



**MÉTODO AUTOMÁTICO DE RECONOCIMIENTO DE VOZ PARA LA  
CLASIFICACIÓN DE VOCALES AL LENGUAJE DE SEÑAS COLOMBIANO.**

**Andrés Santiago Arias Páez  
David Andrés Rubiano Venegas**

Universidad Católica de Colombia  
Facultad de Ingeniería  
Programa de Ingeniería de Sistemas  
Bogotá, Colombia

2018

**MÉTODO AUTOMÁTICO DE RECONOCIMIENTO DE VOZ PARA LA  
CLASIFICACIÓN DE VOCALES AL LENGUAJE DE SEÑAS COLOMBIANO.**

**Andrés Santiago Arias Páez**  
**David Andrés Rubiano Venegas**

Este trabajo de grado es presentado como requisito para optar al título de:  
**Ingeniero de Sistemas**

Asesor: Roger Enrique Guzmán Avendaño  
Msc Ingeniería de Sistemas y Computación  
reguzmab@ucatolica.edu.co

Universidad Católica de Colombia  
Facultad de Ingeniería  
Programa de Ingeniería de Sistemas  
Bogotá, Colombia

2018



## Atribución-NoComercial 2.5 Colombia (CC BY-NC 2.5)

La presente obra está bajo una licencia:  
**Atribución-NoComercial 2.5 Colombia (CC BY-NC 2.5)**

Para leer el texto completo de la licencia, visita:  
<http://creativecommons.org/licenses/by-nc/2.5/co/>

### Usted es libre de:



Compartir - copiar, distribuir, ejecutar y comunicar públicamente la obra  
hacer obras derivadas

### Bajo las condiciones siguientes:



**Atribución** — Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante (pero no de una manera que sugiera que tiene su apoyo o que apoyan el uso que hace de su obra).



**No Comercial** — No puede utilizar esta obra para fines comerciales.

## **Nota de Aceptación**

Aprobado por el comité de grado en cumplimiento de los requisitos exigidos por la Facultad de Ingeniería y la Universidad Católica de Colombia para optar al título de ingeniero de sistemas.

---

Jorge Enrique Carrillo  
**Jurado 1**

---

German Ricardo Rodríguez  
**Jurado 2**

---

Roger Enrique Guzmán Avendaño, Msc.  
**Asesor**

BOGOTÁ D.C, NOVIEMBRE 19 DE 2018.

## **DEDICATORIA**

El presente trabajo investigativo lo dedicamos principalmente a nuestro tutor, por darnos el apoyo y fuerza para continuar este proceso de obtener uno de los anhelos más deseados.

A nuestros padres, por su apoyo, amor, trabajo y sacrificio en todos estos años, gracias a ustedes hemos llegado hasta aquí y convertirnos en lo que somos.

A nuestros amigos que nos han apoyado y han hecho que este trabajo se realice con éxito en especial a aquellos que nos instruyeron durante el camino y compartieron sus conocimientos.

A todas las personas que nos regalaron su voz para dar inicio con el proyecto.

## **AGRADECIMIENTOS**

Agradecemos a todo el apoyo a nuestro tutor Roger Guzmán por guiarnos a lo largo del desarrollo del proyecto.

Gracias a nuestros padres: Jeannette Venegas y José Rubiano; y, Álvaro Arias y Luisa Páez por ser los principales pilares para apoyar nuestros sueños, por confiar y creer en nuestros sueños, por confiar, por los consejos, valores y principios que nos han inculcado.

Agradecemos a nuestros docentes de la Universidad Católica de Colombia, por haber apoyado y compartido sus conocimientos a lo largo de la preparación de nuestra profesión.

A nuestros compañeros: Camila, Evert, Gabriel, Jean, que nos ayudaron con sus conocimientos y ánimos para continuar con las metas trazadas sin desfallecer su gran amistad.

## Resumen

Este trabajo de grado investiga sobre cuál es el mejor modelo para el reconocimiento de voz, para detectar vocales del idioma español haciendo uso del aprendizaje de máquina, también cómo se puede hacer uso del modelo para apoyar la comunicación entre una persona sorda y una oyente haciendo uso del lenguaje de señas colombiano (LSC).

El proyecto fue desarrollado de la siguiente manera: En primer lugar se construyó el conjunto de datos con 2000 audios en formato wav, que contiene la pronunciación de vocales, que se recolectaron a través de un trabajo de campo y se etiquetaron, luego se realizó la eliminación de los silencios que contenían los audios en la etapa llamada preprocesamiento, posteriormente se representan los audios de manera numérica haciendo uso de Coeficientes Cepstrales en las Frecuencias de Mel (MFCC) y la Codificación predictiva lineal (LPC), en la etapa llamada extracción de características, seguidamente se dividió el conjunto de datos en subconjuntos uno de entrenamiento y uno de pruebas en la etapa llamada muestreo, además se usó validación cruzada k-folios, se hizo uso de k vecinos más cercanos (KNN), Redes neuronales artificiales (RNA) y Maquinas de soporte vectorial (SVM) que son modelos de clasificación de aprendizaje supervisado, seguidamente se entrenaron los modelos con el conjunto de entrenamiento y finalmente se obtuvieron reportes poniendo a prueba los modelos entrenados previamente con el conjunto de pruebas.

Los resultados obtenidos del proyecto son el conjunto de datos compuesto de 2000 audios con la pronunciación de las vocales (a, e, i, o y u), el código utilizado y el mejor modelo de clasificación a partir de los reportes obteniendo un rendimiento mayor al 90% de clasificación de pronunciación de vocales.

**Palabras Claves:** Aprendizaje de máquina, Lenguaje de señas colombiano (LSC), modelos de clasificación, reconocimiento de voz.

## Abstract

This grade study investigates what is the best model for speech recognition, to detect vowels of the Spanish language using machine learning, also how you can make use of the classification model to support communication between a Deaf person and a listener using Colombian sign Language (LSC).

The project was developed as follows: First was built the dataset with 2000 audio in WAV format, which contains the pronunciation of vowels, which were collected through a field work and were tagged, then performed the elimination Of the silences that contained the audios in the stage called preprocessing, the audios are then represented numerically using Cepstrales coefficients in the Mel frequencies (MFCC) and the Linear Predictive coding (LPC), in the stage Called extraction of characteristics, then the dataset was divided into one training subsets and one of tests in the stage called sampling, in addition it was used K-folios cross validation, it made use of K nearest Neighbors (KNN), networks Artificial neurons (RNA) and vector support machines (SVM) which are models of supervised learning classification, then the models were trained with the training set and finally, reports were obtained by testing the models Pre-trained with the test set.

The results obtained from the project are the data set composed of 2000 audios with the pronunciation of the vowels (a, E, I, O and u), the code used and the best classification model from the reports obtaining a performance greater than 90%.

**Key words:** Colombian sign language (LSC), classification models, machine learning, voice recognition.



## CONTENIDO

RESUMEN.....	7
ABSTRACT.....	8
ANEXOS.....	13
ACRONIMOS.....	14
INTRODUCCIÓN.....	15
1. GENERALIADES.....	16
1.1. PLANTEAMIENTO DEL PROBLEMA.....	16
1.2. FORMULACIÓN DEL PROBLEMA.....	18
1.3. JUSTIFICACIÓN.....	19
1.4. OBJETIVOS.....	20
1.4.1. Objetivo General.....	20
1.4.2. Objetivos Específicos.....	20
1.5. MARCO REFERENCIAL.....	21
1.5.1 MARCO TEÓRICO.....	21
1.5.2 MARCO CONCEPTUAL.....	28
1.6. ALCANCES Y LIMITACIONES.....	55
1.6.1 ALCANCES.....	55
1.6.2 LIMITACIONES.....	55
1.7. ESTADO DEL ARTE.....	56
1.8. METODOLOGÍA.....	56
1.9. DISEÑO METODOLOGICO.....	68
1.10. INSTALACIONES Y EQUIPO REQUERIDO.....	76
1.11. DISCUSION Y RESULTADOS.....	77
1.12. ESTRATEGIAS DE COMUNICACIÓN Y DIVULGACIÓN.....	88
1.13. CONCLUSIONES.....	89
1.14. TRABAJOS FUTUROS.....	90
1.15. REFERENCIAS BIBLIOGRÁFICAS.....	91
1.16. ANEXOS.....	96

## LISTA DE TABLAS

Tabla 1. Funciones de activación.....	50
Tabla 2. Medidas de desempeño.....	53
Tabla 3. Tercera prueba en un ambiente con ruido. ....	61
Tabla 4. Reportes .....	67
Tabla 5. Métricas de desempeño.....	75
Tabla 6. Métricas del mejor modelo SVM para MFCC.....	77
Tabla 7. Métricas del mejor modelo SVM para LPC .....	78
Tabla 8. Métricas del mejor modelo SVM para MFCC+LPC.....	79
Tabla 9. Comparación de resultados KNN.....	80
Tabla 10. Métricas del mejor modelo KNN para MFCC .....	81
Tabla 11. Métricas del mejor modelo KNN para LPC .....	82
Tabla 12. Métricas del mejor modelo KNN para MFCC+LPC .....	83
Tabla 13. Comparación de resultados NN .....	84
Tabla 14. Métricas del mejor modelo NN para MFCC .....	84
Tabla 15. Métricas del mejor modelo NN para LPC .....	85
Tabla 16. Métricas del mejor modelo NN para LPC.....	86

## LISTA DE FIGURAS

Figura 1 Descuidos y subjetividad. ....	19
Figura 2. Un Hertz.....	22
Figura 3. Sistema vocal .....	23
Figura 4. Clasificación de correos que son spam o no spam.....	25
Figura 5. Regresión lineal del precio de una casa. ....	25
Figura 6. Agrupación de datos con características iguales.....	26
Figura 7. Grupo de datos etiquetados y no etiquetados. ....	26
Figura 8. Aprende por refuerzo.....	27
Figura 9. Conexiones entre neuronas.....	27
Figura 10. Framing.....	29
Figura 11. Ventaneo .....	30
Figura 12. Representación de una onda en función del tiempo .....	31
Figura 13. Matriz de correlación LPC.....	35
Figura 14. Muestreo estratificado.....	37
Figura 15. Selección de datos de prueba y datos de entrenamiento. ....	38
Figura 16. Representación gráfica validación cruzada k-folios .....	39
Figura 17. Representación gráfica validación cruzada aleatoria.....	39
Figura 18. Representación gráfica validación cruzada de dejar un paso.....	40
Figura 19. Representación gráfica del sesgo y varianza .....	41
Figura 20. Relación varianza sesgo.....	41
Figura 21. Clasificación lineal con SVM.....	42
Figura 22. Clasificación no lineal con SVM.....	43
Figura 23. Regresión usando SVM.....	44
Figura 24. Regresión usando Kernel .....	44
Figura 25. Regresión usando Kernel. ....	45
Figura 26. Representación gráfica distancia de manhattan .....	46
Figura 27. Representación gráfica distancia euclidiana .....	46
Figura 28. Proceso de clasificación k vecinos más cercanos .....	47
Figura 29. Estructura general de una neurona.....	48
Figura 30. Red neuronal artificial.....	49
Figura 31. Capas de una red neuronal artificial. ....	51
Figura 32. Estructuras neuronales.....	51
Figura 33. Imagen del Traductor a lenguaje de señas Hetah. ....	56
Figura 34. Interfaz gráfica Centro De Relevo Colombia.....	58
Figura 35. Interfaz gráfica de la aplicación Voz y señas .....	59
Figura 36. Interfaz gráfica Hablando con Julis .....	60
Figura 37. Vibrador .....	60
Figura 38. Comparación de algunos métodos de clasificación .....	62
Figura 39. Diagrama de flujo metodológico 1. ....	64
Figura 40. Diagrama de flujo metodológico 2. ....	65

Figura 41. Diagrama de flujo metodológico 3 .....	65
Figura 42. Diagrama de flujo metodológico 4. ....	66
Figura 43. Diagrama de flujo metodológico 5. ....	66
Figura 44. Diagrama de flujo metodológico 6. ....	67
Figura 45. Construcción del conjunto de datos 1 .....	68
Figura 46. Construcción del conjunto de datos 2.....	69
Figura 47. Construcción del conjunto de datos 5.....	69
Figura 48. Diagrama de eliminación de silencios.....	70
Figura 49. Representación gráfica eliminación de silencios .....	70
Figura 50. Diagrama de flujo MFCC .....	71
Figura 51. Diagrama de flujo LPC.....	72
Figura 52. Representación de las características .....	72
Figura 53. Representacion grafica de la exploracion de parametros .....	74
Figura 54. Matriz de confusión.....	75
Figura 55. Representación en lenguaje de señas colombiano .....	75
Figura 56. Mejor modelo SVM con MFCC .....	78
Figura 57. Mejor modelo SVM con LPC.....	79
Figura 58. Mejor modelo SVM con MFCC+LPC .....	80
Figura 59. Mejor modelo KNN con MFCC .....	81
Figura 60. Mejor modelo KNN con LPC.....	82
Figura 61. Mejor modelo KNN con MFCC+LPC .....	83
Figura 62. Mejor modelo NN con MFCC.....	85
Figura 63. Mejor modelo NN con LPC .....	86
Figura 64. Mejor modelo NN con MFCC+LPC.....	87

## **ANEXOS**

Anexo A: Conjunto de datos. ....	96
Anexo B: Repositorio. ....	96
Anexo C: Artículo científico. ....	96

## ACRONIMOS

**API:** Interfaz de programación de aplicaciones.

**DANE:** Departamento Administrativo Nacional de Estadística.

**DCT:** Transformada de coseno discreta del inglés (Discrete Cosine Transform).

**FFT:** Transformada rápida Fourier del inglés (Fast Fourier Transform).

**INSOR:** Instituto Nacional para Sordos.

**KNN:** K vecinos más cercanos del inglés (k-nearest neighbors algorithm).

**LPC:** Codificación Predictiva Lineal del inglés (Linear prediction coding).

**LSC:** Lenguaje de Señas colombiano.

**LSE:** Lenguaje de señas española.

**LSF:** Lenguaje de señas francesa.

**LSM:** Lenguaje de señas mexicano.

**MFCC:** Coeficientes Cepstrales en las Frecuencias de Mel del inglés (Mel Frequency Cepstral Coefficients).

**RNA:** Redes neuronales artificiales.

**SVM:** Maquinas de soporte vectorial del inglés (Support Vector Machine).

## INTRODUCCIÓN

Desde tiempos inmemorables, el hombre ha dado significado a ciertos elementos que percibe a través de los sentidos, estableciendo entre ellos una asociación, ya sea usando expresiones, señas y gestos para mejorar su comunicación con otros, estos pueden ser de gran ayuda al momento de comunicarse y/o expresar un conjunto de ideas, ya sea indicar un lugar, objeto, persona o describir algo, por ejemplo cuando una persona viaja al extranjero teniendo poco conocimiento del idioma opta por utilizar señas para expresar lo que quiere sin necesidad de hacer uso de palabras, ya sea moviendo una parte de su cuerpo o indicando lo que desea.

Las personas sordas, comenzaron a usar las señas como su principal forma de comunicación; como resultado de esto surgió el lenguaje de señas, el cual está compuesto de estructuras gramaticales definidas para estandarizar lo que significa, facilitando su aprendizaje y comprensión, haciendo uso de la vista, las manos, el cuerpo y los gestos faciales. Sin embargo, pocas personas saben de este idioma, porque generalmente se aprende por la necesidad de comunicarse con amigos o familiares que tengan esta discapacidad, además se requiere bastante esfuerzo y tiempo para poder comunicarse adecuadamente.

Por lo anterior, en este trabajo se propone analizar la eficiencia de diferentes modelos de aprendizaje de máquina para el reconocimiento de vocales por la voz, para ello se realizará un trabajo de campo para grabar la pronunciación de diferentes personas con el objetivo de construir el conjunto de datos, seguidamente se realizara una etapa de pre procesamiento a las voces, donde se eliminara el ruido y sonidos no necesarios, una vez ya pre procesados se extraerán las características, haciendo uso de los Coeficientes Cepstrales en las Frecuencias de Mel y Codificación Predictiva Lineal, seguidamente se separan los datos mediante el muestreo, este paso consiste en determinar los datos que se usarán para entrenamiento y para pruebas; posteriormente se tomaran los datos de entrenamiento para clasificarlos mediante los algoritmos de máquina de soporte vectorial, K vecinos cercanos y redes neuronales además se usa validación cruzada de 5 folios para evitar el sobre entrenamiento y por ultimo para cada método realizar el análisis de precisión de su clasificación por medio de medidas de desempeño como precision, recall y F1 score.

## 1. GENERALIADES

### 1.1. PLANTEAMIENTO DEL PROBLEMA

Según los datos censales del 2005 (DANE e INSOR) en Colombia existen más de 2 millones de personas con algún tipo de limitación, de las cuales aproximadamente 450 mil son sordos que equivalen al 1,1% de la población<sup>1</sup>. Estas personas sordas hacen uso del lenguaje de señas para comunicarse y solo un grupo reducido de lo domina, es por esta razón que dichas personas no pueden relacionarse. Contemplando el hecho de que el 1.1% de la población colombiana es sorda, se puede inferir que la mayoría de las personas en nuestra sociedad son oyentes, este hecho provoca que se ignore el lenguaje de señas, a menos que una persona cercana padezca de esta discapacidad, esto provoca que las personas sordas estén limitadas por barreras al momento de comunicarse y obtener información, esto afecta las posibilidades que tienen para que se puedan incluir en la sociedad, puesto que las instituciones y personas no están adecuadas para ellos, a causa de esto se verán perjudicados en el ámbito educativo y laboral, esta pérdida de oportunidades provoca que la mayoría de las personas sordas sean de estratos 1 y 2<sup>2</sup>, afectando la calidad de vida de estas personas y al estar en esa situación crítica pueden llegar al punto de caer en la delincuencia para poder sobrevivir. Adicional a esto al estar enfermos pueden estar en condiciones de alto riesgo al no poder comunicar los síntomas que tengan y en casos extremos los pueden llevar a la muerte, todos estos casos afectan diferentes derechos de las personas.

Con la comprensión de las necesidades de los sordos y del lenguaje, surge la necesidad de romper la barrera de la comunicación con las demás personas haciendo uso de la tecnología, a causa de este problema surgieron algunos proyectos para mitigar el problema de la comunicación, como es el caso de “Voz &

---

<sup>1</sup> INSOR. Contexto general de la población sorda en Colombia. [en línea]. Bogotá: INSOR. [Citado el 28 marzo, 2018]. Disponible en internet:

<[http://www.insor.gov.co/observatorio/download/Infog\\_pan\\_sordos\\_Col\\_sept2016.pdf](http://www.insor.gov.co/observatorio/download/Infog_pan_sordos_Col_sept2016.pdf)>

<sup>2</sup> CASTRO, M.L., RUIZ, A.J., CÉSAR JIMÉNEZ, J., PATRICIA, N., ESPINOSA. Estadísticas e información para contribuir en el mejoramiento de la calidad de vida de la población sorda colombiana. [en línea]. s.l: CASTRO, M.L., RUIZ, A.J., CÉSAR JIMÉNEZ, J., PATRICIA, N., ESPINOSA [Citado el 7 junio 2018]. Disponible en internet: <[http://www.insor.gov.co/historico/images/boletín\\_observatorio.pdf](http://www.insor.gov.co/historico/images/boletín_observatorio.pdf)>



señas”<sup>3</sup>, esta es una aplicación móvil que traduce el lenguaje de señas mexicano (LSM), esta aplicación solamente tiene una imagen de cada letra representativa en el lenguaje de señas mexicano, el siguiente proyecto es el “Centro de Relevo Colombia”<sup>4</sup>, es una aplicación móvil que permite comunicarse con un intérprete, esta aplicación tiene una disponibilidad limitada por las personas que realizan esta labor, todos estos proyectos han apoyado la comunicación de personas sordas.

A pesar de todas estas incursiones aún no se ha generado una solución definitiva para el problema de la comunicación entre personas sordas y oyentes en Colombia, por lo cual considero pertinente investigar y realizar pruebas para la solución, mediante conocimientos de programación, enfocando la traducción de las vocales del español al LSC, haciendo uso del reconocimiento de voz logrando correctamente la traducción y mostrando la seña indicada para la vocal.

---

<sup>3</sup> Voz y Señas - Traductor LSM. [en línea]. s.l: Voz y Señas - Traductor LSM. [Citado el 28 marzo, 2018]. Disponible en internet: <<http://www.vozysenas.com/>>

<sup>4</sup> Centro de relevo. [en línea]. s.l: Centro de relevo. [Citado el 28 marzo, 2018]. Disponible en internet: <<http://centroderelvo.gov.co/632/w3-channel.html>>

## **1.2. FORMULACIÓN DEL PROBLEMA**

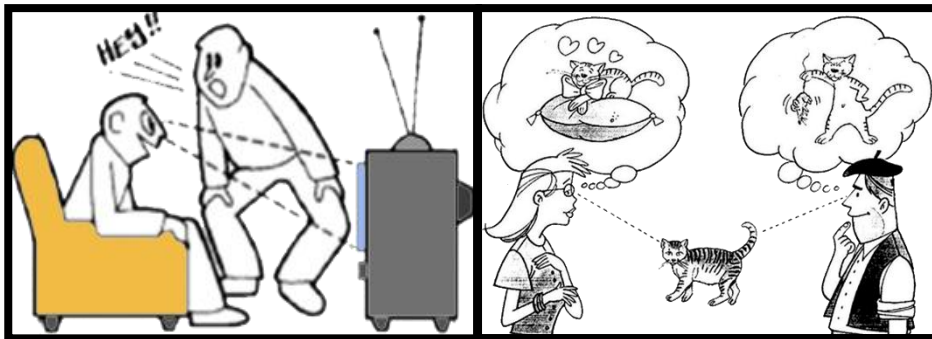
La pregunta de investigación es la siguiente:

¿Cómo se podría apoyar la comunicación mediante sonidos entre una persona sorda y una oyente haciendo uso de la tecnología?

### 1.3. JUSTIFICACIÓN

Actualmente en Colombia, las personas sordas tienden a ser excluidas<sup>5</sup>, ya que la mayoría de la población desconoce el lenguaje de señas a causa de que sólo hay un pequeño porcentaje de personas que hacen uso del lenguaje y no ven la necesidad de aprenderlo<sup>6</sup>, este desconocimiento genera un problema de comunicación entre una persona sorda y una oyente, aun si una persona oyente aprende el lenguaje de señas colombiano existen dificultades que son inherentes al ser humano, como es el ejemplo de la figura 1 donde las personas pueden ser susceptibles a factores que pueden afectar la efectividad de la comunicación como los descuidos, la velocidad de expresión, los estados de ánimo, el estrés o simplemente mayor lentitud para comprender, provocando que el mensaje que se quiere transmitir se altere o se pierda.

