

Sistema de Reconocimiento de Emociones Faciales para medir el Grado de Satisfacción del Cliente mediante Redes Neuronales Convolucionales

Johan Robinson Veramendi Llaulle 1

e-mail: jveramendil@uni.pe

Arisa Samantha Vigo Rojas 2

e-mail: avigor@uni.pe

RESUMEN: *El presente trabajo mostrará cómo puede utilizarse la inteligencia artificial para medir el grado de satisfacción de clientes de una empresa de consumo minorista o de prestación de servicios a través de las emociones faciales. Para ello, se implementó un modelo de aprendizaje profundo con el uso de redes neuronales convolucionales. El proceso de entrenamiento del modelo utilizó conjunto de imágenes públicas y de elaboración propia, las cuales fueron preprocesadas para optimizar el modelo e identificar hasta 7 posibles emociones: enojo, disgusto, miedo, felicidad, tristeza, sorpresa y el estado neutro. La red neuronal fue programada utilizando los recursos de la librería TensorFlow y la arquitectura AlexNet. Se logró obtener una precisión del 71%. Finalmente se creó una aplicación para utilizar el modelo con el uso de una cámara web, en donde se captura la imagen y se muestra el resultado de la evaluación en tiempo real.*

PALABRAS CLAVE: Aprendizaje profundo, inteligencia artificial, emociones faciales, redes neuronales convolucionales, satisfacción de clientes.

ABSTRACT. *This paper will show how artificial intelligence can be used to measure the degree of customer satisfaction of a retail or service delivery company through facial emotions. To this end, a deep learning model was implemented with the use of convolutional neural networks. The model training process used a set of public and self-made images, which were pre-processed to optimize the model and identify up to 7 possible emotions: anger, disgust, fear, happiness, sadness, surprise and the neutral state. The neural network was programmed using the resources of the TensorFlow library and the AlexNet architecture. An accuracy of 71% was achieved. Finally, an application was created to use the model with the use of the webcam, where the image is captured, and the result of the evaluation is displayed in real time.*

Keywords. Deep learning, artificial intelligence, facial emotions, convolutional neuronal networks, customer satisfaction.

1 INTRODUCCIÓN

Las emociones son formas de comunicación del ser humano en respuesta a estímulos internos y externos. Sirven como medio de comunicación universal dada la facilidad de poder ser interpretadas por el ser humano debido a su naturaleza social. De todas las formas de comunicación no verbal, las expresiones faciales emocionales son las que más información ofrecen del estado emocional de otras personas. [1]

El proceso de análisis del comportamiento de una persona bien lo podría realizar una persona designada para tal efecto, invirtiendo tiempo y demás recursos durante el periodo que requiera para llegar a una conclusión sobre el comportamiento de la persona. La ventaja que supone esto es el juicio crítico humano que da valor agregado a la conclusión. Sin embargo, el proceso del reconocimiento de las emociones puede ser encargado a una inteligencia artificial haciendo uso de un modelo computacional de aprendizaje profundo.

Se plantea como hipótesis que el análisis de reconocimiento de emociones faciales con el uso de Inteligencia Artificial mejorará la medición del grado de satisfacción del cliente y por ende se plantea como objetivo principal mejorar la medición del grado de satisfacción del cliente.

2 MATERIALES Y MÉTODOS

2.1 BASE DE DATOS

El conjunto de datos utilizado fue el que perteneció al Desafío de Reconocimiento Facial de Kaggle de 2013 [2]. La base de datos contiene 35887 imágenes de personas y están catalogadas según la emoción que representan: enojo, disgusto, miedo, felicidad, tristeza, sorpresa y el estado neutro.

Las imágenes están en escala de grises y tienen una dimensión de 48*48 pixeles. El conjunto de datos contiene 3 columnas, una que describe la emoción, una que indica los pixeles en forma de cadena, y una columna que indica si es data de entrenamiento o testeo.

