

**UNIVERSIDAD “MAYOR DE SAN SIMÓN”**

**FACULTAD DE CIENCIAS Y TECNOLOGÍA**

**Departamento de Informática y Sistemas**

**Aprendizaje No Supervisado**

**Sistema de Reconocimiento de Emociones**

**INTEGRANTES: Brandon Brayan Goytia Cano**

**Bryan Henry López Maldonado**

**Henry Gabriel Montaño Uriona**

**MATERIA: Ciencia de Datos y Machine Learning**

**DOCENTE: Patricia Erika Rodriguez Bilbao**

**Junio 2022**

**COCHABAMBA - BOLIVIA**

**ÍNDICE**

[**Introducción**](#_kqwdxnt0p6x1) **3**

[**Objetivos**](#_pfa9ozpqhw1j) **3**

[Objetivo general](#_z8k1ebwddce7) 3

[Objetivos específicos](#_c66l3h7srhql) 3

[**Marco teórico**](#_jybtm01tennl) **4**

[Emociones](#_z8qph25ma6ly) 4

[Reconocimiento de las emociones](#_ewovdmbbnw3c) 4

[Selección y extracción de características](#_h64glp58yyul) 4

[Aprendizaje No supervisado](#_z6tm4pkjrwh) 5

# 

# Introducción

Desde hace varios años, el campo de la inteligencia artificial ha tenido un crecimiento acelerado. Los avances de la inteligencia artificial han generado varias tecnologías que involucran la interacción humano-computadora que son de suma importancia. Vivimos en un mundo donde Siri Y Alexa están físicamente más cerca de nosotros que otros humanos.

Como seres humanos, el habla es una de las formas más naturales de expresarnos. Dependemos tanto de él que reconocemos su importancia cuando recurrimos a otras formas de comunicación como correos electrónicos y mensajes de texto donde a menudo usamos emojis para expresar las emociones asociadas a los mensajes. Dado que las emociones juegan un papel vital en la comunicación, la detección y el análisis de las mismas es de vital importancia en el mundo digital actual de la comunicación remota. La detección de emociones es una tarea desafiante, porque las emociones son subjetivas. No existe un consenso común sobre cómo medirlos o categorizarlos.

El reconocimiento de emociones del habla es una colección de metodologías que procesan y clasifican las señales de audios para detectar emociones en ella. Un modelo de este tipo puede encontrar uso en una amplia variedad de áreas de aplicación, como el asistente interactivo basado en voz o el análisis de conversación entre el agente y la persona que llama.

Desde una perspectiva de aprendizaje automático, el reconocimiento de emociones del habla es un problema de clasificación en el que una muestra de entrada (audio) debe clasificarse en unas pocas emociones. Por este motivo en el presente estudio, intentamos realizar un modelo basado en el aprendizaje no supervisado para detectar las emociones del hablante mediante el análisis de características acústicas de los datos de audio grabados.

# Objetivos

## Objetivo general

Desarrollar un modelo para reconocer emociones a partir del habla basado en el aprendizaje no supervisado.

## Objetivos específicos

* Analizar la transmisión, aplicación y manipulación de las señales de voz para poder utilizarlas en el modelo de reconocimiento de emociones.
* Extracción de características de la señal de voz, comparación entre señales de voz, eliminación de ruido.
* Generar un modelo de aprendizaje para el reconocimiento de emociones..
* Evaluar el funcionamiento y rendimiento del modelo.

# Marco teórico

## Emociones

Las emociones son un proceso del organismo humano que se activa mediante la

detección de alguna amenaza, peligro o desequilibrio, esto permite poner en acción los

recursos de cada persona para poner al frente alguna situación. Cada individuo puede

experimentar una emoción de forma diferente debido a las experiencias que viven o la

forma en la que tienen un diálogo con alguna otra persona. A veces las personas pueden

ser gobernadas por las emociones y estas hacen que se tomen acciones positivas o

negativas hacia uno mismo o hacia otra persona, es importante tener un claro control

sobre estas emociones.

## Reconocimiento de las emociones

El reconocimiento de voz es una tecnología que permite que una máquina comprenda el idioma hablado y lo traduzca a un formato legible por máquina. También se conoce como reconocimiento de voz por computadora o reconocimiento de voz automática. Es una forma de comunicarse con una computadora, y una computadora puede realizar una tarea específica según el comando.

El habla es el método más fundamental de comunicación humana. Los humanos pueden detectar naturalmente la emoción en el discurso que se les presenta. Sin embargo, no es tan sencillo para las máquinas. Aquí es donde se destaca la importancia del reconocimiento de emociones del habla (SER). Un sistema SER utiliza archivos que contienen datos de voz y los clasifica en varias emociones independientemente del contenido semántico.