Figura 1 Descuidos y subjetividad.



Fuente. Pvivov. forma\_ver\_gato. [en línea]. s.l: pvivov. [Citado el 30 mayo, 2018]. Disponible en internet: [http://www.pvivov.net/recursos/psicopedagogia/images/forma\\_ver\\_gato.jpg](http://www.pvivov.net/recursos/psicopedagogia/images/forma_ver_gato.jpg).

Por lo que surge la necesidad de encontrar una solución que no solo supere los factores nombrados, sino que supere la barrera del lenguaje, es aquí donde por medio de la inteligencia humana y el avance tecnológico se crean herramientas y métodos para obtener resultados correctos, objetivos y de una manera más rápida.

<sup>5</sup> Noticias RCN. Colombia y el reto que tiene pendiente con los sordos del país. [en línea]. Bogotá: Noticias RCN. [Citado el 28 marzo, 2018]. Disponible en internet: <<https://www.noticiasrcn.com/nacional-pais/colombia-y-el-reto-tiene-pendiente-los-sordos-del-pais>>

<sup>6</sup> INSOR, Op. Cit. p. 12.

## **1.4. OBJETIVOS**

### **1.4.1. Objetivo General**

Implementar un método de clasificación automático de vocales por voz para la representación en el lenguaje de señas colombiano.

### **1.4.2. Objetivos Específicos**

1. Construir un conjunto de datos de pronunciación de vocales con sus respectivas etiquetas.
2. Desarrollar una estrategia de clasificación de vocales por voz, del lenguaje de señas colombiano haciendo uso de aprendizaje de máquina.
3. Comparar los métodos de clasificación haciendo uso de medidas de desempeño como precisión, recall y F1 Score.

## 1.5. MARCO REFERENCIAL

En este capítulo se muestra el marco referencial para el presente proyecto, este se divide en marco teórico y marco conceptual.

### 1.5.1 MARCO CONCEPTUAL

- **El lenguaje de señas.**

Es un lenguaje gestual usado por las personas sordas, tiene como base el uso de movimientos y expresiones haciendo uso de las manos, ojos, rostro y cuerpo<sup>7</sup>.

El Lenguaje de Señas Colombiano (LSC)<sup>8</sup> inicia aproximadamente en el año 1924 cuando se funda la primera institución educativa de sordos en Santafé de Bogotá por Nuestra señora de la sabiduría (internado católico bogotano), el cual tomó elementos del Lenguaje de Señas Francesa (LSF) por la influencia de las monjas que enseñaban en el lugar. Pasados 33 años se funda en Bogotá la primera asociación de sordos del país y al año en Cali la segunda asociación, para este tiempo lo inmigrantes y los colombianos que estudiaron en España introdujeron características del lenguaje de señas española (LSE). En los 70 la presencia de misioneros protestantes de los Estados Unidos provocó una influencia del Lenguaje de Señas Americano (LSA), estas influencias crearon las bases para el LSC. En 1996 se aprueba la ley 324<sup>9</sup> la cual reconoce la existencia oficial del LSC y exige al estado a financiar la formación de intérpretes. A partir de 1998 se realiza una investigación exhaustiva por la Universidad del Valle, en Cali, sobre cómo se comporta, se estructura y se define el lenguaje de señas, dando inicio a la producción de un diccionario de la lengua de señas colombiana, el cual busca ser el conjunto oficial de las señas y su significado para lograr una enseñanza más eficiente y unificada, este se finaliza en el año 2005.

---

<sup>7</sup>INSOR. ¿Qué es la lengua de señas? Portal niños. [en línea]. s.l: INSOR. [Citado el 1 mayo, 2018]. Disponible en internet: <<http://insor.gov.co/ninos/que-es-la-lengua-de-senas/>>

<sup>8</sup> ALEJANDRO OVIEDO. Colombia atlas sordo – Cultura Sorda. [en línea]. s.l: ALEJANDRO OVIEDO. [Citado el 1 mayo, 2018]. Disponible en internet: <<http://www.cultura-sorda.org/colombia-atlas-sordo/>>

<sup>9</sup> CONGRESO DE COLOMBIA. E 1996 Ley 324 de 1996 - Normas a favor de la Población Sorda. Ley. [en línea]. Bogotá: CONGRESO DE COLOMBIA. [Citado el 1 mayo, 2018] Disponible en internet: <[https://puntodis.com/wp-content/uploads/2015/12/Ley\\_324\\_de\\_1996.pdf](https://puntodis.com/wp-content/uploads/2015/12/Ley_324_de_1996.pdf)> p. 1,2,3.

- **Reconocimiento de voz**

Es una disciplina de la inteligencia artificial, el reconocimiento de voz dota a las máquinas con la capacidad de recibir mensajes orales por medio de un micrófono, una vez captado el mensaje se decodifica de forma numérica para realizar un análisis de su significado y así generar una respuesta adecuada<sup>10</sup>.

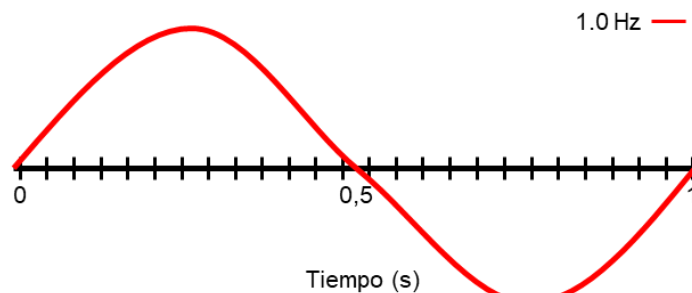
- **Sonido.**

Es un tipo de energía mecánica producida a causa de un movimiento<sup>11</sup> que se transmite por un medio, como el aire, líquidos y sólidos.

- **Hertz (Hz).**

Es unidad de frecuencia del Sistema Internacional de unidades, un Hertz es un ciclo de una onda en un segundo<sup>12</sup>, como se puede ver en la figura 2.

Figura 2. Un Hertz.



Fuente. Los autores.

- **Voz.**

Se produce por la vibración de las cuerdas vocales por medio del aire expulsado por los pulmones mediante la tráquea y laringe, estas vibraciones son moldeadas al pasar por la articulación de la lengua, los labios, el paladar y los dientes<sup>13</sup>.

---

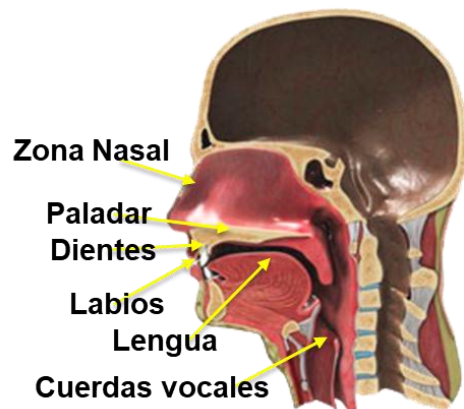
<sup>10</sup> EDUARDO LLEIDA SOLANO, 2000. Definición de frecuencia - Qué es, Significado y Concepto. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <<http://physionet.cps.unizar.es/~eduardo/investigacion/voz/rahframe.html>>.

<sup>11</sup> POLAVIDE.ES, el sonido. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <<http://www.polavide.es/energyluzsonido/sonido.html>>.

<sup>12</sup> SISTEMAS.COM, 2016. Definición de Hertz - Significado y definición de Hertz. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <https://sistemas.com/hertz.php>.

<sup>13</sup> DEFINICION.DE, Definición de voz - Qué es, Significado y Concepto. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <https://definicion.de/voz/>.

Figura 3. Sistema vocal



Fuente. Los autores.

- **Inteligencia Artificial.**

Es la rama de la computación que tiene como objetivo emular la inteligencia humana con base a los conocimientos lógicos y de las ciencias cognitivas. Por esta razón, existen diferentes formas de inteligencia<sup>14</sup>: Actuar como las personas, si una máquina es “inteligente” tendrá la capacidad de comunicarse con una persona de manera natural sin que se reconozca que es una máquina, también conocido como enfoque del test de Turing<sup>15</sup>.

Razonar como las personas, este tipo de inteligencia tiene como objetivo crear máquinas que razonen teniendo como base la forma de razonar de las personas. Razonar racionalmente, es un enfoque que mediante la lógica se emula una formalización del razonamiento. Actuar racionalmente, tiene como enfoque obtener resultados de manera objetiva, un claro ejemplo es un programa diseñado para ganar un juego como el ajedrez.

Existe una clasificación de la inteligencia artificial que considera el objetivo final de investigaciones, estas son:

Inteligencia artificial débil: Expresa que las máquinas solo pueden aparentar que razonan sin llegar al punto de tener conciencia, solo obedecen órdenes y están orientadas a una sola tarea, en la actualidad todas las inteligencias artificiales son de este tipo.

Inteligencia artificial fuerte: Se considera que la máquina puede tener conciencia, estados mentales y las capacidades de la mente humana,

---

<sup>14</sup> FUNDACIÓN GENERAL CSIC. Lychnos cuadernos de la Fundación General CSIC. [en línea]. s.l: Fundación General CSIC. [Citado el 1 mayo, 2018]. Disponible en internet: <[http://www.fgcsic.es/lychnos/es\\_es/articulos/inteligencia\\_artificial](http://www.fgcsic.es/lychnos/es_es/articulos/inteligencia_artificial)>

<sup>15</sup>TECHcetera. ¿Qué es el Test de Turing (y, Google: qué has hecho)? [en línea]. s.l: TECHcetera [Citado el 2 mayo, 2018]. Disponible en internet: <<http://techcetera.co/que-es-el-test-de-turing/>>

pudiendo igualar o superar la inteligencia humana, se puede adaptar a diferentes ambientes, tareas y situaciones. Todavía no existe este tipo de inteligencia artificial, sin embargo, se puede ver ejemplos de cómo podría ser una en la ciencia ficción.

- **Aprendizaje automático.**

Es la rama de la inteligencia artificial<sup>16</sup>, que consiste en proporcionar a las computadoras la habilidad de tomar decisiones a partir de los datos, por medio de algoritmos para la identificación de patrones complejos de un conjunto de datos, estos datos pueden estar etiquetados, en base a esto se pueden realizar acciones, como la predicción comportamientos futuros, agrupación según patrones ocultos o clasificación con respecto a lo que representa los datos a medida que se entrenan, estos algoritmos mejoran la precisión de sus resultados en el tiempo.

Existen los siguientes tipos de aprendizaje automático:

**Aprendizaje supervisado.**

Es un tipo de aprendizaje que se basa en descubrir la relación existente entre unas variables de entrada y unas variables de salida, teniendo como base un conjunto de datos de entrenamiento, donde se entrena al algoritmo con una gran cantidad de ejemplos de cómo deben clasificarse los datos, si se dan las condiciones será capaz de dar un resultado correcto e incluso cuando se le muestre valores que no haya visto, el aprendizaje supervisado es usado para los siguientes problemas:

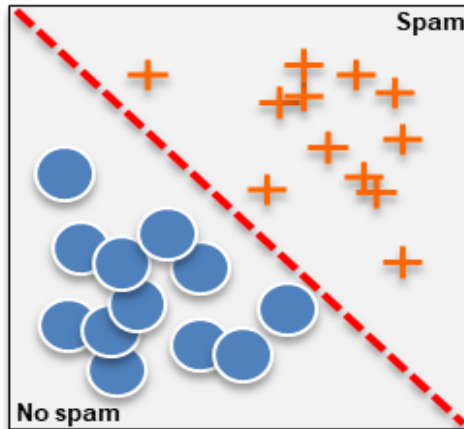
**Problemas de clasificación:** En este tipo de problemas la variable objetivo contiene las categorías, a las cuales se quiere asignar, dependiendo de las características de las variables de entrada como por ejemplo cuando se clasifican los correos como spam o no spam como se ve en la siguiente imagen.

---

<sup>16</sup>DotCSV. ¿Qué es el Aprendizaje Supervisado y No Supervisado? [en línea]. s.l: DotCSV [Citado el 2 mayo, 2018]. Disponible en internet: <<https://www.youtube.com/watch?v=oT3arRRB2Cw>>



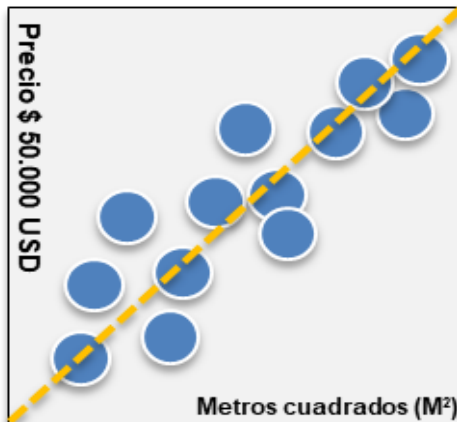
Figura 4. Clasificación de correos que son spam o no spam.



Fuente. Los autores.

**Problemas de regresión:** En este tipo de problemas la variable objetivo es continua y se espera encontrar la línea de tendencia en base a los datos de entrada como por ejemplo a la hora de saber a qué precio vender una casa en base a su tamaño.

Figura 5. Regresión lineal del precio de una casa.



Fuente. Los autores.

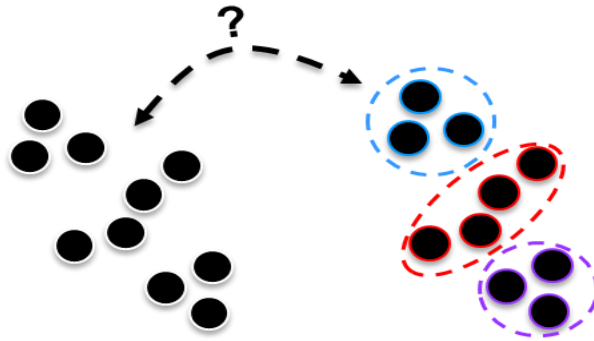
- **Aprendizaje no supervisado.**

Es un modelo de aprendizaje que consigue producir conocimientos únicamente de los datos que se proporcionan como entradas sin necesidad de explicarle al sistema qué resultado queremos obtener<sup>17</sup>, lo que hace el aprendizaje no supervisado es buscar patrones de similitud de los datos de entrada y agruparlos, por ejemplo, los símbolos de un lenguaje.

---

<sup>17</sup> DotCSV. op. Cit. p. 22.

Figura 6. Agrupación de datos con características iguales

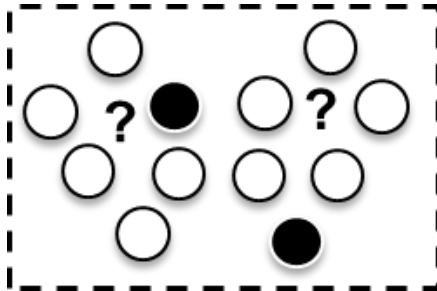


Fuente. Los autores.

- **Aprendizaje semi supervisado.**

Este aprendizaje es una mezcla de los aprendizajes supervisados y no supervisados este hace uso de datos mostrados (datos etiquetados) y nuevos datos no conocidos (datos no etiquetados) con el objetivo de encontrar patrones de similitud<sup>18</sup>, este tipo de aprendizaje puede iniciar con poco conocimiento, así eliminando el costo de tener que etiquetar los datos. En la siguiente imagen se hace una representación visual de los datos usados en este tipo de aprendizaje.

Figura 7. Grupo de datos etiquetados y no etiquetados.



Fuente. Los autores.

- **Aprendizaje por refuerzo.**

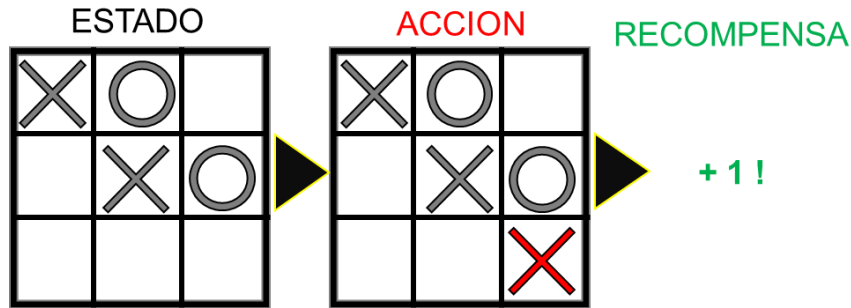
Este aprendizaje está basado en modelos conductistas, consiste en aprender de su entorno donde el entorno es un conjunto de datos que reciben en una etapa llamada entrenamiento, a medida que se entrena este va tomando decisiones y suposiciones para lograr un objetivo, además tendrá a consideración las recompensas positivas y negativas en

---

<sup>18</sup> Maria Mercedes Gomez. ¿Sabes qué es una Machine Learning? - Comunidad e-Learning. [en línea], 2017. [Citado el 3 mayo, 2018]. Disponible en internet: <<http://elearningmasters.galileo.edu/2017/09/21/sabes-que-es-una-machine-learning/>>

el proceso para realizar un comportamiento adecuado<sup>19</sup>. Un ejemplo de este tipo de aprendizaje es AlphaGo Zero, el cual fue entrenado para el juego chino “go” y que ha logrado vencer a profesionales de este juego.

Figura 8. Aprende por refuerzo.

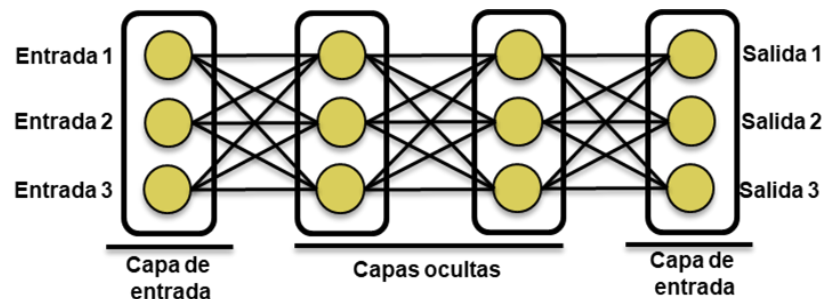


Fuente. Los autores.

- **Aprendizaje profundo.**

Es un conjunto de técnicas y procedimientos algoritmos basados en aprendizaje de máquina para lograr que aprendan similarmente como lo hace un ser humano, estos algoritmos simulan el funcionamiento básico del cerebro mediante neuronas artificiales, están constituidas por capas anidadas de nodos que se interconectan y después de cada nueva experiencia, aprenden reacomodando las conexiones entre los nodos<sup>20</sup>

Figura 9. Conexiones entre neuronas.



Fuente. Los autores.

<sup>19</sup>Fernando Sancho Caparrini. Aprendizaje por refuerzo: algoritmo Q Learning. [en línea]. s.l: Fernando Sancho Caparrini. [Citado el 4 mayo, 2018]. Disponible en internet: <<http://www.cs.us.es/~fsancho/?e=109>>

<sup>20</sup> NAVARRO, B.G., [2015]. Trabajo de Fin de Grado Implementación de Técnicas de Deep Learning Implementation of Deep Learning Techniques. [en línea]. S.l.: [Consulta: 30 octubre 2018]. Disponible en: [https://riull.ull.es/xmlui/bitstream/handle/915/1409/Implementacion de Tecnicas de Deep Learning.pdf?sequence=1](https://riull.ull.es/xmlui/bitstream/handle/915/1409/Implementacion%20de%20Tecnica%20de%20Deep%20Learning.pdf?sequence=1).

El aprendizaje profundo ha exhibido que tiene un gran potencial ya que ha sido capaz de crear varios avances como: Sistemas de vigilancia a gran escala, coches autónomos, generar arte, rostros y muchas cosas más.

## 1.5.2 MARCO TEORICO

A continuación, se explican los principales conceptos acordes al proyecto.

- **Pre-Procesamiento.**

Es una fase en el proceso de aprendizaje automático que consiste en obtener un conjunto de datos útiles para la fase de extracción de características<sup>21</sup>, en el caso del reconocimiento de voz consiste en limpiar la señal de audio de interferencias y otros sonidos no deseados, obtenido una señal, existen los siguientes métodos:

- **Framing**

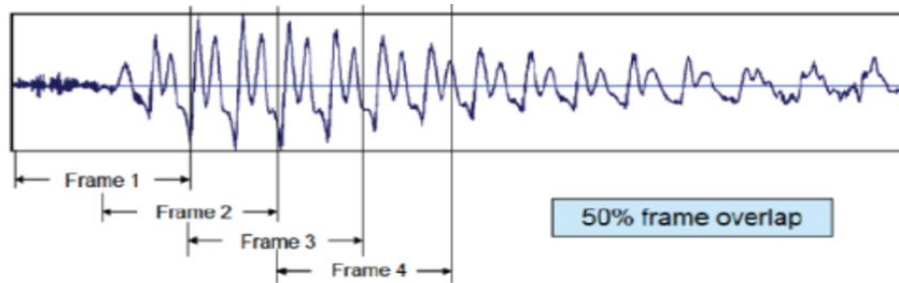
Las señales de voz varían a través del tiempo lo que genera que sea complejo trabajar con toda la señal de voz inmediatamente, por esta razón en el procesamiento de voz se hace uso un proceso llamado Framing que consiste en segmentar la señal de voz en intervalos iguales entre 20 y 25 ms<sup>22</sup>, ya que en este intervalo de tiempo la señal mantiene una amplitud relativamente constante y hay suficiente información en las muestras del audio, para realizar un análisis adecuado de toda la señal, además estos segmentos tienden a estar solapados entre ellos para evitar una pérdida de información que pueden generarse en otros procesos que se realizan en la señal, un ejemplo grafico de este proceso de muestra a continuación en la figura 13, donde la señal de audio está siendo separada en segmentos y están un 50% solapados los segmentos.

---

<sup>21</sup> GARCÍA, S., RA-MÍREZ-GALLEGO, S., LUENGO, J., HERRERA, F. y RAMÍREZ-GALLEGO, S. Big Data monografía Big Data: Preprocesamiento y calidad de datos. [en línea]. s.l: GARCÍA, S., RA-MÍREZ-GALLEGO, S., LUENGO, J., HERRERA, F. y RAMÍREZ-GALLEGO, S. [Citado el 4 mayo 2018]. Disponible en internet: <[http://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133\\_Nv237-Digital-sramirez.pdf](http://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133_Nv237-Digital-sramirez.pdf)> p. 2.

<sup>22</sup> SIGNAL PROCESSING, Q., 2018. Why is each window/frame overlapping? 2016-12-28 [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://dsp.stackexchange.com/questions/36509/why-is-each-window-frame-overlapping>.