La distribución de las imágenes es la siguiente:

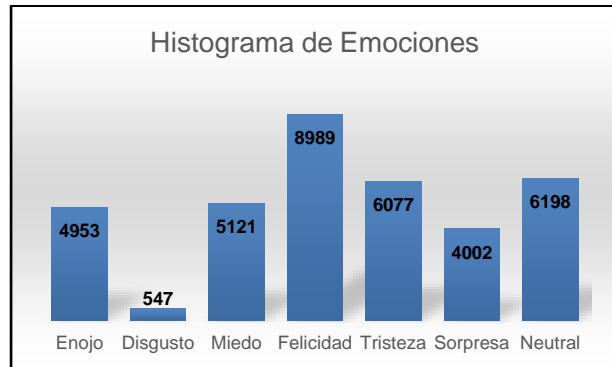


Figura 1. Distribución de imágenes totales

Del conjunto de imágenes totales, las utilizadas para entrenamiento fueron 28709, mientras que para el testeo se utilizaron 7179.

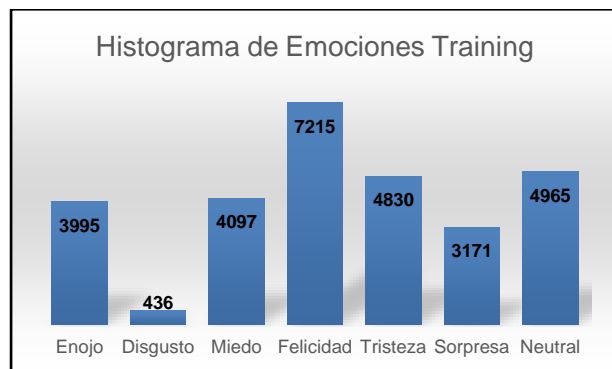


Figura 2. Histograma de data de entrenamiento.

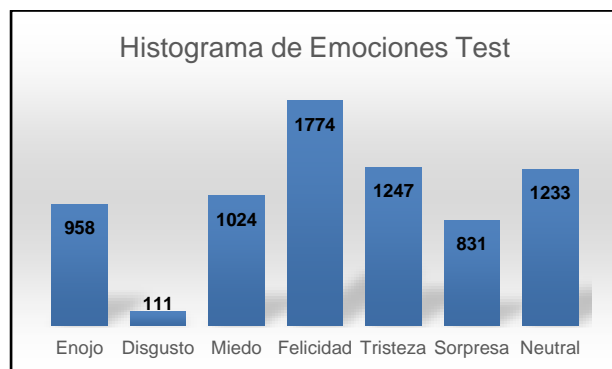


Figura 3. Histograma de data de testeo.

Para poder Visualizar las diferentes emociones, es necesario hacer un tratamiento a las imágenes, esto se puede hacer con el uso de las librerías Numpy y PIL, los pixeles en formato de imagen que se pueden obtener para las distintas emociones son:



Figura 4. Ejemplo de emociones encontradas en el conjunto de datos.

2.2 CONCEPTOS CLAVES

Para poder entender las redes neuronales, es necesario el uso de tres conceptos:

- Capas convolutivas.
- Capas de agrupamiento máximo.
- Capas totalmente conectadas.

2.2.1 CAPA CONVOLUTIVA

Una capa convolutiva es una matriz que representa a la imagen luego de haberle aplicado diferentes filtros. Los filtros permiten identificar patrones en la imagen que se recibe como entrada; estos pueden ser rasgos, bordes, entre otros.

La forma como se obtiene una capa convolutiva es la siguiente:

- Se recibe la imagen como matriz de dimensión $m \times m$, y se cuenta con una matriz que servirá como filtro.
- Se multiplican los valores de las celdas del filtro con las de una sección de la imagen de entrada, se suman los productos y se obtiene un nuevo valor que irá formando la nueva capa convolucionada. Las sumas producto se realizan mientras se recorre el filtro por toda la matriz.

