakis, S. Kotsiantis y V. Tampakas, «Text Classification Using Machine Learning Techniques,» *WSEAS Transactions on Computers*, vol. 4, pp. 966-974, 2005.
- [3] X. Lu y A. K. Jain, *Ethnicity Identification from Face Images*, Michigan, 2004.
- [4] G. Guo y G. Mu, «Joint estimation of age, gender and ethnicity: CCA vs. PLS,» de *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, 2013.
- [5] H. Ding, D. Huang, Y. Wang y L. Chen, «Facial ethnicity classification based on boosted local texture and shape descriptions,» de *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, Shanghai, 2013.
- [6] N. Narang y T. Bourlai, «Gender and ethnicity classification using deep learning in heterogeneous face recognition,» de *2016 International Conference on Biometrics (ICB)*, Halmstad, 2016.
- [7] G. Zaccane, *Getting Started with TensorFlow*, 2016.
- [8] M. Singh, S. Nagpal, M. Vatsa, R. Singh, A. Noore y A. Majumdar, «Gender and ethnicity classification of Iris images using deep class-encoder,» de *2017 IEEE International Joint Conference on Biometrics (IJCB)*, Denver, 2017.
- [9] A. Zarei y D. Mou, «Artificial Neural Network for Prediction of Ethnicity Based on Iris Texture,» de *2012 11th International Conference on Machine Learning and Applications*, Boca Raton, 2012.
- [10] Z. Li, X. Duan, Q. Zhang, C. Wang, Y. Wang y W. Liu, *Multi-ethnic Facial Features Extraction based on Axiomatic Fuzzy Set Theory*, 2017.
- [11] A. Japar, A. Nugroho, J. Purnama y M. Galinium, *Evaluation of feature extraction algorithm for multi-ethnic facial sketch recognition*, 2017.
- [12] S. Masood, S. Gupta, A. Wajid, S. Gupta y M. Ahmad, «Prediction of Human Ethnicity from Facial Images Using Neural Networks,» 2018, pp. 217-226.
- [13] NIST, «National Institute of Standards and Technology,» 25 Enero 2011. [En línea]. Available: <https://www.nist.gov/programs-projects/face-recognition-technology-feret>. [Último acceso: 21 Junio 2019].
- [14] G. P. Mabuza-Hocquet, F. Nelwamondo y T. Marwala, «Ethnicity Prediction and Classification from Iris Texture Patterns: A Survey on Recent Advances,» de *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, 2016.
- [15] X. Duan, C. Wang, X. Liu, Z. Li, J. Wu y H. Zhang, «Ethnic Features extraction and recognition of human faces,» de *2010 2nd International Conference on Advanced Computer Control*, Shenyang, 2010.
- [16] S. Hosoi, E. Takikawa y M. Kawade, «Ethnicity estimation with facial images,» de *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, South Korea, 2004.
- [17] S. K. Jilani, H. Ugail, A. M. Bukar, A. Logan y T. Munshi, «A Machine Learning Approach for Ethnic Classification: The British Pakistani Face,» de *2017 International Conference on Cyberworlds (CW)*, Chester, 2017.
- [18] M. Oncins de Frutos, «Instituto Nacional de Seguridad y Salud en el Trabajo,» [En línea]. Available: <https://bit.ly/2FrAvmJ>. [Último acceso: 16 Enero 2019].
- [19] A. E. Merlo Rivadeneira, «Policía Nacional del Ecuador,» 2018. [En línea]. Available: <https://www.policiaecuador.gob.ec>. [Último acceso: 13 Diciembre 2018].
- [20] F. Olivo Cerda y P. Vásquez Ñaupari, «Policía Nacional del Ecuador,» 2018.



- [En línea]. Available: <https://www.policiaecuador.gob.ec>. [Último acceso: 13 Diciembre 2018].
- [21] J. V. Rodríguez Cuenca, *La Antropología Forense en la Identificación Humana*, Bogotá: Guadalupe Ltda, 2004.
- [22] D. J. Kang, S. G. Ryoo, J. Y. Kim, C. Y. Kim y Y. S. Seo, «Apparatus and method for detecting speaking person's eyes and face». Patente 6.611.613, Agosto 2003.