Figura 10. Framing



Fuente. Design of Digital Blowing Detector [Citado el 28 mayo, 2018]. Disponible en internet: <[https://www.researchgate.net/figure/Signal-Frame-by-Frame-Processing\\_fig2\\_287922921](https://www.researchgate.net/figure/Signal-Frame-by-Frame-Processing_fig2_287922921)>.

- **Ventaneo**

Al realizar el proceso de segmentación a la señal original provoca que en los segmentos estén trucados y presenten características diferentes a la señal original, esto a causa de transiciones bruscas en el inicio y final de los segmentos de la señal, estas transiciones bruscas se conocen como discontinuidades, el efecto que tienen estas en el análisis espectral del segmento es que no se obtendrá el espectro real de la señal sino una versión distorsionada, lo que se conoce como fuga espectral.

El objetivo del ventaneo es disminuir la fuga espectral, disminuyendo la amplitud de las discontinuidades en los segmentos de la señal, el proceso que se lleva a cabo para utilizar el ventaneo es multiplicar los segmentos de la señal de voz, por una ventana de suavizado de longitud finita, cuya amplitud varía suavemente hacia cero en los bordes<sup>23</sup>. Existen diferentes tipos de ventanas de suavizado, un ejemplo es la ventana hamming y se representa con la siguiente fórmula:

$$W(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N} \quad \text{para } n = 0, 1, \dots, N - 1$$

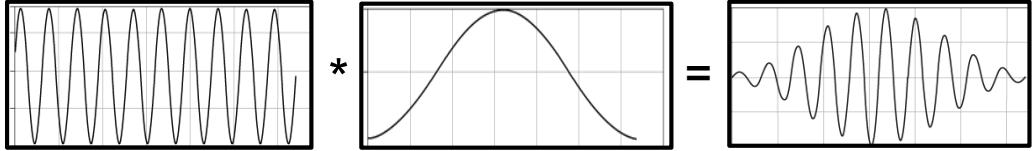
Donde:

- **N** es el número total de muestras en el segmento de señal
- **n** es la muestra actual
- **w(n)** es el valor de la ventana en la muestra n de la señal.

<sup>23</sup> DOCUMENTACION MATLAB, 2018. Hamming window - MATLAB hamming. 2006 [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://www.mathworks.com/help/signal/ref/hamming.html>.

A continuación, se muestra gráficamente el uso de ventana de suavizado, en este caso de tipo haming, a una señal:

Figura 11. Ventaneo



Fuente. [www.physik.uzh.ch](http://www.physik.uzh.ch) [Citado el 28 mayo, 2018]. Disponible en internet: < [https://www.physik.uzh.ch/local/teaching/SPI301/LV-2015-Help/lvanlsconcepts.chm/Windowing\\_Signals.html](https://www.physik.uzh.ch/local/teaching/SPI301/LV-2015-Help/lvanlsconcepts.chm/Windowing_Signals.html) >.

- **Calculo de Energía a Corto Plazo**

Es una técnica usada para calcular la energía/amplitud de la señal a partir del valor de las muestras que esta contiene, al ser usada en los segmentos de la señal se pueden tomar decisiones según el valor obtenido, como, por ejemplo, descartar segmentos que no superen un umbral, esto con el objetivo de eliminar los silencios y enfatizar a los segmentos que contengan información relevante, es definida con la siguiente formula<sup>24</sup>:

$$S(m) = \sum_{n=0}^{N-1} [x(n)]^2$$

Donde:

- **S(m)** es el valor de energía del segmento m de la señal.
  - **N** es el total de muestras en el segmento.
  - **x(n)** es una muestra especifica en la señal.
- **Transformada de Fourier.**  
Es una operación matemática que convierte cualquier función matemática a otro dominio, llamado el dominio de la frecuencia<sup>25</sup>, esta operación permite descomponer una señal periódica en una suma de ondas seno con diferentes frecuencias, fases y amplitudes, se puede definir como:

$$s(t) = a_0 + a_1 \sin(\omega t + \phi_1) + a_2 \sin(2\omega t + \phi_2) + a_3 \sin(3\omega t + \phi_3) + \dots$$

<sup>24</sup> SIGNAL PROCESSING, Q.,2018. Why is each window/frame overlapping? 2016-12-28 [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://dsp.stackexchange.com/questions/36509/why-is-each-window-frame-overlapping>.

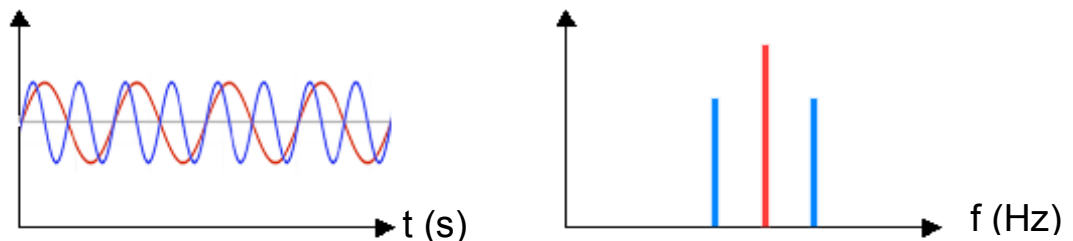
<sup>25</sup>Mriquestions. Transformada de Fourier (FT): preguntas y respuestas en MRI. [en línea]. s.l: Mriquestions. [Citado el 4 mayo, 2018]. Disponible en internet: <<http://mriquestions.com/fourier-transform-ft.html>>

Donde:

- $a_i$  son las amplitudes
- $\phi_i$  son las fases
- $w$  es la frecuencia fundamental

Cómo se puede observar en la figura 6 la transformada de Fourier permite pasar una onda representada en función del tiempo a representarla en el dominio de la frecuencia<sup>26</sup>, caracterizada por el seno y coseno.

Figura 12. Representación de una onda en función del tiempo



Fuente. Los autores.

Es importante conocer la fórmula de Euler que establece la relación entre las funciones trigonométricas y establece que para todo número real  $x$ , siendo  $e$  la base del logaritmo natural e  $i$  es la unidad imaginaria entonces:

$$e^{ix} = \cos(x) + i \sin(x)$$

También se puede representar geoméricamente como una circunferencia con un radio igual a uno en el plano complejo, es medido en el sentido contrario a las agujas del reloj y en radianes. La fórmula solo se cumple si los valores de seno y coseno están en radianes.

La siguiente ecuación es la transformada de Fourier, está compuesta por una función con una frecuencia, que es igual a una integral de menos infinito a infinito de una función en un espacio  $x$  por Euler elevado menos raíz cuadrada de  $-1$  por un espacio  $x$  y una frecuencia.

$$S(\omega) = \int_{-\infty}^{\infty} s(t)e^{-2\pi i\omega t} dt$$

---

<sup>26</sup> JOSÉ MUJICA. Transformada de Fourier. [en línea]. s.l: JOSÉ MUJICA. [Citado el 4 mayo, 2018]. Disponible en internet: <[http://www.escuelasuperiordeaudio.com.ve/articles/fourier discrete.html](http://www.escuelasuperiordeaudio.com.ve/articles/fourier%20discrete.html)>

Donde:

- $S(\omega)$  la señal en el dominio de la frecuencia
- $S(t)$  la señal en el dominio del tiempo
- $e^{-2\pi i\omega t}$  hace uso de la fórmula de Euler

Por su parte la transformada inversa de Fourier está definida como:

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(w)e^{i\omega t} dw$$

Que es usada para realizar la transformación de la señal en el dominio de la frecuencia, al dominio del tiempo.

• **Transformada rápida de Fourier (FFT).**

Es una operación matemática para calcular la transformada discreta de Fourier (DFT), este algoritmo elimina una gran parte de los cálculos repetitivos que se aplican a la DFT permitiendo que el cálculo sea más rápido<sup>27</sup>. La ecuación es la misma transformada de Fourier, pero con una sumatoria, donde la función está dada por una lista de N valores y donde k es una frecuencia discreta.

$$x_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, \dots, N - 1$$

Donde:

- $N$  es el número total de muestras
- $n$  es la muestra actual,  $n \in \{0 \dots N - 1\}$
- $x_n$  el valor de la señal en el tiempo n
- $k$  frecuencia actual,  $k$  de 0Hz a  $N - 1$ Hz
- $x_k$  cantidad de la frecuencia k en la señal (amplitud y fase), es un numero complejo
- $\frac{n}{N}$  es el porcentaje de tiempo
- $2\pi k$  es la velocidad en *radianes/segundo*
- $e^{-i2\pi k \frac{n}{N}}$  representa que tan lejos se movió, a través de la ruta circular, en contra de las manecillas de reloj para esa velocidad y tiempo.

---

<sup>27</sup> Slide Player. La Transformada Rápida de Fourier. [en línea]. s.l: Slide Player [Citado el 8 mayo, 2018]. Disponible en internet: <<http://slideplayer.es/slide/1715431/>>



- **Transformada de coseno discreta (DCT).**

Es una ecuación matemática que se basada en la DFT, expresa una secuencia de varios puntos como el resultado de la suma de diferentes señales<sup>28</sup>, se diferencia de la transformada discreta de Fourier en que se descompone la señal únicamente con la suma de cosenos, es generalmente usada para comprimir voz e imágenes, la ecuación típicamente más utilizada es la siguiente.

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=1}^N y(k) \frac{1}{\sqrt{1 + \delta_{k1}}} \cos\left(\frac{\pi}{2N} (2k - 1)(n - 1)\right)$$

Donde:

- $f_j$  representa la frecuencia sinusoidal
- $x$  representa una secuencia de señal
- $k$  es la posición de la secuencia de la señal<sup>29</sup>.

Es importante usar la transformada discreta de coseno ya que para tiene la cualidad de des correlacionar información que ocasiona el solapamiento en los segmentos del audio.

- **Extracción de características.**

Este paso consiste en identificar los componentes que son buenos de la señal de audio ya pre procesada, esta extracción se puede hacer mediante algoritmos algunos de estos están basados en la percepción auditiva humana, existen los siguientes métodos:

- **Codificación predictiva lineal (LPC).**

Es un modelo simplificado para la producción de voz, este algoritmo es unos de los codificadores estandarizados más antiguos, que funciona a baja velocidad de bits inspirada en observaciones de las propiedades básicas de las señales de voz y representa un intento de imitar el mecanismo de producción de habla humana<sup>30</sup>, es utilizado como una herramienta de extracción de característica, es muy usado para modelar vocales, tiene como principio que la voz es generada de una fuente a

---

<sup>28</sup> Ruye Wang. Discrete Cosine Transform. [en línea]. s.l: Ruye Wang. [Citado el 9 mayo, 2018]. Disponible en internet: <[http://fourier.eng.hmc.edu/e101/lectures/Image\\_Processing/node13.html](http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node13.html)>

<sup>29</sup> DOCUMENTACIÓN MATLAB, 2018. Discrete cosine transform - MATLAB dct. [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://www.mathworks.com/help/signal/ref/dct.html>.

<sup>30</sup> IRCAM. Introducción - Codificación predictiva lineal. [en línea]. s.l: IRCAM [Citado el 10 mayo, 2018]. Disponible en internet: <<http://support.ircam.fr/docs/AudioSculpt/3.0/co/LPC.html>>

través de una función de transferencia, a continuación se muestra el modelo que es equivalente a una señal producida por una ecuación diferencial del habla:

$$s(n) = \sum_{k=1}^p a_k s(n-k) \pm Gu(n) \text{ para } n \in \{1 \dots N-1\}$$

Donde:

- $s(n)$  es la predicción realizada por la ecuación respecto al habla.
- $p$  es el orden de la ecuación diferencial.
- $a_k$  son coeficientes del filtro.
- $s(n-k)$  es la señal de voz.
- $u(n)$  es la señal de excitación o ruido blanco.
- $G$  es un parámetro histórico equivalente al valor del error cuadrático medio.

El  $\pm$  es arbitrario siempre y cuando sea consistente<sup>31</sup>. Esta ecuación tiene como objetivo simular el sistema del habla y predecir el valor de la muestra actual a partir de las  $k$  muestras anteriores, para obtener un buen resultado de esto se busca minimizar el error cuadrático medio:

$$e(n) = s(n) - \sum_{k=1}^p a_k s(n-k)$$

El error se basa en la diferencia del valor real de la señal en la muestra  $n$  y la predicción realizada por la ecuación diferencial.

Para obtener la ecuación diferencial se puede realizar a través del método de autocorrelación, que tiene como objetivo analizar si existe patrones que se repitan en la señal, para ello se tiene que resolver el siguiente conjunto de ecuaciones lineales:

$$\sum_{k=1}^p a_k R(i-k) = R(i) \quad 1 \leq i \leq p$$

---

<sup>31</sup> PELEG, N. Linear Prediction Coding. [en línea]. s.l: [Citado el 10 mayo, 2018]. Disponible en internet: <<http://cs.haifa.ac.il/~nimrod/Compression/Speech/S4LinearPredictionCoding2009.pdf>> p. 7.

Donde:

$$R(k) = \sum_{m=0}^{N-1-k} s(m)s(m+k)$$

El conjunto de ecuaciones lineales a resolver se puede representar en forma de matriz como se muestra a continuación:

Figura 13. Matriz de correlación LPC

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \dots \\ R(p) \end{bmatrix}$$

Fuente. AL-JUNAID, Hessa, SAIF, Amina Mohamed y ALWAZZAN, Fatima Yacooq, 2016. Design of Digital Blowing Detector. *International Journal of Information and Electronics Engineering* [en línea], vol. 6, no. 3, pp. 180-184. [Consulta: 1 noviembre 2018]. ISSN 20103719. DOI 10.18178/IJIEE.2016.6.3.620. Disponible en: <<http://www.ijee.org/index.php?m=content&c=index&a=show&catid=65&id=698>>.

Este sistema de ecuaciones se puede resolver con el procedimiento recursivo Levinson-Durbin ya que la matriz de correlación se Toeplitz.

- **Coefficientes Cepstrales de las frecuencias de Mel (MFCC).**

Son valores para la representación la voz basándose en la captación del oído humano, que tiende a identificar mejor las variaciones en frecuencias bajas a frecuencias altas, esto se entiende como el tono. Los MFCC muestran las características locales de la señal de voz asociadas al tracto vocal, incluida la lengua, los dientes, entre otras.<sup>32</sup>

<sup>32</sup> LLORENTE, C.R., ROBERTO, D., CHICOTE, B., JUAN, D., MONTERO MARTÍNEZ, M., JAVIER, D., GUARASA, M., RUBÉN, D., SEGUNDO, S., SECRETARIO, H., JUAN, D, MONTERO, M, SUPLENTE, M., FERNÁNDEZ, D.F. y CALIFICACIÓN, M. PROYECTO FIN DE CARRERA. [en línea]. S.L: LLORENTE, C.R., ROBERTO, D., CHICOTE, B., JUAN, D., MONTERO MARTÍNEZ, M., JAVIER, D., GUARASA, M., RUBÉN, D., SEGUNDO, S., SECRETARIO, H., JUAN, D, MONTERO, M, SUPLENTE, M., FERNÁNDEZ, D.F. y CALIFICACIÓN, M. PROYECTO FIN DE CARRERA.

Mediante el MFCC se identifica los componentes de la señal de audio que son buenos para determinar el contenido lingüístico y descartar todo lo demás. Para calcular el MFCC primero se segmenta la señal en intervalos cortos: Una señal de audio puede cambiar constantemente, por esta razón se realiza la segmentación a la señal en segmentos de 20 a 40 ms, ya que si es más corto no se tiene suficientes muestras para obtener una estimación espectral confiable y si es más larga la señal cambia demasiado afectara el proceso de análisis espectral, además de esto los segmentos están solapados, para evitar perdida de información de la señal, a estos segmentos se les aplica una ventana de suavizado, en este caso haming para mejorar el resultado análisis espectral mitigando la fuga espectral<sup>33</sup>.

Se pasa la señal que está en el dominio del tiempo al dominio de la frecuencia por medio de la transformada rápida de fourier para todos los segmentos, ya con la señal en el espacio de frecuencia se realiza la transformación a la escala de frecuencia de mel, que equivalen al tono, siendo representados con la siguiente formula:

$$M(f) = 1125 \ln \left( 1 + \frac{f}{700} \right) \text{ o } M(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)$$

De manera más general la formula se puede representar como:

$$m = C \log \left( 1 + \frac{f}{f_0} \right)$$

Donde:

- $C$  es una constante.
- $f$  es la frecuencia a la que se quiere realizar la transformación a escala de mel.
- $f_0$  es la frecuencia de corte.

Esto es resultado de experimentos que realizaron la medición de como escucha el ser humano, de estos resultados se ha comprendido que el

---

[Citado el 10 mayo, 2018]. Disponible en internet: <<http://lorien.die.upm.es/barra/pfcs/2007-carmenr/docs/proyecto.pdf>> p. 77.

<sup>33</sup> Mel Frequency Cepstral Coefficient (MFCC) tutorial. 2012 [en línea], [sin fecha]. [Consulta: 30 octubre 2018]. Disponible en: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>.

oído escucha de manera lineal las frecuencias hasta la frecuencia de corte( $f_0$ ) y por encima de la frecuencia de corte pasa a una escala logarítmica, la frecuencia de corte se elige entre el rango de 600Hz and 1000Hz<sup>34</sup>, para conocer el valor de la constante se hace uso de la siguiente formula:

$$C = \frac{1000}{\log(1 + 1000/f_0)}$$

Ya con la señal en la escala de mel se realiza la transformada discreta de coseno para volver los coeficientes obtenidos a el dominio del tiempo para obtener los coeficientes cepstrales.

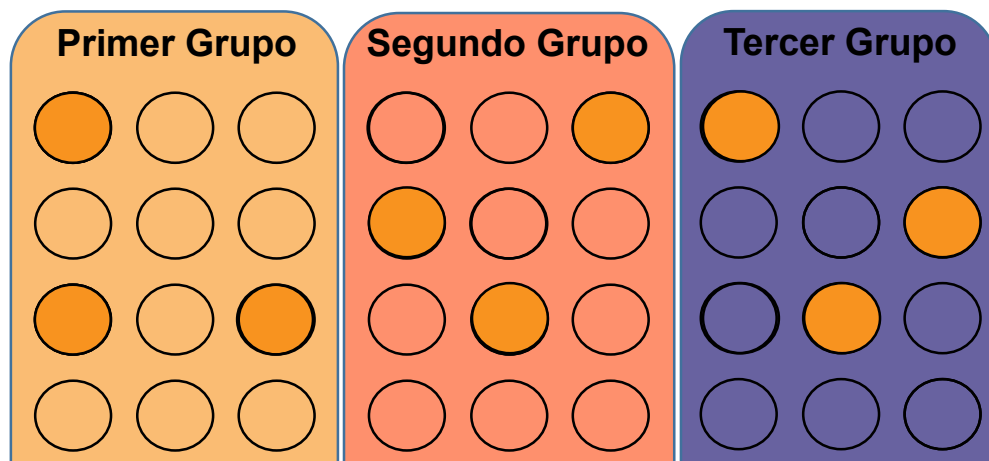
- **Muestreo.**

Este paso consiste en determinar los datos que se usarán para entrenamiento y para pruebas un método usado es el siguiente:

- **Muestreo estratificado.**

Es un tipo de muestreo probabilístico el cual consiste en dividir los datos en diferentes subgrupos y luego seleccionar aleatoriamente sujetos en forma proporcional<sup>35</sup>, esto es especial especialmente útil para balancear las clases y así asegurar un buen entrenamiento adecuado con respecto a las diferentes clases que tiene como objetivo clasificar los modelos de clasificación del aprendizaje de maquina supervisado.

Figura 14. Muestreo estratificado



Fuente. Los autores.

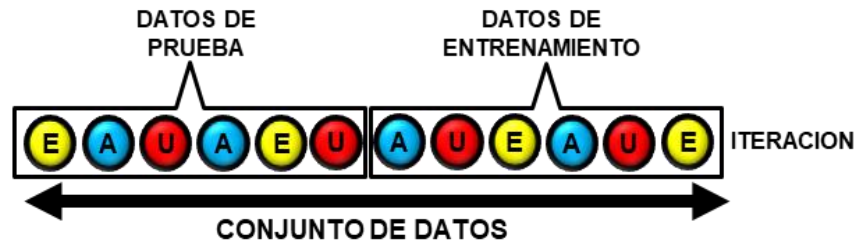
<sup>34</sup> speech processing - The origin of constants in mel-scale formula - Signal Processing Stack Exchange. 2018 [en línea], [sin fecha]. [Consulta: 30 octubre 2018]. Disponible en: <https://dsp.stackexchange.com/questions/46209/the-origin-of-constants-in-mel-scale-formula>.

<sup>35</sup>

- **Validación cruzada.**

Es un método estadístico para evaluar y comparar algoritmos de aprendizaje dividiendo los datos en dos segmentos: uno utilizado para entrenar un modelo y el otro usado para validar el modelo<sup>36</sup>.

Figura 15. Selección de datos de prueba y datos de entrenamiento.



Fuente. Los autores.

La idea principal detrás de la validación cruzada es que cada muestra en nuestro conjunto de datos tiene la oportunidad de ser probado, se realiza un cruce de etapas de entrenamiento y validación en rondas sucesivas.

El error de la validación cruzada es definido por:

$$E = \frac{1}{N} \sum_{i=1}^N E_i$$

Donde:

- $N$  es el número de iteraciones.
- $E_i$ , es el error individual de la iteración.

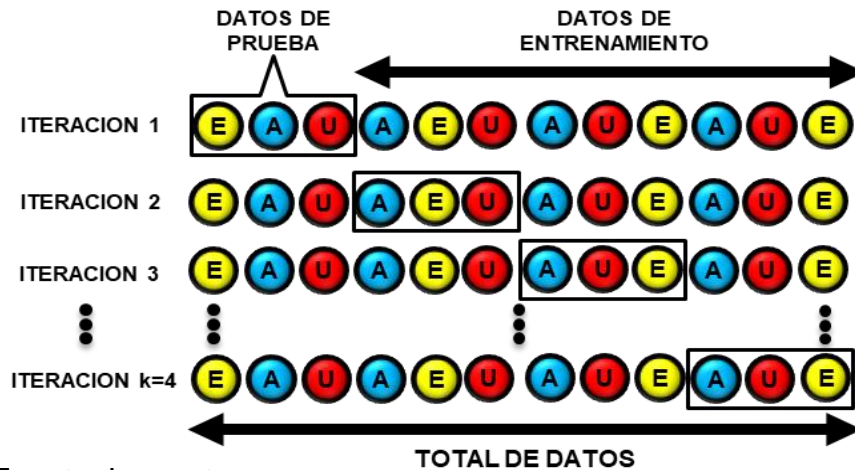
- **Validación cruzada de k-folios.**

Los datos de muestra se dividen en  $K$  subconjuntos, en el que uno de esos subconjuntos se utiliza como datos de pruebas y los otros como datos de entrenamiento. Se realiza este proceso  $k$  veces, cambiando cual es el subconjunto de pruebas, ya cuando se realiza las pruebas con todos los subconjuntos, se analiza la media aritmética para elegir el mejor<sup>37</sup>. es un método muy bueno, pero tiene un gran costo computacional, generalmente se hace con  $k=10$ .