Para crear un sistema de reconocimiento de emociones, se pueden emplear modelos de aprendizaje automático y aprendizaje profundo. Esto implica extraer características significativas de los datos sin procesar y utilizarlos para hacer que las máquinas entiendan las emociones humanas a través del entrenamiento de modelos. Durante el proceso de entrenamiento, el modelo aprende a clasificar la información y producir los resultados deseados mientras mantiene un cierto nivel de precisión.

La tarea de reconocimiento de emociones del habla se divide en dos secciones principales: selección y extracción de características, y clasificación.

### Selección y extracción de características

En este estudio se realizó el uso de tres características extraídas de los datos de audio, MFCC (coeficientes cepstrales de frecuencia de Mel), espectrograma de Mel y chroma.

* **MFCC (coeficientes cepstrales de frecuencia de Mel)**

En el análisis convencional de señales de tiempo, cualquier componente periódico (por ejemplo, ecos) aparece como picos agudos en el espectro de frecuencia correspondiente (es decir, espectro de Fourier. Esto se obtiene aplicando una transformada de Fourier en la señal de tiempo). Cualquier característica del cepstrum se obtiene aplicando la transformada de Fourier en un espectrograma. La característica especial de MFCC es que se toma en una escala Mel, que es una escala que relaciona la frecuencia percibida de un tono con la frecuencia real medida. Escala la frecuencia para que coincida más estrechamente con lo que el oído humano puede escuchar. La envolvente del espectro de potencia temporal de la señal del habla es representativa del tracto vocal y MFCC representa con precisión esta envolvente.

* **Espectrograma de Mel**

Se calcula una transformada rápida de Fourier sobre segmentos de ventana superpuestos de la señal, y obtenemos lo que se denomina espectrograma. Este es solo un espectrograma que representa la amplitud que se mapea en una escala de Mel.

* **Chroma.**

Un vector Chroma es típicamente un vector de características de 12 elementos que indica cuánta energía de cada clase de tono está presente en la señal en una escala cromática estándar.

## Aprendizaje No supervisado

El aprendizaje no supervisado se asocia con el aprendizaje sin supervisión o capacitación. Esto se realiza sin necesidad de etiquetar objetos. El objetivo del aprendizaje no supervisado es identificar las relaciones implícitas que puedan tener los datos. En el aprendizaje no supervisado, los algoritmos se entrenan con datos que no están etiquetado ni clasificado. En el aprendizaje no supervisado, el agente necesita aprender patrones sin valores de salida correspondientes.

Clasificación y regresión fueron los dos modelos usados en aprendizaje supervisado y los modelos de aprendizaje no supervisado son: **clustering, asociación y reducción de la dimensionalidad.**

### Clustering

La agrupación en clusters es una técnica de minería de datos que agrupa datos no etiquetados en grupos basados en similitudes y diferencias. Por ejemplo, el algoritmo de K-means divide puntos de datos similares en grupos, donde el valor “K” indica el tamaño y la granularidad del agrupamiento. Este método es útil para la segmentación del mercado, la compresión de imágenes y otras aplicaciones.

Otro tipo de método de aprendizaje no supervisado es la asociación, que emplea varias reglas para descubrir relaciones entre variables en un conjunto de datos. Estas técnicas se utilizan comúnmente en el análisis de la cesta de la compra y en los motores de recomendación, como las sugerencias “Los clientes que compraron esta cosa también compraron esta otra cosa”.

Cuando el número de características/dimensiones en un conjunto de datos es demasiado grande, se utiliza la reducción de la dimensionalidad como técnica de aprendizaje. Mantiene la integridad de los datos mientras reduce la cantidad de entradas de datos a un tamaño manejable. Esta técnica se usa con frecuencia en la etapa de preprocesamiento de datos, por ejemplo, cuando los codificadores automáticos eliminan el ruido de los datos visuales para mejorar la calidad de la imagen.

El uso de conjuntos de datos etiquetados es la principal diferencia entre los dos enfoques. El aprendizaje no supervisado no utiliza datos de entrada y salida etiquetados, mientras que el aprendizaje supervisado si lo hace. En lugar de hacer predicciones iterativas sobre conjuntos de datos y ajustar la respuesta correcta con un algoritmo de entrenamiento de aprendizaje, los modelos de aprendizaje no supervisados funcionan por sí solos para descubrir la estructura inherente de los datos no etiquetados.

l