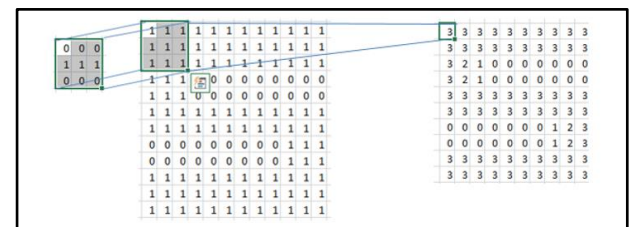


Figura 5. Aplicación de filtro de convolución.

2.2.2 CAPA DE AGRUPAMIENTO MÁXIMO

Una capa de agrupamiento máximo implica aplicar una operación de agrupamiento por el máximo valor a la capa convolutiva, y así obtener una nueva capa que resuma las características principales de la anterior.

La operación de agrupamiento máximo implica seleccionar una sección de dimensión $n \times n$ de una capa convolutiva de dimensión $m \times m$, donde $n < m$, y obtener el mayor valor de la matriz $n \times n$.

Se recorre toda la capa convolutiva con la matriz de sección $n \times n$, y se obtiene una nueva capa denominada Capa de agrupamiento. La capa agrupada tiene como objetivo reducir la dimensión de la matriz original manteniendo las características importantes que se obtuvieron como resultado de los filtros.

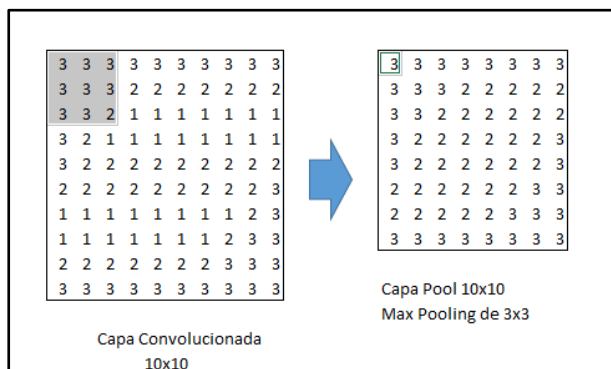


Figura 6. Ejemplo de agrupamiento máximo.

2.2.3 CAPA TOTALMENTE CONECTADA

Es la capa que conecta las neuronas artificiales y permite realizar el proceso de clasificación. La capa densa de la red está inspirada en la manera en que las neuronas transmiten señales a través del cerebro. Se necesita una gran cantidad de características de entrada y funciones de transformación a través de capas conectadas con pesos entrenables. [3]

2.3 MODELO DE SOLUCIÓN

Se deberá tener el conjunto de imágenes que sirva como entrenamiento debidamente catalogado. Para el desarrollo del modelo, serán 7 las emociones que se utilizarán: miedo, felicidad, tristeza, asombro, enojo y el estado neutral.

Como primer punto, se debe conseguir el conjunto de datos óptimos, para ello es necesario hacer un tratamiento de la data que se recibe como entrada y seleccionar los datos de entrenamiento y de testeo. Dado que la data de entrada contiene imágenes de diferentes

personas en diferentes posiciones es necesario hacer un filtro de aquellas imágenes con las que realmente se puede trabajar.

Para realizar este filtro se hace un recorte de los rostros con el uso de Haar-Cascade. Haar-cascade es un recurso de la librería OpenCV, y permite identificar rostros a partir de una imagen que recibe como entrada. La imagen es recortada para poder trabajar solo con el rostro de la persona en análisis (ver figura 7).