<sup>36</sup> Friedrich. ¿Qué es la validación cruzada en el aprendizaje automático? [en línea]. s.l: Friedrich [Citado el 10 mayo, 2018]. Disponible en internet: <<https://www.quora.com/What-is-cross-validation-in-machine-learning>>

<sup>37</sup> Ingsistemastelesup. Validación cruzada. [en línea]. s.l: Ingsistemastelesup. [Citado el 10 mayo, 2018]. Disponible en internet: <<https://ingsistemastelesup.files.wordpress.com/2017/03/validacion-cruzada.pdf>> p. 1,2.

Figura 16. Representación gráfica validación cruzada k-folios



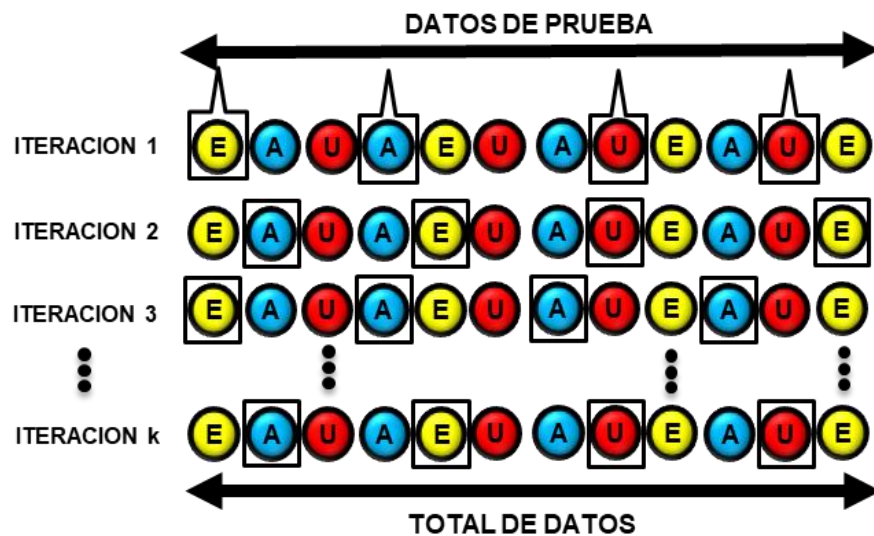
Fuente. Los autores.

En la imagen anterior se puede ver un ejemplo de validación cruzada de k-veces, cuando  $k=4$ , como en cada iteración el subconjunto de pruebas es diferente para comprobarlos.

- **Validación cruzada aleatoria.**

Este método consiste en dividir aleatoriamente los datos que van a pertenecer al conjunto de entrenamiento y al de prueba<sup>38</sup>, esto se itera k veces como se puede observar en la siguiente imagen.

Figura 17. Representación gráfica validación cruzada aleatoria



Fuente. Los autores.

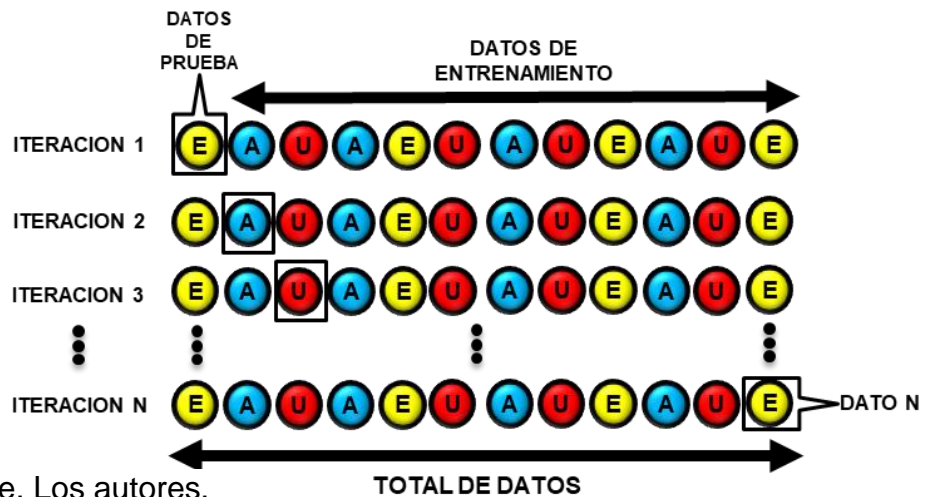
<sup>38</sup> IRCAM. op. cit, p. 30.

En este método algunas muestras pueden quedar sin ser evaluadas y otras pueden ser evaluadas más de una vez. No hay un número de iteraciones definidas.

- **Validación cruzada de dejar un paso.**

Los datos son separados de forma que para cada iteración tengamos una sola muestra para los datos de prueba y los demás datos son los datos de entrenamiento<sup>39</sup>, como se puede observar en la figura 15.

Figura 18. Representación gráfica validación cruzada de dejar un paso.



Fuente. Los autores.

Se debe realizar una iteración por cada dato, esto provoca que el costo computacional sea muy alto, sin embargo, el error que genera este tipo de validación es bajo.

- **Dilema sesgo varianza.**

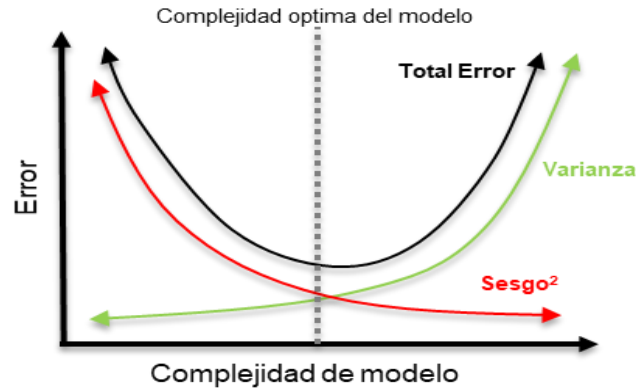
El ajuste ideal del modelo captura las tendencias en los datos lo suficiente como para ser razonablemente preciso y generalizable a un conjunto diferente de puntos de la misma fuente, para hallar la complejidad óptima para el modelo, siendo el punto donde el error total es mínimo<sup>40</sup>.

<sup>39</sup> IRCAM. op. cit, p. 30.

<sup>40</sup> JASON BROWNLEE. Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning. [en línea]. s.l: JASON BROWNLEE [Citado el 12 mayo, 2018]. Disponible en internet: <<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>>



Figura 19. Representación gráfica del sesgo y varianza



Fuente. Los autores.

El error total es definido de la siguiente manera:

$$Error\ Total = sesgo^2 + varianza + error\ irreducible$$

**Sesgo (bias):** Es la diferencia entre el valor esperado y el valor real, un alto sesgo es una sobre generalización del modelo, provocando que el modelo sea demasiado rígido para adoptar la tendencia de los datos de entrenamiento.

**Varianza:** Representa la dispersión de los datos con respecto al valor real, una alta varianza es producida por el sobreajuste del modelo, lo que es igual al error producido por la complejidad del modelo para ajustarse lo más posible a los datos de entrenamiento, al realizar este acercamiento a las tendencias de los datos de entrenamiento pierde la capacidad de generalización.

Figura 20. Relación varianza sesgo.



Fuente. Los autores.

**Error irreducible:** es un ruido que es generado en el proceso de obtención de los datos, es generado por factores como el error humano y la incertidumbre de medición.

- **Algoritmos de clasificación.**

Este paso consiste en comparar los segmentos extraídos del audio con unos ya predefinidos y clasificarlos según una característica, existen los siguientes métodos:

- **Máquina de soporte vectorial (SVM).**

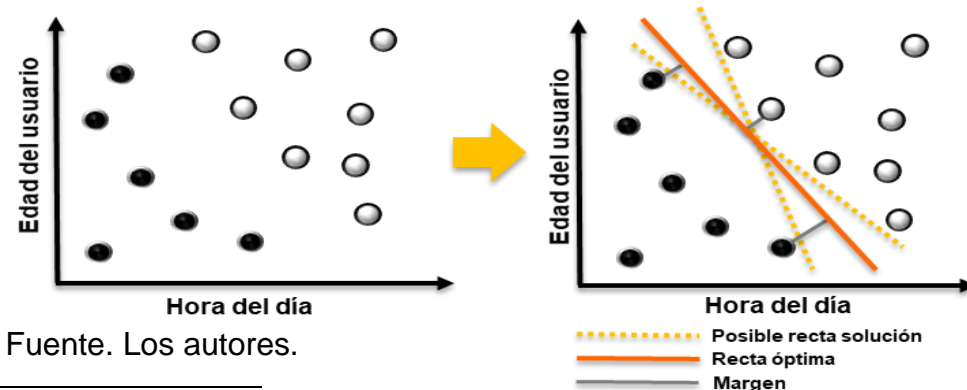
Es un conjunto de métodos relacionados, creados para resolver problemas de clasificación y regresión, es usado en el aprendizaje supervisado<sup>41</sup>.

Una mejora del SVM es llamado, optimización mínima secuencial (SMO), es un método que optimiza la programación cuadrática mediante la descomposición de subproblemas<sup>42</sup>.

En el SVM se realiza una primera etapa de entrenamiento, se le da un conjunto de datos en forma de pares donde se ejemplifique la solución al problema, la segunda fase es donde entra en acción para resolver problemas.

El SVM en problemas de clasificación lineales genera un hiperplano, el cual es definido como una línea en más de 3 dimensiones, ajustado al espacio vectorial original, a partir del conjunto de datos en la etapa de entrenamiento para maximizar el margen entre los elementos y mitiga el error de clasificación como se muestra en la siguiente figura.

Figura 21. Clasificación lineal con SVM.



<sup>41</sup> JANA ÁLVAREZ. Machine Learning y Support Vector Machines Analítica web. [en línea]. s.l: JANA ÁLVAREZ. [Citado el 12 mayo, 2018]. Disponible en internet: <<http://www.analiticaweb.es/machine-learning-y-support-vector-machines-porque-el-tiempo-es-dinero-2/>>

<sup>42</sup> ALBERTO, P. y VALERO, T. Extracción de Información con Algoritmos de Clasificación. [en línea]. s.l: ALBERTO, P. y VALERO, T. [Citado el 12 mayo, 2018]. Disponible en internet: <[https://ccc.inaoep.mx/~mmontesg/tesis\\_estudiantes/TesisMaestria-AlbertoTellez.pdf](https://ccc.inaoep.mx/~mmontesg/tesis_estudiantes/TesisMaestria-AlbertoTellez.pdf)> p. 20.

Existen casos en que los valores no se pueden separar de manera adecuada con un hiperplano, para ello svm busca minimizar el error que está definido con la siguiente función objetivo:

$$\text{minimize } \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^N \xi_i$$

Sujeto a:

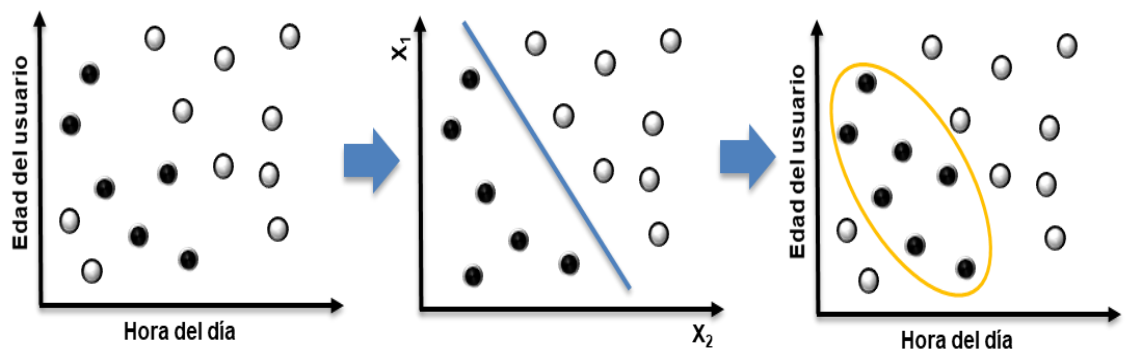
$$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 - \xi \quad \text{para } i=1, \dots, N$$

Donde:

- $N$  es el número de características.
- $\xi$  es la distancia de un punto que se ha clasificado incorrectamente con respecto al hiperplano, se entiende como la variable de penalización.
- $C$  es el parámetro de regularización.
- $\vec{w}$  es el vector con los pesos de las variables y la magnitud equivale al impacto de remover esa variable.
- $\|\vec{w}\|^2$  representa el uso de la norma L2.

Cuando el problema de clasificación es linealmente no separable en el espacio dimensional original, se hace uso de las funciones de kernel, las cuales son funciones matemáticas que superan el problema al pasar a un espacio dimensional mayor en el cual es posible obtener una respuesta del hiperplano lineal de manera fácil, luego se vuelve al espacio dimensional original teniendo, como se muestra en la figura 25.

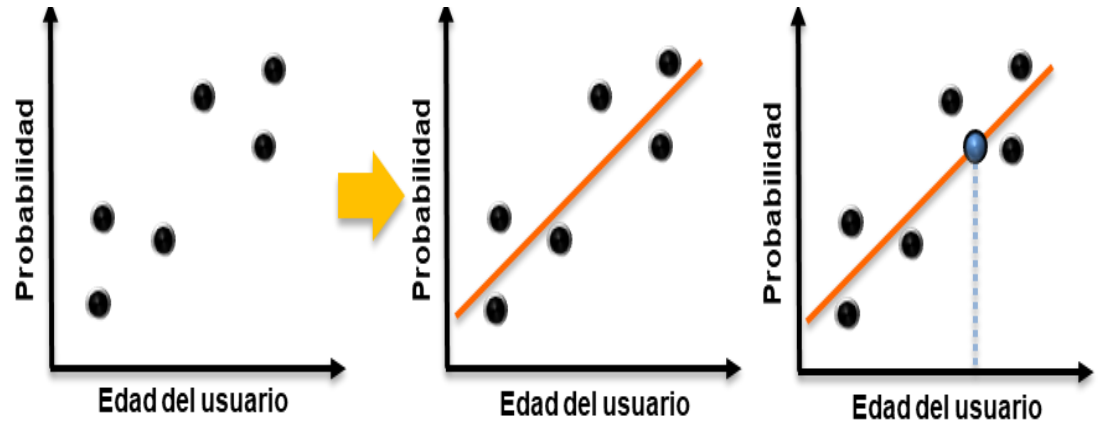
Figura 22. Clasificación no lineal con SVM.



Fuente. Los autores.

SVM en problemas de regresión genera una línea de tendencia minimizando el error para futuros valores, como se observa en la figura 26.

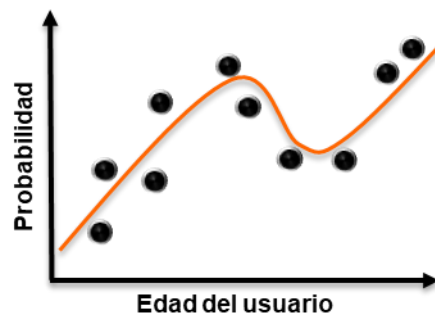
Figura 23. Regresión usando SVM.



Fuente. Los autores.

Al igual que con los problemas de clasificaciones, al no poder generar una solución lineal, hace uso de una función de tipo kernel para hallar la solución, para ello cambia el espacio dimensional original a otro espacio de características, donde si se pueden generar el hiperplano que representa la línea de tendencia de los datos, para luego volver al espacio de dimensional original dando como resultado soluciones como se puede observar en la figura 27.

Figura 24. Regresión usando Kernel



Fuente. Los autores.

Es necesario contemplar correctamente las variables que se van a usar en el SVM, ya que dependiendo de estas se aumenta la dimensión y la complejidad. Los kernels más utilizados son función de base el linear, función de base radial o también conocida como gauss y polinomial. Estos

también hacen uso del parámetro  $\gamma$  que representa cuanta influencia tiene un único ejemplo de entrenamiento, es de gran importancia para la creación de buenos modelos de svm elegir unos buenos valores de  $\gamma$  y el parámetro  $C$ .

- **K vecinos más cercanos (KNN)**

Los K vecinos más cercanos son un método de aprendizaje supervisado, en el que teniendo un conjunto de casos ( $N$ ) como base, cada caso se compone de un conjunto de variables, (datos  $X_1, \dots, X_n$ ) con su respectiva clase( $C$ ) y para cada nuevo caso con su conjunto de variables (datos) tiene como objetivo establecer a que clase corresponde<sup>43</sup>, como se muestra a continuación:

Figura 25. Regresión usando Kernel.

		$X_1$	...	$X_j$	...	$X_n$	$C$
$(\mathbf{x}_1, c_1)$	1	$x_{11}$	...	$x_{1j}$	...	$x_{1n}$	$c_1$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$(\mathbf{x}_i, c_i)$	$i$	$x_{i1}$	...	$x_{ij}$	...	$x_{in}$	$c_i$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$(\mathbf{x}_N, c_N)$	$N$	$x_{N1}$	...	$x_{Nj}$	...	$x_{Nn}$	$c_N$
$\mathbf{x}$	$N + 1$	$x_{N+1,1}$	...	$x_{N+1,j}$	...	$x_{N+1,n}$	?

Fuente. Los autores.

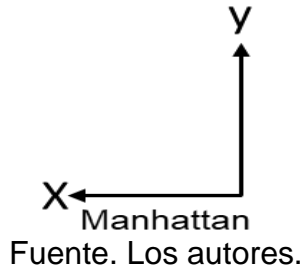
Para clasificar se calcula la distancia entre el nuevo caso con respecto a todos los casos base, las funciones de distancia que se usan son Manhattan, Euclidiana y Minkowsky, la distancia de manhattan se define como:

$$d(x, x_r) = \sum_{j=1}^n |x_{rj} - x_j|$$

Se basa en la distancia que existe entre un par de puntos es la suma de las distancias con respecto a cada coordenada individualmente como se ve gráficamente:

<sup>43</sup> AVINASH NAVLANI, 2018. KNN Classification using Scikit-learn (article) - DataCamp. 2018-08-2 [en línea]. [Consulta: 11 noviembre 2018]. Disponible en: <<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>>.

Figura 26. Representación gráfica distancia de manhattan

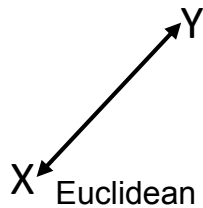


Por su parte, la distancia euclidiana se define como:

$$d(x, x_r) = \sqrt{\sum_{j=1}^n (x_{rj} - x_j)^2}$$

La distancia hace referencia a una línea recta entre los puntos, gráficamente es:

Figura 27. Representación gráfica distancia euclidiana



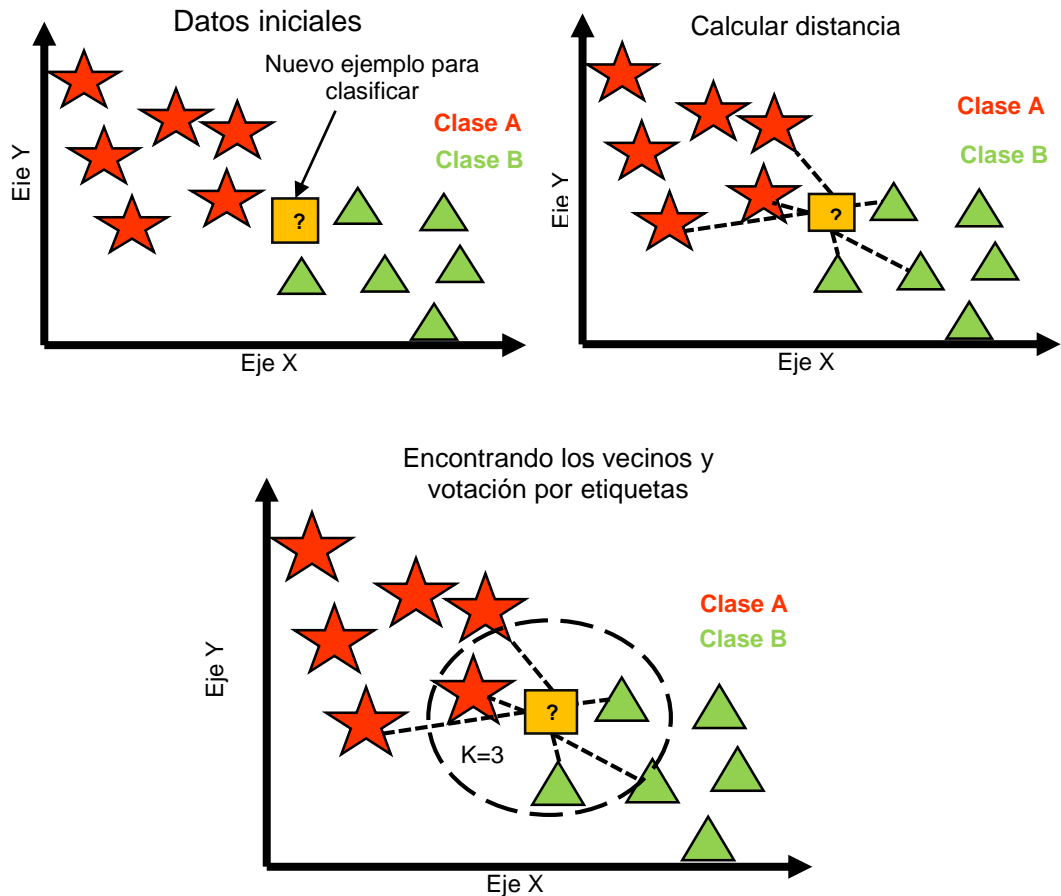
Y la distancia de Minkowski:

$$d(x, x_r) = \sqrt[p]{\sum_{j=1}^n (|x_{rj} - x_j|)^p}$$

El resultado organiza los casos con respecto a los más cercanos al punto y se tienen en cuenta los K casos más cercanos, la variable k que se

establece empíricamente, asignándole la clase más frecuente entre los  $k$  casos<sup>44</sup>. Como se puede apreciar en la siguiente imagen:

Figura 28. Proceso de clasificación  $k$  vecinos más cercanos



Fuente. Los autores.