- [23] J. Casas Anguita, J. R. Repullo Labrador y J. Donado Campos, «La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I),» *Atención Primaria*, vol. 31, n° 8, pp. 527-538, 2003.
- [24] P. López Roldán y S. Fachelli, *Metodología de la Investigación Social Cuantitativa*, Barcelona, 2015.
- [25] B. Villacís y D. Carrillo, «País atrevido: la nueva cara sociodemográfica del Ecuador,» *Instituto Nacional de Estadística y Censos (INEC) - Analitika*, 2012.
- [26] L. M. Dikovskiy Riobóo, *Estadística Básica*, Estelí, 2008.
- [27] A. Morillas, «Plataforma de Apoyo a la Docencia Presencial,» 20 Abril 2010. [En línea]. Available: <https://bit.ly/2Frtbrg>. [Último acceso: 26 Enero 2019].
- [28] C. Blanco, *Encuesta y Estadística: Métodos de Investigación Cuantitativa en Ciencias Sociales y Comunicación*, Córdoba: Brujas, 2011.
- [29] L. Vladimirov, «TensorFlow Object Detection API Tutorial,» [En línea]. Available: <https://bit.ly/2KGqI5i>. [Último acceso: 16 Febrero 2019].
- [30] V. Le, J. Brandt, Z. Lin, L. Boudev y T. Huang, «Helen Dataset,» [En línea]. Available: <http://www.ifp.illinois.edu/~vuongle2/helen/>. [Último acceso: 1 Julio 2019].
- [31] L. DeBruine y B. Jones, «Figshare: Face Research Lab London Set,» [En línea]. Available: [https://figshare.com/articles/Face\\_Research\\_Lab\\_London\\_Set/5047666](https://figshare.com/articles/Face_Research_Lab_London_Set/5047666). [Último acceso: 1 Julio 2019].
- [32] L. Moreno, «Desarrolloweb,» [En línea]. Available: <https://desarrolloweb.com/articulos/1778.php>. [Último acceso: 10 Julio 2019].
- [33] F. Prieto, *Antropometría Facial a Partir de Múltiples Vistas: Una propuesta metodológica*, Manizales, 2008.
- [34] J. Xu, «Deep Learning Italia,» 8 Junio 2018. [En línea]. Available: <https://bit.ly/2RvU6Hr>. [Último acceso: 15 Marzo 2019].
- [35] C. O. i. Context, «COCO: Common Objects in Context,» [En línea]. Available: <http://cocodataset.org/#home>. [Último acceso: 12 Abril 2019].
- [36] TensorFlow, «Github,» [En línea]. Available: <https://bit.ly/2EtFjJP>. [Último acceso: 22 Marzo 2019].
- [37] R. Walpole, R. Myers, S. Myers y K. Ye, *Probabilidad y Estadística para Ingeniería y Ciencias*, Naucalpan de Juárez: Pearson Educación de México, 2012.
- [38] C. Lehmann, *Geometría Analítica*, Nueva York: Limusa, 1980.
- [39] Digital Foto Red, «Digital Foto Red,» [En línea]. Available: <https://bit.ly/2FsqKF1>. [Último acceso: 25 Junio 2019].
- [40] INEC, «INEC: Buenas cifras, mejores vidas,» [En línea]. Available: <https://bit.ly/2l05wRK>. [Último acceso: 25 Junio 2019].
- [41] J. Avilés, H. Toapanta, P. Morillo y D. Vallejo-Huanga, «Figshare:Dataset of Ethnic Facial Images of Ecuadorian People,» 28 Junio 2019. [En línea]. Available: <https://doi.org/10.6084/m9.figshare.8266730>. [Último acceso: 3 Julio 2019].
- [42] J. Brownlee, «Machine Learning Mastery,» 25 Enero 2019. [En línea]. Available: <https://bit.ly/2Ng2dJB>. [Último acceso: 27 Junio 2019].
- [43] H. Zulkifli, «Towards Data Science,» 21 Enero 2018. [En línea]. Available: <https://bit.ly/2s4UAuK>. [Último acceso: 27 Junio 2019].