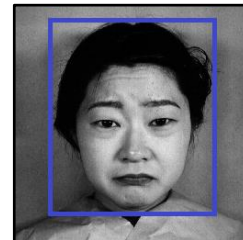


Figura 7. Reconocimiento facial utilizando Haar Cascade

Una vez obtenidas las imágenes obtenidas, se almacenan dos arreglos del formato Numpy (Numpy es una librería de Python que también se utilizará). Un arreglo contendrá las imágenes filtradas, y el otro arreglo contendrá las etiquetas de las respectivas emociones.

El segundo paso es construir la red neuronal en la que se utilizará la data de entrenamiento. Para este proceso se utilizará como referencia la arquitectura AlexNet. [4]

La red neuronal tendrá la siguiente arquitectura:

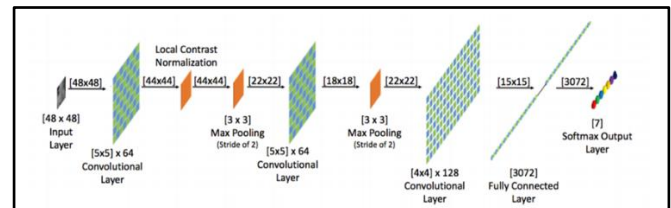


Figura 8. Arquitectura de la red neuronal. [5]

```
def build_network(self):
    # Smaller 'AlexNet'
    # https://github.com/tflearn/tflearn/blob/master/examples/images/alexnet.py
    print('[+] Building CNN')
    self.network = input_data(shape = [None, SIZE_FACE, SIZE_FACE, 1])
    self.network = conv_2d(self.network, 64, 5, activation = 'relu')
    self.network = local_response_normalization(self.network)
    self.network = max_pool_2d(self.network, 3, strides = 2)
    self.network = conv_2d(self.network, 64, 5, activation = 'relu')
    self.network = max_pool_2d(self.network, 3, strides = 2)
    self.network = conv_2d(self.network, 128, 4, activation = 'relu')
    self.network = dropout(self.network, 0.3)
    self.network = fully_connected(self.network, 3072, activation = 'relu')
    self.network = fully_connected(self.network, len(EMOTIONS), activation = 'softmax')
    self.network = regression(self.network,
        optimizer = 'momentum',
        loss = 'categorical_crossentropy')
    self.model = tflearn.DNN(
        self.network,
        checkpoint_path = SAVE_DIRECTORY + '/emotion_recognition',
        max_checkpoints = 1,
        tensorboard_verbose = 2
    )
    self.load_model()
```

Figura 9. Código de la red neuronal

Como se aprecia en la figura 8, la red neuronal contendrá 2 capa convolucionales, 2 capas de agrupamiento máximo y una capa totalmente conectada que permitirá realizar la clasificación de la emoción. Luego de cada capa convolutiva, se aplicará una capa de agrupamiento máximo o capa max pool. La capa max pool se encargará de reducir las dimensiones de la matriz, de modo que dicha matriz reducida logre representar a la matriz anterior tomando y conservando sus atributos principales.

El proceso de entrenamiento se realizó en una laptop Lenovo ideapad 310 con un procesador Corei5 de quinta generación con 12GB de memoria RAM. La red fue construida utilizando los recursos de la librería Tensor Flow.

El paso final es obtener una interfaz que permita correr el modelo y evaluarlo utilizando la cámara web como fuente de entrada. Esta aplicación debe mostrar las probabilidades de la emoción, así como la etiqueta correspondiente a la emoción determinada en tiempo real.

2.4 DISEÑO DEL SISTEMA EXPERTO

Se diseñará un sistema experto que permita la aplicación de la red neuronal convolutiva a través de una aplicación implementada en Python.

El sistema experto contendrá dos principales módulos, uno para el reconocimiento de las emociones y la satisfacción del cliente, y otro para poder visualizar estadísticos acerca de la información recolectada.

El sistema permitirá acceder a la cámara web con que esté trabajando el computador, y procesará las imágenes de los rostros de las personas que estén frente a ella. El sistema además guardará la información generada durante las interacciones con las personas con el fin de poder generar estadísticos y consultas acerca de las emociones de los clientes relacionados al proceso de atención y satisfacción.