Para el ejemplo anterior se le asignaría al nuevo caso la clase B ya que hay más casos de esa clase para  $K=3$ .

- **Redes neuronales artificiales (RNA).**

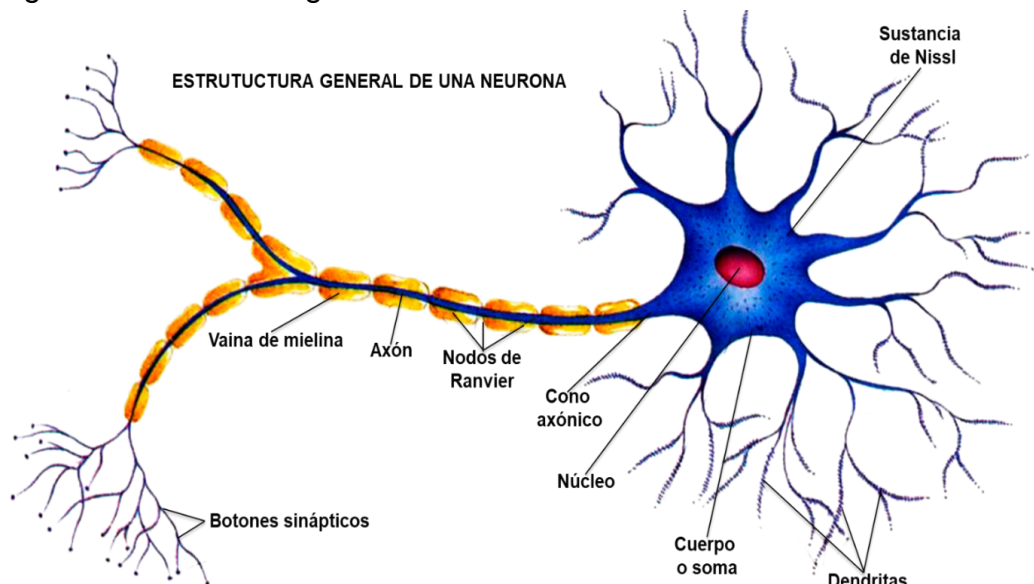
Las RNA tienen como objetivo de emular características propias de los cerebros humanos, como la capacidad de memorizar y de asociar hechos.

<sup>44</sup> AVINASH NAVLANI, 2018. KNN Classification using Scikit-learn (article) - DataCamp. 2018-08-2 [en línea]. [Consulta: 11 noviembre 2018]. Disponible en: <<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>>.

Es un sistema que a través de la experiencia es capaz de adquirir conocimiento, así como lo hacen los humanos<sup>45</sup>.

Las neuronas biológicas por medio de la unión entre ellas la cual denominada sinapsis, puede transmitir información desde el axón hasta las dendritas de la neurona siguiente de forma direccionada y en un solo sentido al llegar al terminal nervioso provoca que se liberen neurotransmisores provocando que la neurona se excite o inhibe dependiendo de si se recibe la información y generando un tipo de respuesta.

Figura 29. Estructura general de una neurona.



Fuente. Vignette. Morfología de la neurona.JPG. [en línea]. s.l: Vignette. [Citado el 3 junio, 2018]. Disponible en internet: <[https://vignette.wikia.nocookie.net/psicologia-145/images/8/83/Morfologia\\_de\\_la\\_neurona.JPG/revision/latest?cb=201511>22205840&path-prefix=es](https://vignette.wikia.nocookie.net/psicologia-145/images/8/83/Morfologia_de_la_neurona.JPG/revision/latest?cb=201511>22205840&path-prefix=es)>

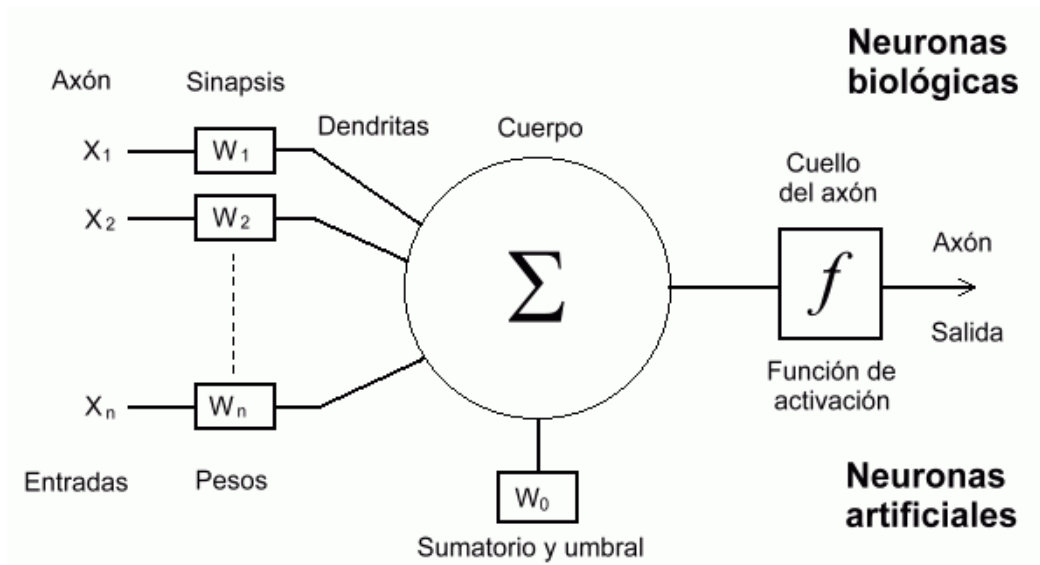
Las neuronas artificiales reciben diferentes entradas  $X$  las cuales tienen relacionadas un peso sináptico  $W$ , adicional se tiene una función modificadora de los pesos ( $w_0$ ) que se va adaptando para llegar a la respuesta óptima, se realiza la sumatoria de las entradas con sus respectivos pesos para obtener la entrada neta (Net) con la cual se utilizara

<sup>45</sup> NICOLÁS SÁNCHEZ ANZOLA. Vista de Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD/COP spot intradiario ODEON. [en línea]. s.l: NICOLÁS SÁNCHEZ ANZOLA. [Citado el 12 mayo, 2018]. Disponible en internet: <<https://revistas.uexternado.edu.co/index.php/odeon/article/view/4414/5256>>



en la función de activación (  $f$  ) que dependiendo del umbral dado provoca que la neurona permanece inactiva o pase su salida  $Y$  a las neuronas siguientes<sup>46</sup>. Se representa el modelo de una neurona artificial.

Figura 30. Red neuronal artificial.



Fuente. Universidad de Murcia. [en línea] [Citado el 1 junio, 2018]. Disponible en: < <http://www.um.es/LEQ/Atmosferas/Ch-VI-3/6-3-8.GIF>.>

También puede representarse de manera matemática, principalmente se obtiene la entrada neta, la cual va a ser utilizada:

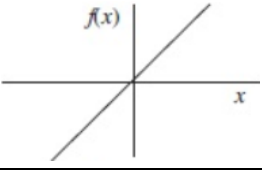
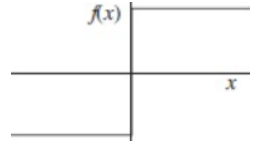
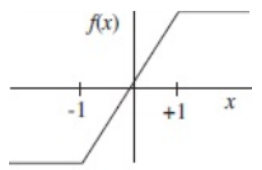
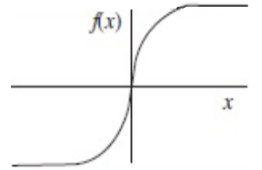
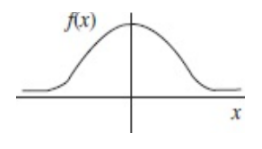
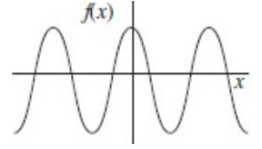
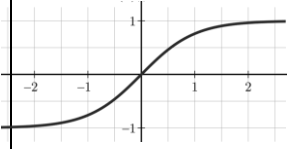
$$z = \sum_{i=1}^N x_i w_{ji} + \theta_j$$

Las principales funciones de activación<sup>47</sup> usadas son:

<sup>46</sup> Fernando Sancho Caparrini. Redes Neuronales: una visión superficial. [en línea]. s.l: Fernando Sancho Caparrini. [Citado el 12 mayo, 2018]. Disponible en internet: <<http://www.cs.us.es/~fsancho/?e=72>>

<sup>47</sup> Ibid. p. 37.

Tabla 1. Funciones de activación.

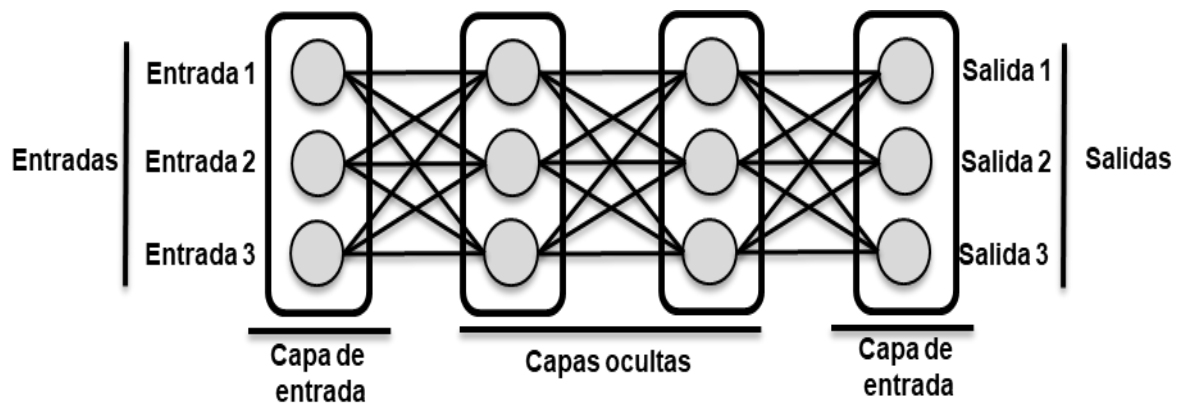
	Función	Rango	Grafica
Identidad	$y = x$	$[-\infty, \infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, 1\}$ $\{0, 1\}$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } l \leq x \leq -l \\ 1, & \text{si } x > l \end{cases}$	$[-1, 1]$	
Sigmoidal	$y = \frac{1}{1+e^{-x}}$	$[0, 1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0, 1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, 1]$	
Tangente hiperbólica	$y = \tanh(x)$	$[-1, 1]$	

Fuente. Los autores.

Las redes neuronales artificiales son capaces de detectar patrones, ya que se trata de imitar a las neuronas biológicas, conectadas entre sí y trabajando en conjunto, aprendiendo sobre el proceso. dadas unas entradas existe una forma de combinar las neuronas para predecir sus resultados, esta combinación está dada a partir del entrenamiento de las redes neuronales ya que es la parte crucial para que el algoritmo sea

preciso en sus resultados<sup>48</sup>. Las salidas están dadas según tres funciones: La función de propagación, función de activación y función de transferencia. La distribución de las neuronas en una RNA está generalizada en tres capas: Capa de entradas, recibe las variables de entrada; las capas ocultas, son las capas internas a la red por lo que no tienen contacto directo con el exterior, pueden ser más de una, contener varias neuronas y estar conectadas de diferentes maneras; y la capa de salidas, son el conjunto de neuronas que transportan la información procesada por la red neuronal al exterior.

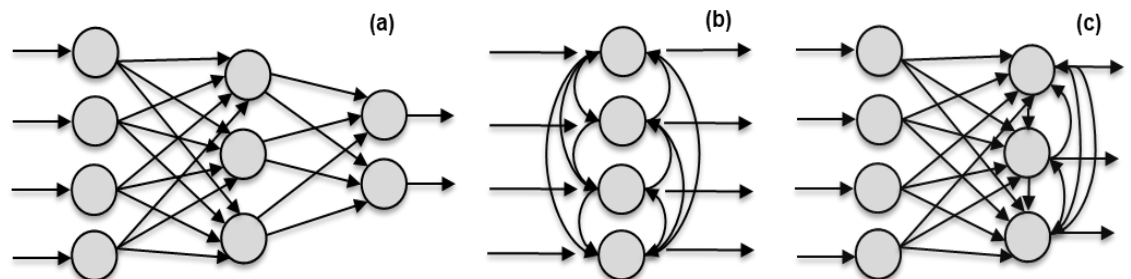
Figura 31. Capas de una red neuronal artificial.



Fuente. Los autores.

Las conexiones de las entre las neuronas son definidas por la estructura de la red neural en lo que puede ser propagación hacia adelante (feedforward), que consiste en que las salidas de las neuronas son las entradas en la capa siguiente de la red neuronal y propagación hacia atrás (feedback), que consiste en que las salidas de las neuronas pueden ser entradas de neuronas de capas anteriores o la misma capa.

Figura 32. Estructuras neuronales.



Fuente. Los autores.

<sup>48</sup> Diego Calvo. Red neuronal Convolutiva CNN. [en línea]. s.l: Diego Calvo. [Citado el 4 junio 2018]. Disponible en internet: <<http://www.diegocalvo.es/red-neuronal-convolutiva-cnn/>>

En las RNA el conocimiento está representado el peso de las conexiones entre las neuronas y el proceso de aprendizaje genera un cambio en el peso de estas conexiones lo que se podría representar como:  $w(t+1)$ , es el peso actualizado;  $w(t)$ , peso actual  $\Delta w$ , variación del peso sináptico.

$$w(t + 1) = w(t) + \Delta w(t)$$

Un aspecto importante y que se puede utilizar para diferenciar las reglas de aprendizaje se basa en los aprendizajes online y offline, teniendo como similitudes que se realiza una etapa de entrenamiento y otra de pruebas, la diferencia es que se mantienen fijos los pesos sinópticos en el entrenamiento off line, mientras que en el entrenamiento online varían cuando se presenta nueva información en el sistema.

Para realizar el proceso de aprendizaje la variación del peso sináptico es necesario utilizar el error global de la red neuronal, definido como error cuadrático medio, es el error que producen las neuronas salidas de la red ante los patrones de aprendizaje que se utilizan en el entrenamiento. siendo P, número de patrones de entrenamiento; N, número de neuronas en la capa de salida;  $y_j$ , es el valor de la salida de la red neuronal y  $d_j$ , es el valor de la salida deseada.

$$Error\ Global = \frac{1}{2p} \sum_{k=1}^p \sum_{j=1}^N (y_j^{(k)} - d_j^{(k)})^2$$

Para realizar la variación en los pesos con el error global se define como:

$$\Delta w_{ji} = k \frac{\partial Error\ Global}{\partial w_{ji}}$$

- **Medidas de desempeño**

En esta etapa se realiza el análisis de las clasificaciones realizadas en el modelo para hallar el porcentaje de exactitud en ellos, la matriz de confusión proporciona es una herramienta de visualización del aprendizaje supervisado<sup>49</sup>, en cada columna se muestra el número de clasificaciones por cada clase y facilitan identificar si el sistema está confundiendo las clases y

---

<sup>49</sup> RENUKA JOSHI. Precisión recuperación y puntaje de F1: interpretación de las medidas de rendimiento - Exsilio Blog. [en línea]. S.L: RENUKA JOSHI. [Citado el 17 mayo, 2018]. Disponible en internet: <http://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>

proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos no tan buenos:

Tabla 2. Medidas de desempeño.

	Negativo	Positivo
Negativo	Verdadero negativo	Falso positivo
Positivo	Falso negativo	Verdadero positivo

Fuente. Los autores.

Las métricas se realizan con respecto a cada clase, teniendo como base la Tabla 2, a continuación, se explicará con más detalle los casos de clasificación:

Verdaderos positivos y verdaderos negativos son las observaciones que se clasifican correctamente.

**Verdaderos positivos:** Son los casos en que el modelo de clasificación asigna adecuadamente la clase a la observación referente, en este caso, las observaciones de la clase positivo se clasifican correctamente como positivo.

**Verdaderos negativos:** Son los casos en los que el modelo clasifica correctamente las observaciones con respecto a las otras clases que no son la referente, en este caso teniendo como referente la clase positivo, clasifico correctamente las observaciones como negativo, es decir, clasifico correctamente la otra clase.

Falsos positivos y falsos negativos, son las observaciones que se clasifican incorrectamente.

**Falsos positivos:** Son los casos en los que el modelo de clasificación se equivoca asignándole a la observación la clase referente, cuando esta es incorrecta, por ejemplo, en este caso negativo es la clase que se debería asignar, sin embargo, el algoritmo los clasifica como positivo.

**Falsos negativos:** Son los casos cuando el modelo asigna otra clase a las observaciones que en realidad son de la clase referente, en este caso clasificar como negativo, las observaciones que en realidad son positivo.

- **Precisión**

Esta métrica muestra el porcentaje de las observaciones que el modelo de clasificación le asignó la clase referente, son verdaderos<sup>50</sup>, un alto porcentaje de precisión se relaciona con la baja tasa de falsos positivos.

$$\begin{aligned} Precision &= \frac{Verdadero\ positivo}{Verdadero\ positivo + falso\ Positivo} \\ &= \frac{Verdadero\ positivo}{Total\ de\ positivos\ predichos} \end{aligned}$$

- **Recall.**

Es la métrica que muestra el porcentaje de observaciones de la clase referente que fueron clasificados correctamente por el modelo, es decir, que tan bueno es el modelo para clasificar las observaciones de la clase referente, un alto porcentaje de recall se relaciona con una baja tasa de falsos negativos.

$$\begin{aligned} Recall &= \frac{Verdadero\ positivo}{Verdadero\ positivo + falso\ negativo} \\ &= \frac{Verdadero\ positivo}{Total\ de\ positivos\ actuales} \end{aligned}$$

- **F1 Score**

Es la media armónica entre las métricas de precisión y el recall.

$$F1 = 2 \left( \frac{Precision \cdot Recall}{Precision + Recall} \right)$$

F1 score suele ser más útil cuando se tiene una distribución de clases desigual.

Si el costo de los falsos positivos y los falsos negativos es muy diferente, es mejor considerar tanto la precisión como recall.

---

<sup>50</sup> KOO PING SHUNG. Accuracy Precision, Recall or F1? – Towards Data Science. [en línea]. S.L: KOO PING SHUNG. [Citado el 17 mayo, 2018]. Disponible en internet: <<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>>

## **1.6. ALCANCES Y LIMITACIONES**

### **1.6.1 ALCANCES**

Inicialmente se realizará un trabajo de campo para obtener la pronunciación de vocales de diferentes personas con el objetivo de crear el conjunto de datos.

Seguidamente se desarrollará un experimento a partir del conjunto de datos haciendo uso de Python y Matlab donde se implementará métodos para las etapas de pre procesamiento, extracción de características, muestreo y clasificación, además se realizará el análisis del desempeño de los modelos implementados.

### **1.6.2 LIMITACIONES**

Para alcanzar los objetivos propuestos, se tomarán en cuenta las siguientes limitaciones:

- Se construirá el conjunto de datos con la pronunciación de las vocales de aproximadamente 2000 audios.
- Las pronunciaciones de vocales serán en el idioma español.
- Se limitará al uso del lenguaje de señas colombiano (LSC).
- El proyecto se realizó en la ciudad de Bogotá, Colombia.
- Los audios se obtuvieron de personas que dieron su consentimiento para la grabación.

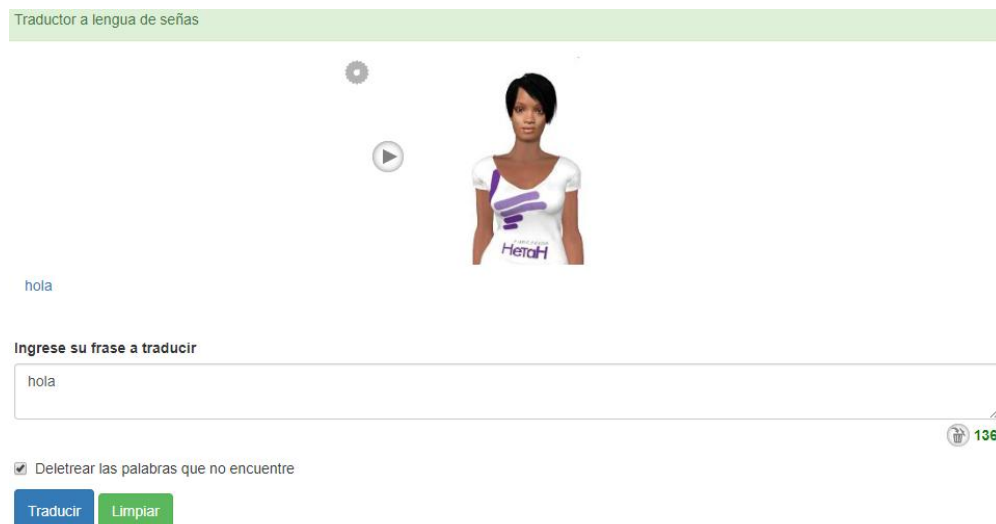
## 1.7. ESTADO DEL ARTE

En esta sección se realizó una búsqueda de investigaciones relacionadas entre los años 2016 – 2018 y diferentes incursiones que se han desarrollado para mitigar la problemática de la comunicación de las personas sordas, las incursiones son las siguientes:

### Hetah.

Es una fundación colombiana que tiene como objetivo identificar, superar y resolver los problemas de la humanidad mediante el desarrollo y la aplicación de tecnología <sup>51</sup>, entre las herramientas tecnológicas de la fundación están: El traductor de texto a braille, el traductor de texto a lenguaje de señas, y el diccionario de lenguaje de señas. El desarrollo de las herramientas es compartido bajo licencia de código abierto en la página de la fundación que permite contribuciones en los avances de estas herramientas.

Figura 33. Imagen del Traductor a lenguaje de señas Hetah.



Fuente. HETAH. Hetah, [en línea]. s.l: HETAH. [Citado el 1 junio, 2018].  
Disponibile en: <<http://hetah.net/>>

El traductor a lenguaje de señas fue desarrollado por el ingeniero de sistemas Jorge Enrique Leal <sup>52</sup>, este traductor permite ingresar texto hasta un máximo de 140 caracteres para su traducción, haciendo uso del diccionario de señas en el cual hay 3063 señas colombianas y 37 venezolanas, las señas son realizadas

<sup>51</sup> MANUEL BENÍTEZ. emtic – Hetah traductor al lenguaje de signos y a Braille. [en línea]. s.l: MANUEL BENÍTEZ. [Citado el 30 abril, 2018]. Disponible en internet: <<https://enmarchaconlastic.educarex.es/244-emic/herramientas-2-0/1296-hetah-traductor-al-lenguaje-de-signos-y-a-braille>>

<sup>52</sup> Ibid, p. 16.



por un avatar llamado “Iris”. Como se observa en la figura 2, si la palabra a traducir no se encuentra en el diccionario tiene la opción de traducir letra por letra de la palabra, si se desactiva se omite su traducción, además permite ajustar la velocidad en que el avatar realiza las señas, por una opción lenta, normal y rápida, sin embargo, para hacer uso de este traductor es necesario conectarse a internet.

