

**UNIVERSIDAD “MAYOR DE SAN SIMÓN”**

**FACULTAD DE CIENCIAS Y TECNOLOGÍA**

**Departamento de Informática y Sistemas**

**Aprendizaje No Supervisado**

**Sistema de Seguridad Basado en el Reconocimiento de Voz**

**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**

**Mayo 2022**

**COCHABAMBA - BOLIVIA**

# Introducción

Hoy en día las personas tienen la necesidad de resguardar y de saber quien ha tenido acceso a sus viviendas, información, etc. Por ejemplo, en los condominios en los que los residentes salen y entran a toda hora y para poder acceder a sus viviendas necesitan verificar su identidad con el portero del condominio.

También en muchas empresas y organizaciones es necesario restringir el acceso de los empleados a cierta información o zonas de la empresa. Por ejemplo, documentación importante, áreas peligrosas, áreas donde se encuentra equipo sensible, etc.

Debido a todas estas situaciones se creó los sistemas de seguridad, que es la actividad principal de la seguridad física e informática, con el afán de facilitar el acceso y la seguridad de que solo las personas autorizadas tengan acceso a los ambientes e información.

Los sistemas de seguridad han ido evolucionando a medida que avanza la tecnología, los primeros sistemas de seguridad fueron los sistemas de identificación por teclado, luego aparecieron los Sistemas de identificación por proximidad y radiofrecuencia, donde se usa tarjetas para identificarse. El problema de estos sistemas es que los usuarios podrían compartir la clave o perder la tarjeta de acceso dejando los sistemas inservibles.

Debido a estos problemas y con el avance en la inteligencia artificial se crearon los sistemas de seguridad Biométricos, usando cualidades únicas de las personas como las huellas dactilares, reconocimiento facial, reconocimiento de iris y retina o reconocimiento de voz.

La ventaja de los sistemas de seguridad por voz es que, a partir de una muestra de voz, un software se encarga de realizar la autenticación, reduciendo el riesgo de intrusos o accesos no autorizados.

Es por eso que el presente trabajo se basará en sistemas de seguridad por voz, utilizando para su desarrollo Inteligencia artificial y técnicas de machine learning, más específicamente la técnica de Aprendizaje No Supervisado, que encuentra patrones ocultos o estructuras intrínsecas en los datos de entrada.

# Objetivos

1. **Objetivo general**

Desarrollar un sistema de seguridad basado en el reconocimiento de voz.

1. **Objetivos específicos**
2. Analizar la transmisión, aplicación y manipulación de las señales de voz para poder utilizarlas en el sistemas de seguridad.
3. Extracción de características de la señal de voz, comparación entre señales de voz, eliminación de ruido.
4. Evaluar el funcionamiento y rendimiento del sistema.
5. Analizar los resultados obtenidos sobre la implementación de Machine Learning.

# Marco teórico

1. **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.**

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.

**Algoritmos aprendizaje no supervisado**

* Análisis de Correspondencias Simple y Múltiple
* Escalamiento Multidimensional
* Análisis Clúster Jerárquico y No Jerárquico
* Análisis de componentes principales
* Análisis Factorial

1. **Reconocimiento por voz**

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 reconocimiento de voz tiene tres etapas, **pre procesamiento, recolección y extracción de características, y decodificación.**

**Preprocesamiento:** Descomponer la señal acústica en pequeñas ventanas/segmentos o marcos de aproximadamente 20 a 25 milisegundos es la primera etapa del reconocimiento de voz. Debido a que la señal de voz de cada fonema es relativamente constante durante aproximadamente 10 a 20 milisegundos, elegir estas ventanas al azar hace que sea poco probable que cada ventana capture un fonema. Los investigadores usan ventanas superpuestas para asegurarse de que algunas ventanas capturen fonemas. Además algunos filtros se usan para poder reducir el ruido y eliminar frecuencias bajas y altas.

**Recolección y extracción de características de voz:**: Luego viene la etapa de extracción de características, donde las medidas espectrales se definen en un conjunto de parámetros que representan las cualidades acústicas de las unidades fonéticas.

Estos factores pueden incluir:

* nasalidad: la presencia o ausencia de resonancia nasal
* fricación: la presencia o ausencia de excitación aleatoria en la voz
* ubicación de formantes: las frecuencias de las primeras tres resonancias
* clasificación de sonido tonal y no tonal

Segmentación y etiquetado del procesamiento es donde el algoritmo busca descubrir regiones estables con pocos cambios en las características, que se etiquetan en función de qué tan bien se ajusta la característica en la región a las unidades fonéticas individuales resultando en una caracterización de red de fonemas del habla.

**Decodificación:** El objetivo del proceso de decodificación es determinar la cadena de palabras más probable en función de las características acústicas que se han observado.

En lugar de estimar directamente las probabilidades de las cadenas de palabras dadas las características observadas, la mayoría de los sistemas utilizan el Teorema de Bayes.

La mejor respuesta es la cadena de palabras con la probabilidad de modelo acústico más alta multiplicada por la probabilidad de modelo de lenguaje.