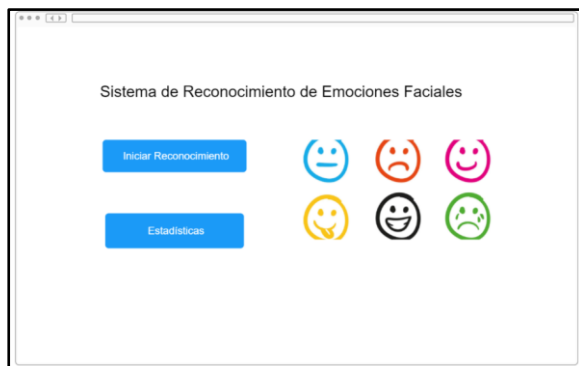


Figura 9. Interfaz principal

2.4.1 ARQUITECTURA DE LA RED

La aplicación será una aplicación de escritorio desarrollada en Python, utilizando como motor de base de datos MySQL.

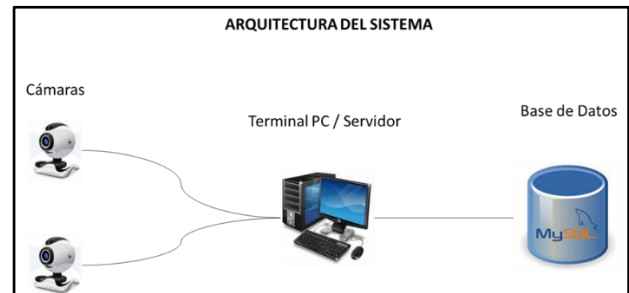


Figura 10. Arquitectura del sistema experto.

El sistema capturará las imágenes con las emociones de las personas a través de cámaras instaladas y dichas imágenes en video serán procesadas por el modelo de reconocimiento ya entrenado. Finalmente interactuará con una capa adicional que almacena las emociones reconocidas en una Base de Datos MySQL para posteriormente hacer el análisis de la información recolectada.

2.4.2 MODELO DE DATOS

Para poder generar el módulo de reportes estadísticos y poder almacenar la información recolectada durante la interacción con el cliente, se generó un modelo de datos sencillos que permita el almacenamiento (ver figura 11).

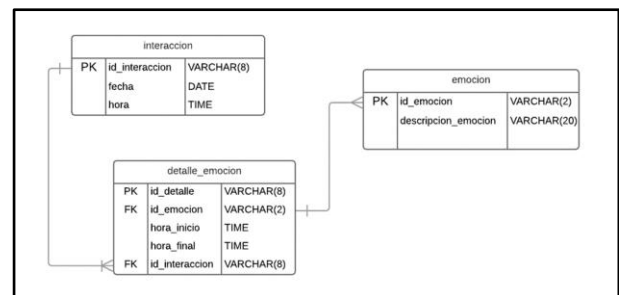


Figura 11. Modelo de datos de la aplicación.

3 RESULTADOS

Se realizaron 10 iteraciones con la data de entrenamiento, con cerca de 11250 imágenes válidas que fueron según el filtro realizado previamente. Se obtuvo un accuracy de 0.5316.

Los pesos entrenados en la red neuronal generan un archivo de cerca de 400 MB de espacio. Este archivo es necesario para que la red pueda clasificar las emociones.

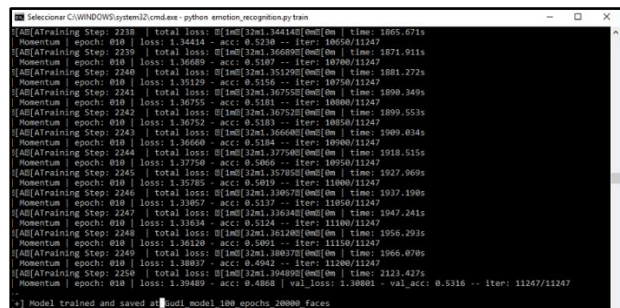


Figura 12. Entrenamiento de la red.