### **Centro de Relevo Colombia.**

Es un sistema que permite contactarse con un intérprete que hace de intermediario, permitiendo que una persona sorda y una oyente se puedan comunicar, este sistema puede ser usado mediante la aplicación móvil o del navegador.

El sistema de centro de relevo Colombia<sup>53</sup> cuenta con tipos de servicios para comunicarse, el primero es el “relevo de llamada” donde una persona sorda contacta a un intérprete, esta comunicación es por medio de una video llamada y con la posibilidad de apoyarse con un chat, el intérprete contesta la llamada y hace de intermediario para lograr la comunicación, también puede ser usado por una persona oyente para llamar a una persona sorda, ya que es un servicio bidireccional, su limitante son los horarios los cuales son de 6 de la mañana a 12 de la noche, el segundo servicio es “SIEL (sistema de información eléctrico colombiano)” por medio de un dispositivo móvil, facilita la comunicación de personas en el mismo espacio, las empresas e instituciones públicas podrán tener como apoyo un intérprete con el sistema del centro de relevo Colombia, el último servicio es por medio de “WhatsApp” permite que la persona sorda envíe videos cortos del LSC hacia un intérprete, o una persona oyente envía una nota de voz para resolver consultas, adicionalmente tienen un servicio de reclamos, quejas, peticiones, sugerencias y denuncias.

---

<sup>53</sup> Centro de relevo. Relevo de llamadas. [en línea]. Bogotá: Centro de relevo. [Citado el 30 abril, 2018]. Disponible en internet: <<http://centroderelvo.gov.co/632/w3-channel.html>>

Figura 34. Interfaz gráfica Centro De Relevo Colombia



Fuente: GECKO SAS. Centro de Relevo Colombia Aplicaciones en Google Play. [en línea]. s.l: GECKO SAS. [Citado el 28 abril, 2018].

Disponible en internet:

<<https://play.google.com/store/apps/details?id=com.app.relevo>>

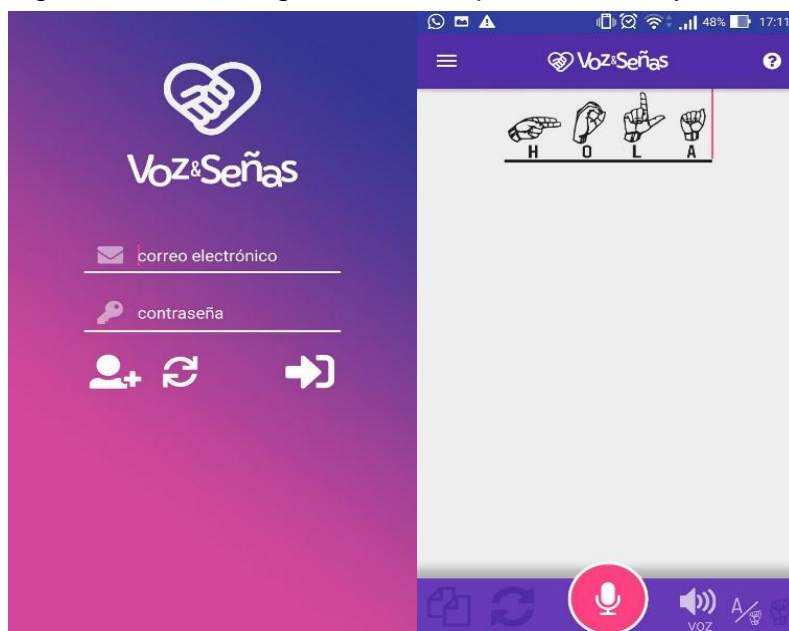
### **Voz & señas.**

Es una App móvil<sup>54</sup> la cual tiene como función traducir de texto a lenguaje de señas mexicano, favoreciendo la comunicación entre una persona sorda y una persona oyente, dentro de sus usos sirve como intérprete.

Las personas sordas pueden escribir en texto lo que desean decir y por un sintetizador de voz poderse comunicar con los de su alrededor, las personas pueden captar la voz por medio de una interfaz de programación de aplicaciones (API) de Google y pasarlo a texto, o simplemente escribir el texto y automáticamente se muestra su equivalente a lenguaje de señas mexicano, sin embargo, se traducen las letras de las palabras.

<sup>54</sup> Voz y Señas - Traductor LSM, Op. Cit. p. 12.

Figura 35. Interfaz gráfica de la aplicación Voz y señas



Fuente. INSTITUTO EN CIENCIAS PEDAGÓGICAS SC. Voz y Señas - Apps en Google Play. [en línea]. s.l: INSTITUTO EN CIENCIAS PEDAGÓGICAS SC. [Citado el 28 abril, 2018]. Disponible en internet: [https://play.google.com/store/apps/details?id=com.fatimsoft.aldo.vozysenas&hl=es\\_419](https://play.google.com/store/apps/details?id=com.fatimsoft.aldo.vozysenas&hl=es_419).

Este es un proyecto organizado por el Instituto de Pedagogía Aplicada en conjunto con Tecno Prótesis y Bienestar Incluyente A.C. es financiada por el apoyo de instituciones y donativos.

### **Hablando con Julis.**

Es una aplicación móvil<sup>55</sup> que ha sido creada con un método inclusivo de aprendizaje enfocado hacia las personas de 3 a 85 años de edad, esta aplicación tiene cuatro principales beneficios: Educación regular, bilingüismo, analfabetismo y discapacidad, su propósito es lograr una comunicación total permitiendo leer y escribir un mensaje, a medida que avanza su uso la persona se irá ejercitando en la pronunciación, vocabulario, intención comunicativa, lectura y escritura, adicionalmente poseen pedagogos para enseñar cómo usar la aplicación, algunos de los beneficios que aporta la aplicación son los siguientes: aprendizaje de conceptos por medio de imágenes, comunicación total, leer y escribir mediante imágenes, pronunciación mediante escucha y repeticiones, aumentar el vocabulario el nivel va aumentando según el proceso pedagógico, la lectura se facilita por medio de la relación que se hace con cada

<sup>55</sup> Google Play. Hablando con Julis. [en línea]. s.l: Hablando con Julis. [Citado el 30 abril, 2018]. Disponible en internet: <<https://play.google.com/store/apps/details?id=io.cordova.julistalkes&hl=es>>

imagen, la escritura por medio de la relación imagen-palabra-voz, mejorar la ortografía y disponible en español e inglés, la aplicación requiere móviles con sistema operativo Android 4.4 y versiones posteriores y adicionalmente una suscripción mensual de 20 dólares para las personas de escasos recursos no podrían tener hacer uso de esta aplicación, ya que aunque es muy buena no pueden invertir en este tipo de aplicaciones por limitaciones económicas.

Figura 36. Interfaz gráfica Hablando con Julis



Fuente. Google Play. ¡Hablando con Julis! [en línea]. s.l: Google Play. [Citado el 28 mayo, 2018]. Disponible en internet: <<https://play.google.com/store/apps/details?id=io.cordova.julistalkes&hl=es>>

### **Vibrador.**

Es un proyecto desarrollado por dos jóvenes paisas, el vibrador es una manilla electrónica que permite a las personas sordas transitar con tranquilidad por las calles; esta manilla es capaz de convertir las ondas sonoras de pitos de autos y motos en señales vibratorias o luminosas que avisan a los usuarios disminuyendo las posibilidades de accidentalidad.

Figura 37. Vibrador



Fuente. Colombia.com. [Citado el 28 mayo, 2018]. Disponible en internet:<<https://www.colombia.com/tecnologia/ciencia-y-salud/sdi/76572/el-vibrador-invento-que-beneficia-a-la-comunidad-sorda>>

Es un reloj liviano y su alcance es de 10 metros suficiente para que la persona tenga tiempo de reaccionar y evitar ser atropellada.

### Investigaciones

Las investigaciones relacionadas entre los años 2016 – 2018 son las siguientes:

Se ha encontrado que: La transformada de Fourier es usada ampliamente en el pre procesamiento de la señal, los coeficientes cepstrales de Mel se aplican para lograr una mejor extracción de característica y que las redes neuronales son muy usadas en la clasificación.

En la investigación Gil L J, Castillo F y Flórez R D<sup>56</sup>, hicieron uso del DTW (Alineamiento temporal dinámico), cruce por cero, HMM (Modelo oculto de Markov) y las redes neuronales profundas con el objetivo de controlar una silla de ruedas mediante comandos de voz, de estos métodos el que dio mejores resultados fueron las redes neuronales, durante las pruebas con mujeres y hombres se logró una exactitud del 99.36% y un 98.29% en la siguiente tabla se muestra la exactitud y la sensibilidad de la tercera prueba en un lugar con ruido.

Tabla 3. Tercera prueba en un ambiente con ruido.

Parámetro	Prueba #3 73 dB(A) hasta 85 dB(A)	
	Mujeres	Hombres
Exactitud	99,36%	98,29%
Sensibilidad	Superior al 94,99% en siete de los comandos. El resto 100%	Superior al 92,4% en catorce de los comandos. El resto 100%

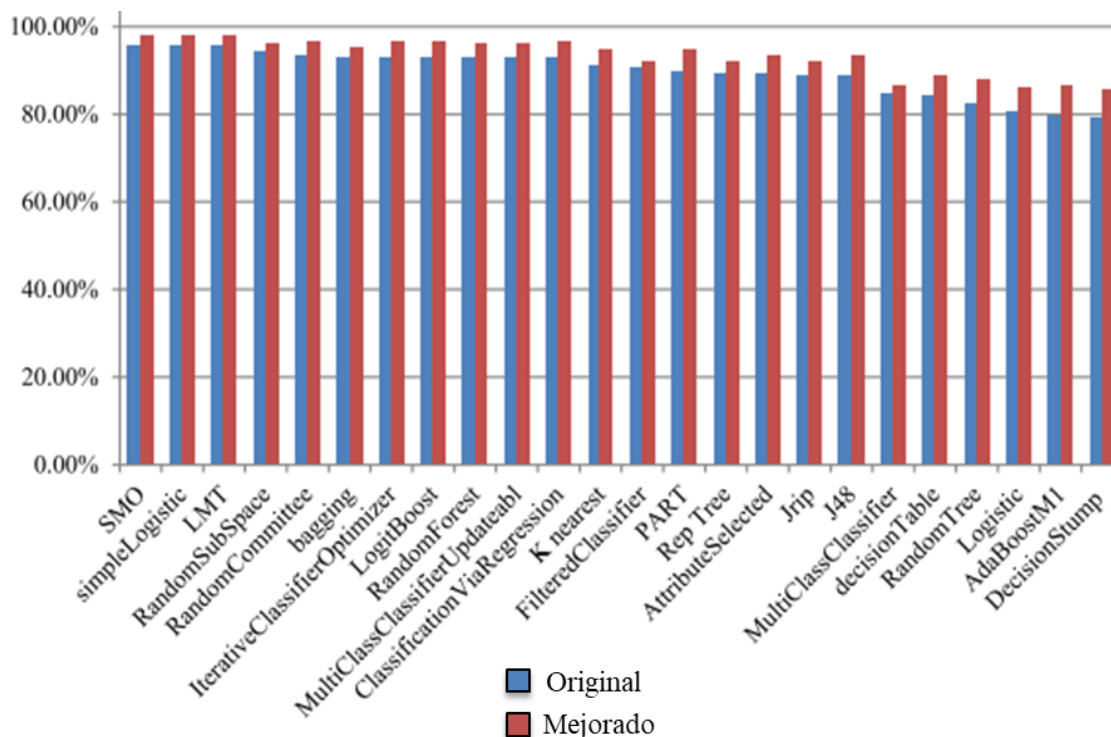
Fuente KLAYLAT, S., OSMAN, Z., HAMANDI, L. y ZANTOUT, R. Enhancement of an Arabic Speech Emotion Recognition System. International Journal of Applied Engineering Research. 2018. p. 4,7.

Otros autores como Klaylat Samira, Osman Ziad, Hamandi Lama y Zantout Rached desarrollaron un proyecto para reconocer tres tipos de emociones (feliz, enojado, sorprendido) a partir de la voz<sup>57</sup>, ellos utilizaron treinta y cinco modelos

<sup>56</sup> GIL, L.J., CASTILLO, L.F. y FLÓREZ, R.D. Reconocimiento de comandos de voz en español orientado al control de una silla de ruedas. UIS Ingenierías, Revista de la facultad de ingeniería físico mecánicas [en línea]. S.L: GIL, L.J., CASTILLO, L.F. y FLÓREZ, R.D. Disponible en internet: <<http://search.ebscohost.com/login.aspx?direct=true&db=fua&AN=121143584&lang=es&site=ehost-live>> p. 10.

<sup>57</sup> KLAYLAT, S., OSMAN, Z., HAMANDI, L. y ZANTOUT, R. Enhancement of an Arabic Speech Emotion Recognition System. International Journal of Applied Engineering Research. 2018. p. 4,7.

de clasificación algunos de ellos son: SMO, simple logistic, LMT, random subspace, random committee, bagging, iterative classifier optimizer, logitBoost, random forest, multiclass classifier updateable, classification via regression, k nearest, filtered classifier, PART, Rep tree, attribute selected, Jrip, j48, decision table, random tree y el mejor de ellos fue el optimización mínima secuencial (SMO), como se observa en la figura 30 logró una precisión del 95.52%.  
 Figura 38. Comparación de algunos métodos de clasificación



Fuente. LATINUS, Marianne y BELIN, Pascal, 2011. Human voice perception. Current Biology [en línea], vol. 21, no. 4, pp. [Citado el 18 mayo, 2018]. Disponible en: <http://linkinghub.elsevier.com/retrieve/pii/S096098221001701X>.>

En la investigación de Gómez Julieth, Simancas José, Acosta Melisa, Meléndez Farid y Vélez Jaime<sup>58</sup>, el análisis rápido de Fourier permite eliminar una gran parte de los cálculos repetitivos que se aplican a la DFT esto permitió que se procesa rápido la voz y los coeficientes cepstrales de Mel pueden dar un resultado óptimo para realizar la clasificación por medio de redes neuronales, sus resultados fueron de un 87.5% de aciertos. También según Dwi Murman, Nishizaki Ichiro, Hayashida Tomohiro y Sekizaki Shinya las redes neuronales

<sup>58</sup> GÓMEZ, J., SIMANCAS, J., ACOSTA, M., MELÉNDEZ, F. y VÉLEZ, J. Algoritmo de reconocimiento de comandos voz basado en técnicas no-lineales. [en línea]. S.L: GÓMEZ, J., SIMANCAS, J., ACOSTA, M., MELÉNDEZ, F. y VÉLEZ, J. Disponible en internet: <<http://repositorio.cuc.edu.co/xmlui/handle/11323/904>> p. 8,17.

profundas con la clasificación MFCC interactúan eficientemente mediante una mejora del rendimiento<sup>59</sup>, el MFCC relaciona la frecuencia de un tono a su frecuencia real de manera similar los humanos diferencian mejor pequeños cambios de tono en bajas frecuencias, esta escala hace que estas características se ajusten más a lo que los humanos escuchan, la ecuación será la siguiente:

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right)$$

Los pasos para usar la ecuación son: Muestrear las señales en tramos de 20 a 40 ms donde f es la frecuencia de los tonos que componen la señal de voz. Después de extraer las características para la implementación de la red neuronal se usó el Matlab para crear entrenar la red y por último para cada palabra se usaron 30 muestras donde 20 fueron para entrenamiento y 10 para pruebas, una vez optimizado se hicieron pruebas, sus resultados fueron de un 99.38% de éxito.

Otros autores como Clemente Eduardo, Vargas Alcira, Olivier Alejandra, Kirschning Ingrid optaron por usar redes neuronales y modelos ocultos de markov<sup>60</sup>, en este caso la exactitud más alta de las redes fue de un 99.1% y de los modelos de un 93%, los recursos aproximados.

---

<sup>59</sup> DWI, M., NISHIZAKI, I., HAYASHIDA, T. y SEKIZAKI, S. Deep Belief Network Optimization in Speech Recognition. International Conference on Sustainable Information Engineering and Technology (SIET). 2017. p. 3,4,5.

<sup>60</sup> CLEMENTE, E., VARGAS, A., OLIVIER, A., KIRSCHNING, I. y CERVANTES, O. Entrenamiento y Evaluación de reconocedores de Voz de Propósito General basados en Redes Neuronales feed-forward y Modelos Ocultos de Markov Eduardo. 2018. p. 8.

## 1.8. METODOLOGÍA

Para dar inicio al proyecto se realizó una investigación sobre las técnicas que se utilizan en diferentes proyectos relacionados con el reconocimiento de voz y lenguaje de señas, estas se clasificaron en cinco fases:

1. Construcción del conjunto de datos.
2. Pre-Procesamiento
3. Extracción de característica
4. Muestreo y validación cruzada.
5. Métodos de clasificación.

Adicionalmente se eligieron los modelos de clasificación con mayor índice de exactitud de los diferentes proyectos.

Las siguientes fases se usarán para hacer el proceso de reconocimiento de voz: **Construcción del conjunto de datos:** En esta etapa se graban, organizan y etiquetan los audios, para ello se realizó un trabajo de campo que consiste en grabar la pronunciación de vocales de varias personas, alrededor de 135 personas para un total de 2000 audios.

Figura 39. Diagrama de flujo metodológico 1.



Fuente. Los autores.

**Preprocesamiento:** Una vez terminado el conjunto de datos se realiza un proceso a cada audio con el objetivo de recortar el audio y obtener solamente el segmento donde se pronuncia la vocal, un audio puro, para ello se aplica un método de eliminación de silencios.



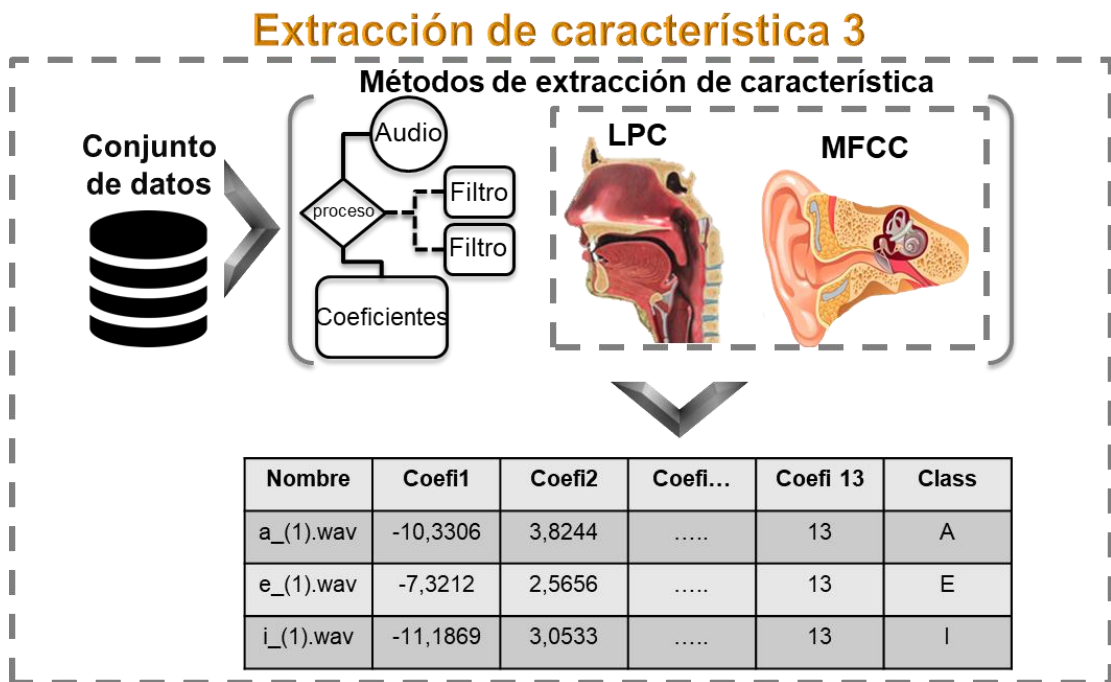
Figura 40. Diagrama de flujo metodológico 2.



Fuente. Los autores.

**Extracción de características:** En esta etapa se extraerán las características más relevantes del audio, con el fin de obtener un vector de características de valores numéricos para ellos se aplicaron dos métodos codificación predictiva lineal y coeficientes centrales de Mel.

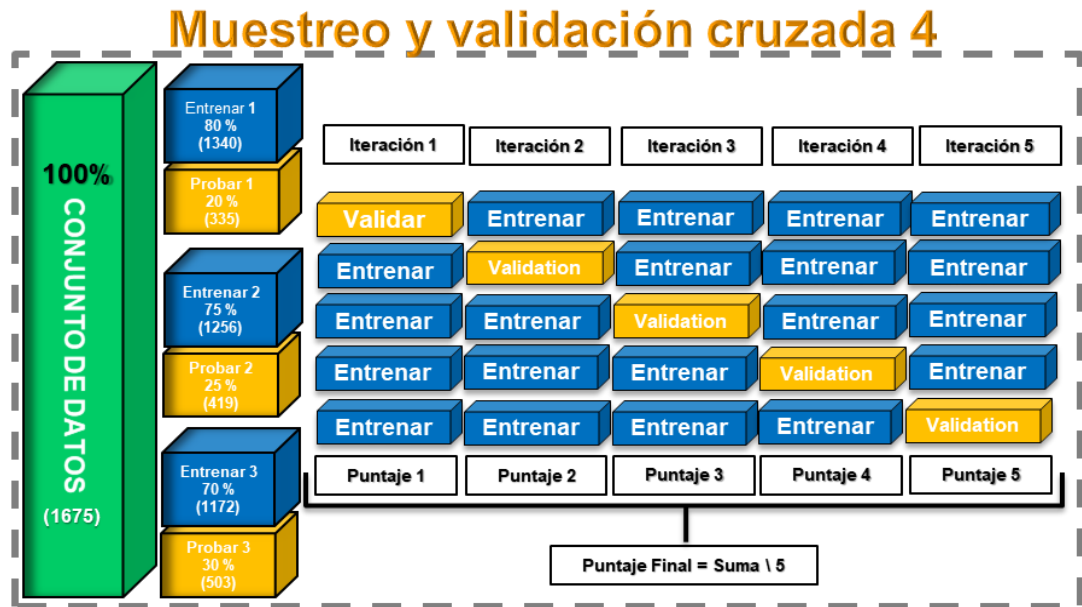
Figura 41. Diagrama de flujo metodológico 3



Fuente. Los autores.

**Muestreo y validación cruzada:** En esta etapa se clasificará el conjunto de datos en dos grupos, los datos de entrenamiento y datos de prueba, dentro de esta etapa se harán uso de métodos de clasificación y la obtención de la precisión de los métodos mediante medidas de desempeño.

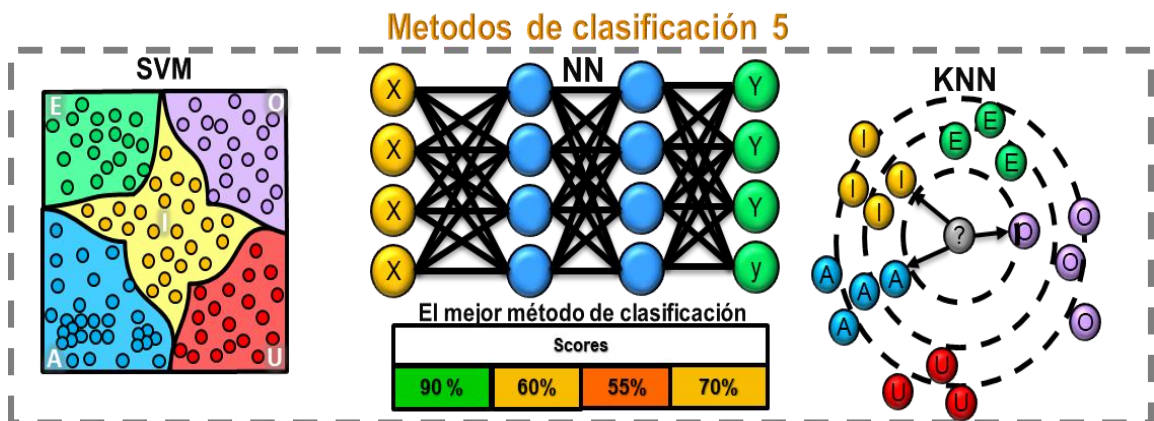
Figura 42. Diagrama de flujo metodológico 4.



Fuente. Los autores.

**Clasificación:** En esta etapa se realizará la etapa de entrenamiento del algoritmo de clasificación con el conjunto de audios, para esto se hizo uso métodos de clasificación: Maquinas de soporte vectorial, Redes neuronales y k vecinos cercanos. Posterior a esto se realizará la etapa de pruebas, en la cual se utilizará un conjunto de datos diferentes a los de entrenamiento para comprobar si los clasifica correctamente en sus respectivas clases.

Figura 43. Diagrama de flujo metodológico 5.



Fuente. Los autores.

**Salida:** En esta etapa se obtendrá la equivalencia del audio a la vocal correspondiente del lenguaje de señas colombiano.

Figura 44. Diagrama de flujo metodológico 6.



Fuente. Los autores.

**Reportes:** En esta etapa se hará uso de las medidas de desempeño como: precisión, recall y F1 Score para realizar el análisis de la exactitud que tienen los modelos en sus predicciones, adicionalmente se hará uso de las matrices de confusión las cuales permiten comparar los datos predichos vs los datos reales.

Tabla 4. Reportes

%Training %Validation	Method	SVM -Linear		
		precision	recall	f1-score
70%-30%	MFCC	0.93	0.92	0.92
75%-25%	MFCC	0.94	0.94	0.94
80%-20%	MFCC	0.93	0.93	0.93

Fuente. Los autores.

## 1.9. DISEÑO METODOLOGICO

### Construcción del conjunto de datos.

Para dar inicio al proyecto inicialmente se construyó un conjunto de datos de la pronunciación de vocales haciendo uso de una grabadora Sony ICD-UX560F y en formato wav, para esta tarea se realizó:

- un trabajo de campo donde se grabó a 135 personas pronunciando las vocales (a,e,i,o y u) y en 3 intervalos de tiempo(1s, 2s y 4s), se recopilaron 2000 audios.
- El etiquetado del conjunto de datos, en este proceso se revisaban los audios recolectados previamente y se le asignaba la etiqueta correspondiente según la clase a la que pertenecía en este caso son 5 clases que representan las vocales.

A continuación, se muestra una evidencia del trabajo de campo realizado, fueron tomadas en la casa del deporte en suba.

Figura 45. Construcción del conjunto de datos 1



Fuente. Los autores.

Figura 46. Construcción del conjunto de datos 2.



Fuente. Los autores.

Figura 47. Construcción del conjunto de datos 5.

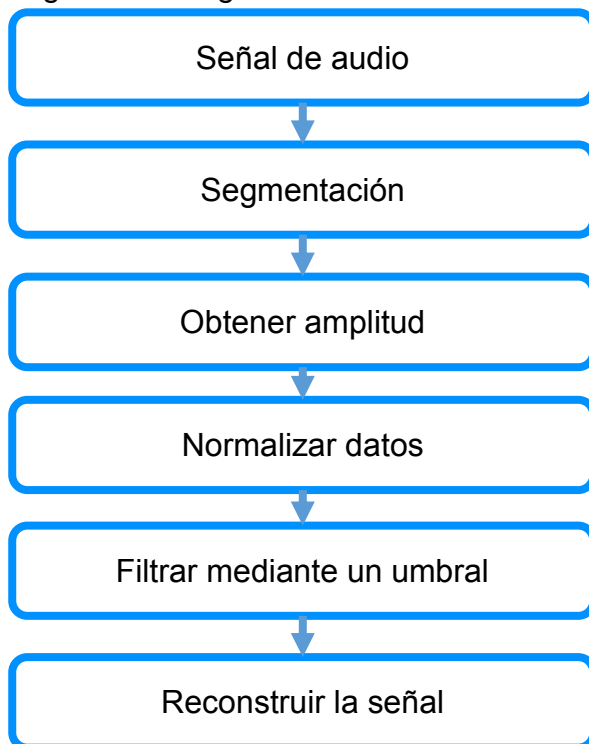


Fuente. Los autores.

## Preprocesamiento.

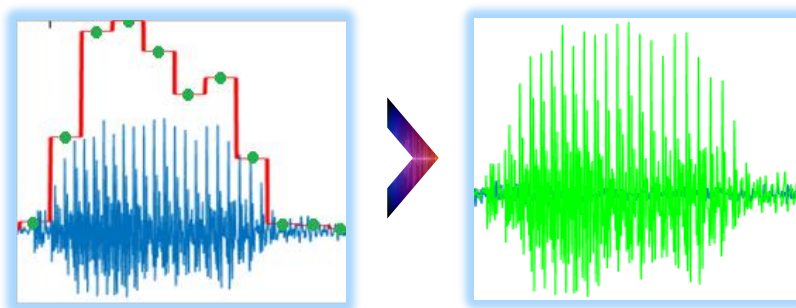
En esta etapa se busca limpiar los diferentes audios del conjunto de datos de información innecesaria en los audios y enfatizar en la pronunciación para este proceso se utilizó el método de eliminación de silencios con los siguientes pasos: segmentación del audio en intervalos de 25ms, obtención de la amplitud de los segmentos haciendo uso de Cálculo de energía a corto plazo, normalización de los datos, especificación un umbral y reconstruir la señal con los segmentos que superaran el umbral deseado.

Figura 48. Diagrama de eliminación de silencios



Fuente. Los autores.

El resultado del audio pre procesado se observa en la siguiente figura.  
Figura 49. Representación gráfica eliminación de silencios



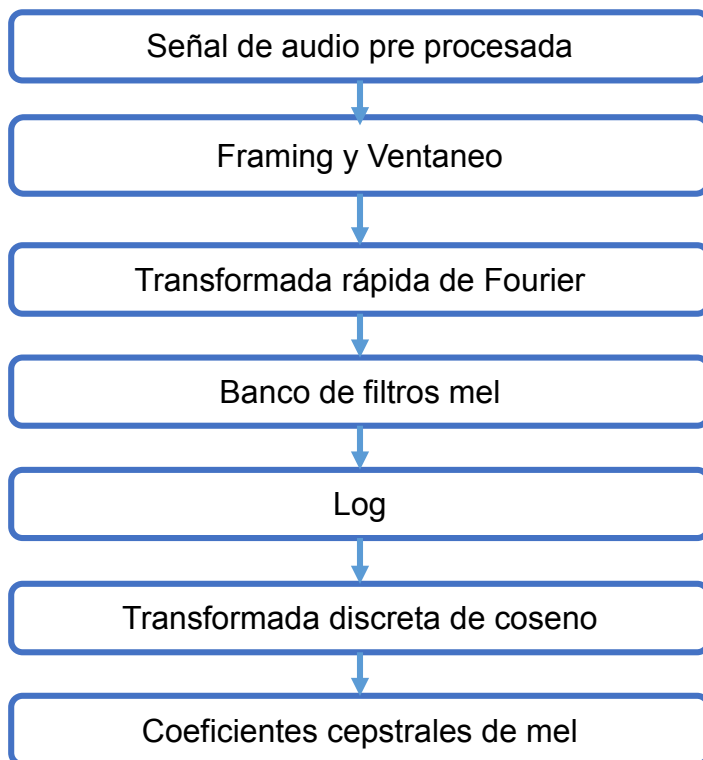
Fuente. Los autores.

### Extracción de características.

En esta etapa se extraerán las características, que son la representación numérica de la señal, del conjunto de datos que ya ha pasado por la etapa de preprocesamiento, para esto se hizo uso de los coeficientes centrales de Mel (MFCC) y codificación predictiva lineal (LPC) los siguientes métodos y sus respectivos pasos:

- MFCC: Inicialmente se ingresará el audio pre procesado donde se la aplicará framing el cual consiste en segmentar el audio en intervalos de ms seguidamente a cada intervalo se le aplicará un ventaneo que consiste en atenuar el inicio y final de cada segmento de audio, seguidamente se le aplicará la transformada rápida de Fourier esta permite transformar el audio de dominio del tiempo (segundos) a dominio de la frecuencia (Hz) una vez ya transformado se le aplicaran una serie de filtros especiales llamados filtros de mel, después se agruparan los segmentos de audio y se les aplicara log y la transformada discreta de coseno por último el resultado final será una matriz de características de 13 columnas que representa el audio ingresado.

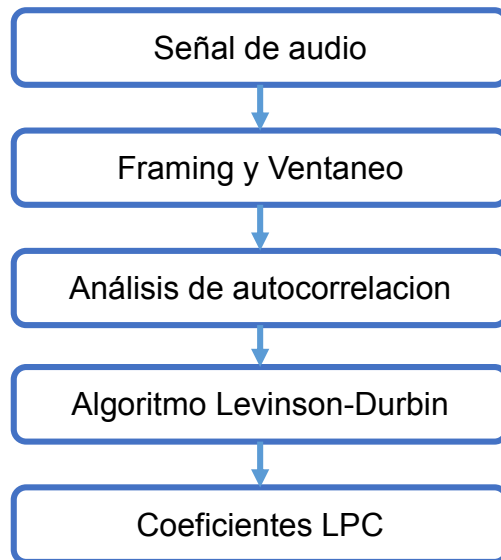
Figura 50. Diagrama de flujo MFCC



Fuente. Los autores.

- LPC: Inicialmente se ingresará el audio ya pre procesado donde se la aplicará framing el cual consiste en segmentar el audio en intervalos de ms seguidamente a cada intervalo se le aplicará un ventaneo que consiste en atenuar el inicio y final de cada segmento de audio, seguidamente se realizará un análisis de autocorrelación y se le aplicará un algoritmo llamado Levinson-Durbin el cual dará como resultado un vector de 47 elementos representativo del audio ingresado.

Figura 51. Diagrama de flujo LPC



Fuente. Los autores.

Una vez obtenidos los coeficientes, se guardan en un archivo de texto, junto con su respectiva etiqueta, así para todos los audio del conjunto de datos, además de los coeficientes individuales de los métodos de extracción de características, MFCC y LPC, se creó un archivo de texto con las características combinadas de ambos métodos.

Figura 52. Representación de las características



LPCs.txt MFCCs.txt TOTAL.txt  
 ha\_1\_ (1).wav, -10.3306, 3.8244, -0.5348, 0.7163, -0.2104, 0.0722,  
 -0.2085, -0.3572, -0.0408, 0.1984, 0.0001, -0.1015, 0.0027, a

Fuente. Los autores.



### **Muestreo, validación cruzada y clasificación.**

Inicialmente se tomarán dos grupos de datos de tres diferentes tamaños para testear y entrenar:

- 80% entrenamiento y 20% pruebas
- 75% entrenamiento y 25% pruebas
- 70% entrenamiento y 30% pruebas

Primero se realiza un muestreo estratificado del conjunto de datos, para obtener el conjunto de entrenamiento y pruebas, para asegurar el balance de las diferentes clases, se realizando pruebas con diferentes distribuciones.

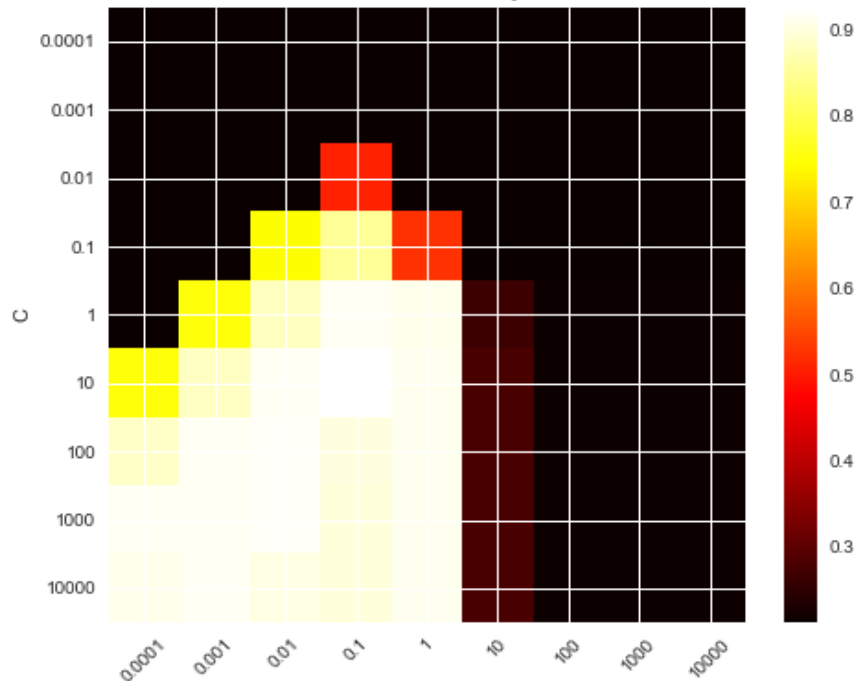
Ya con los conjuntos de entrenamiento y pruebas se crean los modelos de clasificación haciendo uso de los siguientes algoritmos:

- Máquinas de soporte vectorial (SVM)
  - Función de base radial (RBF)
  - Lineal
  - Polinomial
- Redes neuronales (NN)
- K vecinos más cercanos (KNN)

También es importante decir que se realizó la exploración de parámetros en los diferentes modelos de clasificación y se utilizó la validación cruzada 5 folios para evitar el sobreentrenamiento, los diferentes modelos de clasificación fueron entrenados con el mismo conjunto de entrenamiento para poder realizar la comparación.

A continuación, se observa la gráfica de la exploración de parámetros en las máquinas de soporte vectorial, donde el color más claro representa un mejor desempeño por parte de la combinación de los parámetros, la representación visual se usó en los modelos de clasificación SVM, ya que los demás modelos de clasificación tenían demasiados parámetros para poderlos representar gráficamente.

Figura 53. Representación grafica de la exploracion de parametros



Fuente. Los autores.

Ya con el modelo entrenado se utiliza el conjunto de pruebas para observar cómo es su desempeño con datos que no ha visto previamente.

### Reportes.

Teniendo como base las predicciones realizadas por el modelo entrenado con el conjunto de pruebas, se realizan las siguientes medidas de desempeño:

- Precisión
- Recall
- f1 score

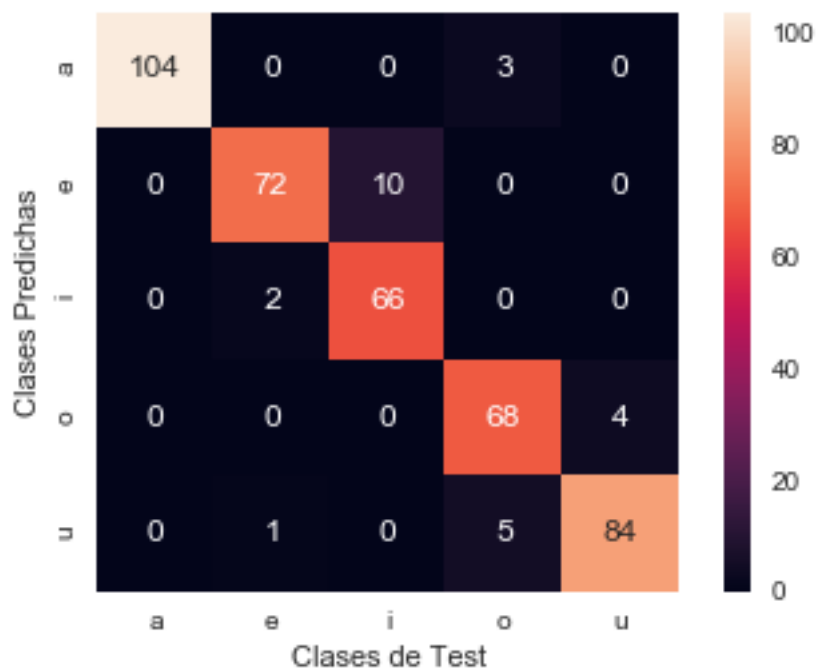
Con el objetivo de conocer la exactitud de los modelos seguidamente del conjunto de pruebas se tomará el porcentaje restante para comparar las vocales predichas con las reales para así obtener una matriz de confusión, adicionalmente matrices de calor donde se podrá observar los mejores parámetros de los métodos de clasificación.

Tabla 5. Métricas de desempeño

precision	recall	f1-score	support	Cantidad de datos
a	0.97	1.00	0.99	104
e	0.88	0.96	0.92	75
i	0.97	0.87	0.92	76
o	0.94	0.89	0.92	76
u	0.93	0.95	0.94	88
<b>avg/total</b>	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>	<b>419</b>

Fuente. Los autores.

Figura 54. Matriz de confusión



Fuente. Los autores.

Después se muestra la representación en el lenguaje de señas colombiano de la clase que el mejor modelo de clasificación le asigna a un audio.

Figura 55. Representación en lenguaje de señas colombiano



Fuente. Los autores.

## 1.1. INSTALACIONES Y EQUIPO REQUERIDO

En esta sección se describen los equipos requeridos directamente en la realización del proyecto.

- Computador usado para el desarrollo y pruebas incluyendo tareas de documentación y codificación con las siguientes características:
- Procesador Intel (R) Core (TM) i7 de sexta generación.
- Memoria RAM de 32 GB.
- 50 GB de espacio libre en disco duro.
- Grabadora Sony ICD-UX560F con las siguientes características:
- Modo de grabación en LPCM formato WAV y MP3.
- Memoria interna de 4 GB.
- Filtro de grabación NCF y LCF.
- Micrófono estéreo.

Las instalaciones requeridas para el desarrollo del proyecto son:

- Instalaciones de la universidad católica de Colombia, incluyendo salones y salas de cómputo.
- Residencia personal con acceso a internet y servicios públicos.

Lenguajes de programación:

- **Python 3.6**, librerías usadas para el desarrollo del experimento: Os, Numpy, Pandas, Matplotlib, Seaborn, Collections, time y sklearn (Kfold, svm, GridSearchCV, Confusión\_matrix, train\_test\_split, classification\_report, cross\_val\_score, joblib, shuffle, MLPClassifier, preprocessing y neighbors).
- **Matlab 2018**, librerías usadas para el desarrollo del experimento: mfcc, lpc, round, sum, reshape, audioread, audiowrite y fopen.

## 1.10. RESULTADOS

Se realizó el análisis de desempeño de los resultados obtenidos con respecto a los diferentes modelos de clasificación, primero se realiza la comparación de los modelos con respecto a las características:

### Máquinas de soporte vectorial

Tabla 5. Comparación de resultados SVM

%entrenamiento- %pruebas	Método	SVM-Lineal			SVM-RBF			SVM-polinomial		
		P	R	F1	P	R	F1	P	R	F1
70%-30%	MFCC	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	LPC	0.80	0.80	0.80	0.54	0.53	0.53	0.45	0.46	0.45
	MFCC+LPC	0.93	0.92	0.92	0.93	0.93	0.93	0.92	0.92	0.92
75%-25%	MFCC	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	LPC	0.82	0.82	0.82	0.54	0.54	0.54	0.50	0.50	0.50
	MFCC+LPC	0.94	0.93	0.93	0.94	0.94	0.94	0.93	0.92	0.92
80%-20%	MFCC	0.93	0.93	0.93	0.95	0.95	0.95	0.93	0.93	0.93
	LPC	0.80	0.80	0.80	0.55	0.56	0.55	0.49	0.50	0.49
	MFCC+LPC	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93

Fuente. Los autores.

Siendo P= precisión, R=recall y F1= f1-score.