Al ejecutar la aplicación, se obtuvieron las siguientes capturas:

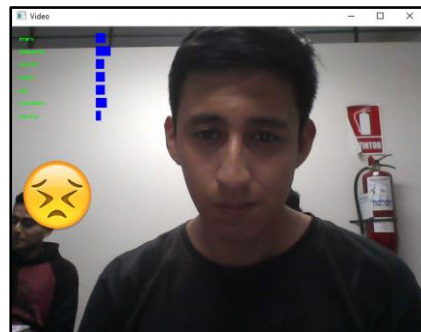


Figura 13. Captura de emoción de disgusto.

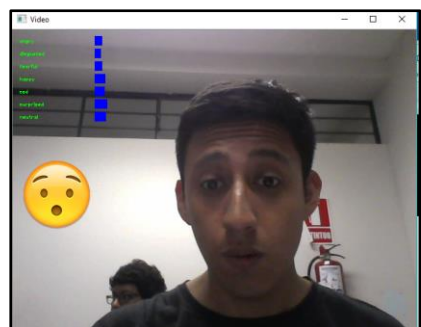


Figura 14. Captura de emoción de asombro.

4 DISCUSION

En base a los resultados obtenidos, se logró crear un sistema que a través de un modelo de inteligencia artificial logra identificar la emoción más probable.

El sistema además permite identificar las emociones en tiempo real con una precisión del 71% según los resultados obtenidos a partir de la comparación de los datos de testeo con los obtenidos por el modelo. Durante el proceso de pruebas además se identificó como necesario una correcta iluminación del ambiente para una clasificación más exacta.

Los datos obtenidos durante la evaluación son almacenados para un posterior análisis acerca del comportamiento durante un cierto periodo de tiempo.

Este análisis permitirá finalmente a los negocios que estudien a sus clientes, determinar cómo se sintió durante todo el proceso de atención y si además tuvo algún cambio de ánimo notorio producto de la atención recibida.

5 CONCLUSIONES

En base a los resultados obtenidos, se logró crear un sistema que a través de un modelo de inteligencia artificial logre identificar la emoción más probable.

- El sistema permite identificar hasta 7 emociones en tiempo real con una precisión del 71%
- El sistema ayudará a los negocios a identificar la satisfacción de los clientes de manera automática y más precisa con los indicadores que se obtengan del sistema.
- Las emociones faciales permiten obtener un indicador más confiable si se compara con lo que pueda expresar un cliente a través de una encuesta.
- El sistema es fácil de implementar en cualquier negocio, dado que solo se necesita de una cámara web y una pc.
- El sistema puede ser escalable y ser llevado a la nube sin problemas dada la arquitectura con la que fue diseñada.
- Puede aplicarse y desarrollarse el reconocimiento de emociones faciales en muchos más ámbitos, tales como reproducción de música según el estado de ánimo, recomendaciones de películas, entre otros.

6 REFERENCIAS

- [1] B. García-Rodríguez, A. F., "Procesamiento emocional de las expresiones faciales", REVISTA DE NEUROLOGÍA, 609-617. 2008
- [2] ICML 2013 Workshop in Challenges in Representation Learning, GA. "Dataset: Facial Emotion Recognition (FER2013)", 2013.
- [3] M. Castrill, D. Suarez, S. Hernandez y L. Navarro, «Face and Facial Feature Detection Evaluation,» Third International



Sistema de Reconocimiento de Emociones Faciales para medir el Grado de Satisfacción del Cliente mediante Redes Neuronales Convolucionales

Johan, Veramendi – Autor1
Arissa, Vigo – Autor2

Conference on Computer Vision Theory and Applications, VISAPP08, 2008.

- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc., 2012.