Analizando los resultados obtenidos se identifican los mejores resultados con respecto a las diferentes características obtenidas:

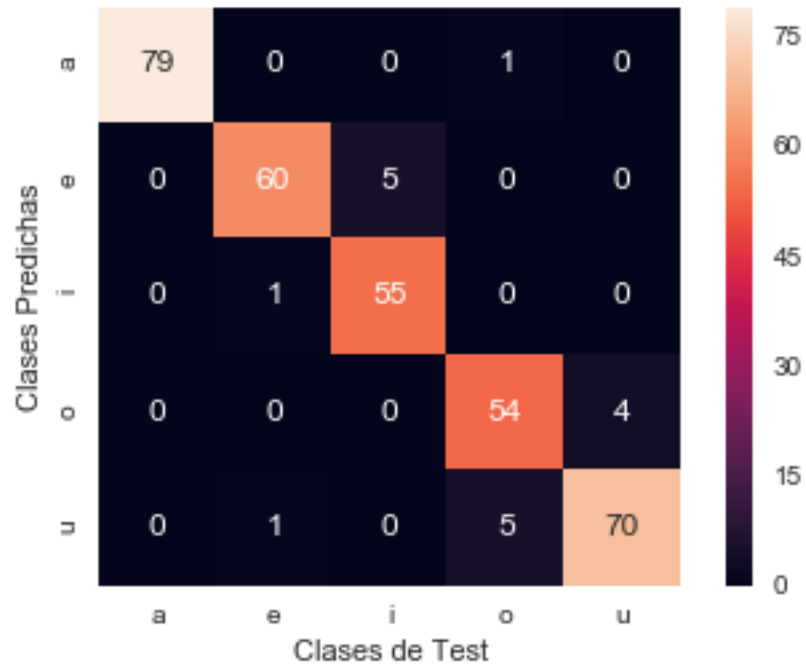
- **MFCC**
  - Distribución de los conjuntos de entrenamiento y pruebas de 80% de entrenamiento y 20% para pruebas
  - Kernel RBF
  - El valor C=10 y gamma 0.1

Tabla 6. Métricas del mejor modelo SVM para MFCC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.99	1.00	0.99	79
e	0.92	0.97	0.94	62
i	0.98	0.92	0.95	60
o	0.93	0.90	0.92	60
u	0.92	0.95	0.93	74
<b>avg/total</b>	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>	<b>335</b>

Fuente. Los autores.

Figura 56. Mejor modelo SVM con MFCC



Fuente. Los autores.

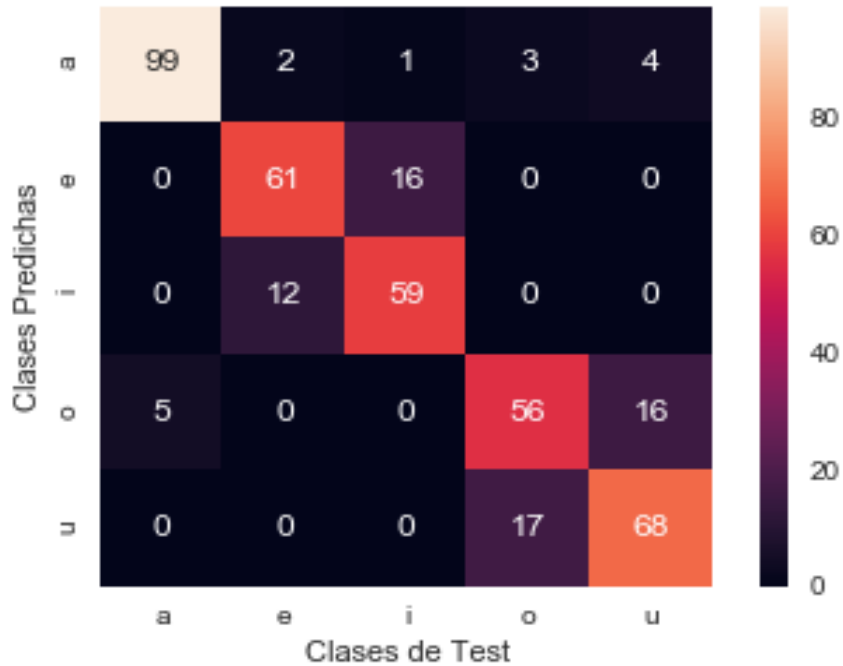
- **LPC**
  - Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
  - Kernel lineal
  - El valor C=10000

Tabla 7. Métricas del mejor modelo SVM para LPC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.91	0.95	0.93	104
e	0.79	0.81	0.80	75
i	0.83	0.78	0.80	76
o	0.73	0.74	0.73	76
u	0.80	0.77	0.79	88
<b>avg/total</b>	<b>0.82</b>	<b>0.82</b>	<b>0.82</b>	<b>419</b>

Fuente. Los autores.

Figura 57. Mejor modelo SVM con LPC



Fuente. Los autores.

- **MFCC+LPC**

- Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
- Kernel rbf
- El valor C=100 y gamma=0.001

Tabla 8. Métricas del mejor modelo SVM para MFCC+LPC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.99	0.99	0.99	104
e	0.87	0.95	0.90	75
i	0.94	0.87	0.90	76
o	0.96	0.89	0.93	76
u	0.92	0.97	0.94	88
<b>avg/total</b>	0.94	0.94	0.94	419

Fuente. Los autores.

Figura 58. Mejor modelo SVM con MFCC+LPC



Fuente. Los autores.

**K vecinos más cercanos:**

Se realizó el análisis de los resultados haciendo uso de los k vecinos más cercanos:

Tabla 9. Comparación de resultados KNN

%entrenamiento- %pruebas	Método	KNN		
		precision	recall	f1-score
70%-30%	MFCC	0.92	0.92	0.92
	LPC	0.33	0.35	0.33
	MFCC+LPC	0.83	0.83	0.83
75%-25%	MFCC	0.92	0.92	0.92
	LPC	0.39	0.39	0.38
	MFCC+LPC	0.84	0.84	0.84
80%-20%	MFCC	0.93	0.93	0.93
	LPC	0.35	0.36	0.35
	MFCC+LPC	0.84	0.84	0.84

Fuente. Los autores.



Se identificaron que los mejores modelos haciendo uso de K vecinos fueron los siguientes:

- **MFCC**
  - Distribución de los conjuntos de entrenamiento y pruebas de 80% de entrenamiento y 20% para pruebas
  - Algoritmo = ball tree
  - El número de vecinos =4
  - distancia = euclidiana
  - pesos en base a la distancia y no de manera uniforme.

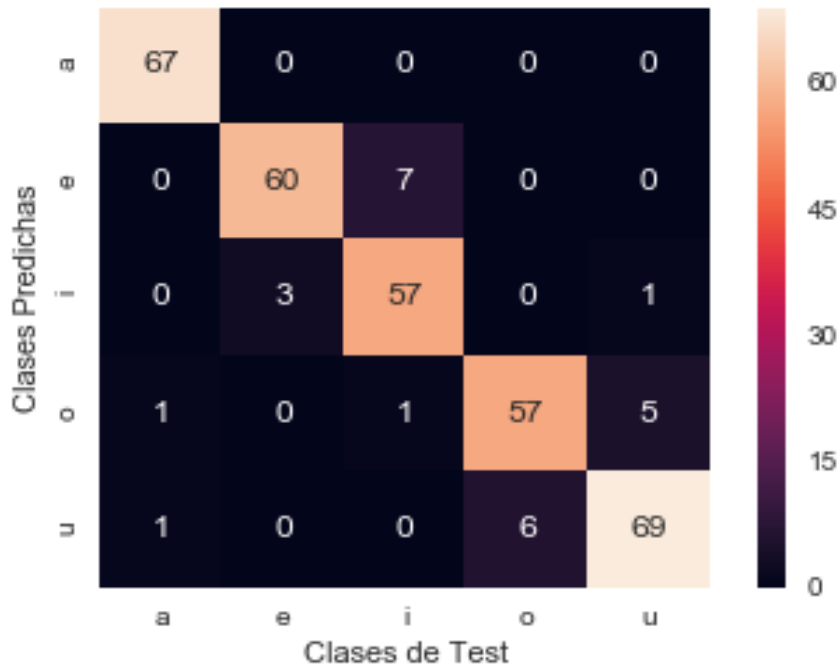
Tabla 10. Métricas del mejor modelo KNN para MFCC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	1.00	0.97	0.99	69
e	0.90	0.95	0.92	63
i	0.93	0.88	0.90	65
o	0.89	0.90	0.90	63
u	0.91	0.92	0.91	75
<b>avg/total</b>	<b>0.93</b>	<b>0.93</b>	<b>0.93</b>	<b>335</b>

Fuente. Los autores.

Figura 59. Mejor modelo KNN con MFCC

Matriz de confusion KNN 20% Test 80% Train MFCCs



Fuente. Los autores.

- **LPC**

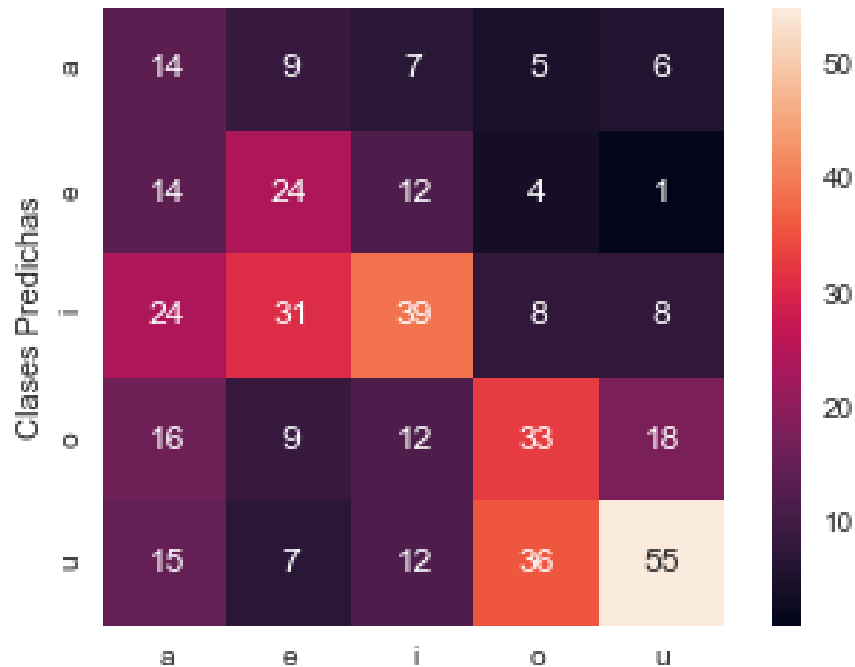
- Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
- Algoritmo = ball-tree
- El número de vecinos =50
- distancia = Minkowsky (10)
- pesos en base a la distancia y no de manera uniforme.

Tabla 11. Métricas del mejor modelo KNN para LPC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.34	0.17	0.23	83
e	0.44	0.30	0.36	80
i	0.35	0.48	0.41	82
o	0.38	0.38	0.38	86
u	0.44	0.62	0.52	88
<b>avg/total</b>	<b>0.39</b>	<b>0.39</b>	<b>0.38</b>	<b>419</b>

Fuente. Los autores.

Figura 60. Mejor modelo KNN con LPC



Fuente. Los autores.

- **MFCC+LPC**

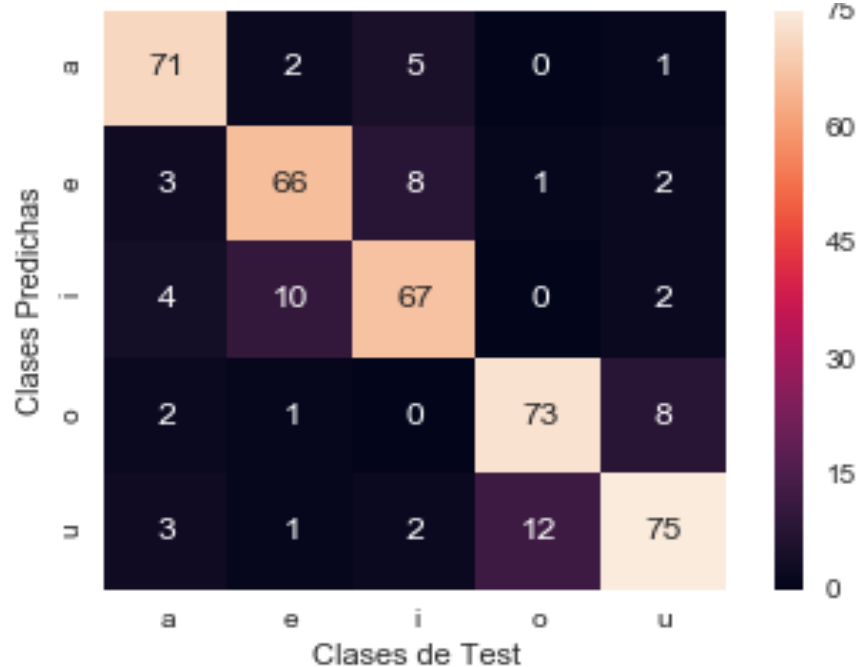
- Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
- Algoritmo = ball tree
- El número de vecinos =15
- distancia = Minkowsky (6)
- pesos en base a la distancia y no de manera uniforme.

Tabla 12. Métricas del mejor modelo KNN para MFCC+LPC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.90	0.86	0.88	83
e	0.82	0.82	0.82	80
i	0.81	0.82	0.81	82
o	0.87	0.85	0.86	86
u	0.81	0.85	0.83	88
<b>avg/total</b>	<b>0.84</b>	<b>0.84</b>	<b>0.84</b>	<b>419</b>

Fuente. Los autores.

Figura 61. Mejor modelo KNN con MFCC+LPC



Fuente. Los autores.

Redes neuronales:

Se realiza el análisis de los resultados haciendo uso de las redes neuronales:

Tabla 13. Comparación de resultados NN

%entrenamiento- %pruebas	Método	NN		
		Precision	Recall	F1-score
70%-30%	MFCC	0.92	0.92	0.92
	LPC	0.74	0.74	0.74
	MFCC+LPC	0.90	0.90	0.90
75%-25%	MFCC	0.93	0.93	0.93
	LPC	0.73	0.73	0.73
	MFCC+LPC	0.93	0.93	0.93
80%-20%	MFCC	0.92	0.92	0.92
	LPC	0.75	0.74	0.74
	MFCC+LPC	0.90	0.90	0.90

Fuente. Los autores.

A continuación, se evidencian los reportes de los mejores desempeños obtenidos de los 3 tipos de métodos de clasificación.

los mejores modelos desempeño con las características fueron los siguientes:

- **MFCC**

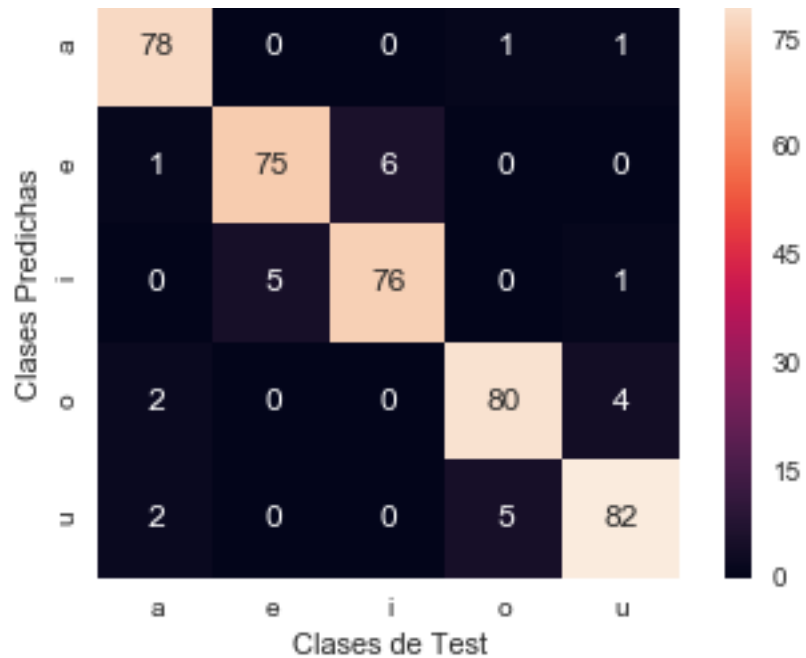
- Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
- Función de activación = logística
- Tamaño de los lotes= 150
- 2 capas y 50 neuronas por capa
- Ratio de aprendizaje =1
- Optimizador = lbfgs

Tabla 14. Métricas del mejor modelo NN para MFCC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.97	0.94	0.96	83
e	0.91	0.94	0.93	80
i	0.93	0.93	0.93	82
o	0.93	0.93	0.93	86
u	0.92	0.93	0.93	88
<b>avg/total</b>	<b>0.93</b>	<b>0.93</b>	<b>0.93</b>	<b>419</b>

Fuente. Los autores.

Figura 62. Mejor modelo NN con MFCC



Fuente. Los autores.

### LPC

- Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
- Función de activación = identidad
- Tamaño de los lotes= 150
- 2 capas y 50 neuronas por capa
- Ratio de aprendizaje =1
- Optimizador = lbfgs

Tabla 15. Métricas del mejor modelo NN para LPC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.82	0.93	0.87	69
e	0.74	0.67	0.70	63
i	0.75	0.77	0.76	65
o	0.62	0.75	0.68	63
u	0.81	0.61	0.70	75
<b>avg/total</b>	<b>0.75</b>	<b>0.74</b>	<b>0.74</b>	<b>335</b>

Fuente. Los autores.

Figura 63. Mejor modelo NN con LPC



Fuente. Los autores.

- **MFCC+LPC**

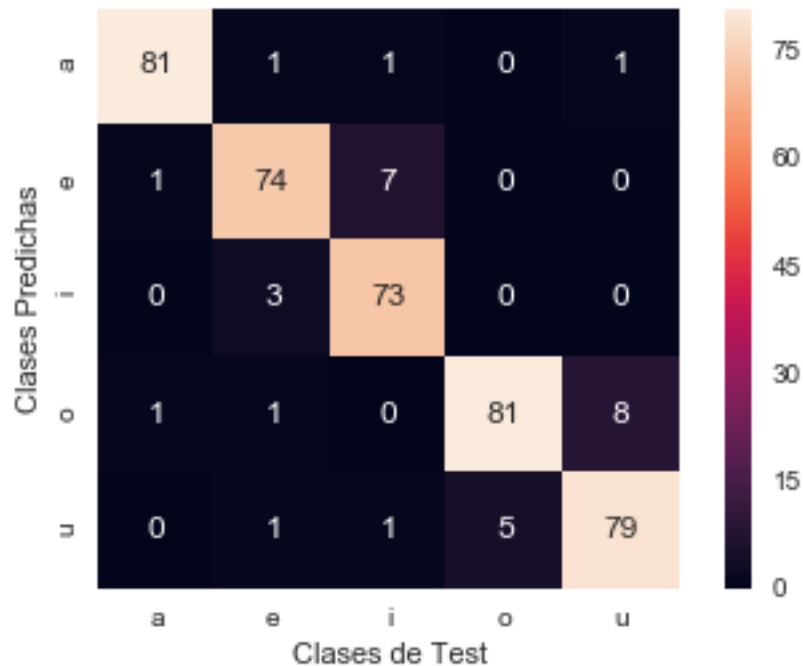
- Distribución de los conjuntos de entrenamiento y pruebas de 75% de entrenamiento y 25% para pruebas
- Función de activación = logística
- Tamaño de los lotes= 150
- 2 capas y 50 neuronas por capa
- Ratio de aprendizaje =1
- Optimizador = lbfgs

Tabla 16. Métricas del mejor modelo NN para LPC

Clase	Precision	Recall	F1-score	Cantidad de datos
a	0.96	0.98	0.97	83
e	0.90	0.93	0.91	80
i	0.96	0.89	0.92	82
o	0.89	0.94	0.92	86
u	0.92	0.90	0.91	88
<b>avg/total</b>	<b>0.93</b>	<b>0.93</b>	<b>0.93</b>	<b>419</b>

Fuente. Los autores.

Figura 64. Mejor modelo NN con MFCC+LPC



Fuente. Los autores.

Para finalizar también se realiza un artículo científico donde se demuestran el proceso y resultados obtenidos del proyecto, se puede encontrar el enlace a este como anexo C.

### 1.11. DISCUSIONES

Durante el desarrollo del proyecto se observó que:

- Las vocales que tienden a confundir los modelos de clasificación son la "E" junto con la "I" y la "O" junto con la "U", ya que a veces se pronuncian de manera similar.
- El uso de las características obtenidas de MFCC en los clasificadores dio muy buen rendimiento, mientras que el uso de las características LPC varió enormemente con respecto al clasificador utilizado, finalmente la concatenación de las características MFCC y LPC dio mejores resultados que LPC, pero menos que MFCC.
- Se observó que cuanto mayor es la cantidad de datos destinados a entrenar el modelo, mejor rendimiento se obtiene.

## **1.12. ESTRATEGIAS DE COMUNICACIÓN Y DIVULGACIÓN**

Para dar visibilidad a los resultados de este proyecto se busca hacer uso de los espacios de socialización de trabajos de grado, además mediante el repositorio institucional de la universidad católica de Colombia en donde se tienen almacenados los trabajos de grado aprobados, adicionalmente la sustentación del proyecto ante los jurados y el artículo de investigación.



### 1.13. CONCLUSIONES

- Haciendo uso del reconocimiento de voz se puede mitigar la brecha que existe en la comunicación entre una persona sorda y uno oyente, ya que por parte de la persona oyente puede transmitir su mensaje al lenguaje de señas colombiano para que la persona sorda pueda entenderlo.
- Comprendiendo mejor la traducción entre el español y el lenguaje de señas colombiano, se podría ofrecer subtítulos en lenguaje de señas permitiendo que puedan entender lo que se dice en videos o audios para tener el mismo derecho de disfrutar estos contenidos que una persona oyente.
- El proceso de la construcción del conjunto de datos es una tarea que conlleva mucho tiempo y esfuerzo ya que requirió de un trabajo de campo donde se solicitaba a una gran cantidad de personas permiso para poder grabarlas pronunciando las vocales, además se tuvo en cuenta que en el espacio que se grabó hubiera el menor ruido posible para evitar problemas a la hora de pre procesar el audio.
- El modelo de clasificación que obtuvo mejores resultados fue el MFCC que en todos los casos tenía desempeños superiores al 90% en las diferentes métricas de desempeño, se observó también que los resultados con LPC no dieron resultados muy altos en la mayoría de los casos, sin embargo, haciendo uso de la concatenación de las MFCC+LPC se obtuvieron buenos resultados, pero sin lograr superar a MFCC.
- Se observó que la mayoría de los modelos de clasificación obtuvieron mejores resultados con una distribución del conjunto de datos del 25% para el conjunto de testeo y 75% para el conjunto de entrenamiento.

## 1.14. TRABAJOS FUTUROS

Se recomienda que para posibles trabajos futuros se tengan en cuenta los siguientes puntos:

1. Aplicar nuevas técnicas para el Reconocimiento de voz en ambientes ruidosos: La comunicación se realiza en cualquier momento y lugar, es por eso que se deben realizar un análisis de las técnicas y métodos para obtener un pre procesamiento eficiente.
2. Reconocimiento automático de palabras: ampliar el alcance del proyecto para poder tener la capacidad de identificar palabras, iniciando principalmente con las más utilizadas para poder abarcar vocabulario y estar más cerca de una solución que pueda mitigar la brecha de comunicación.
3. Análisis del lenguaje de señas colombiano y el español: Es importante comprender la estructura gramatical de las frases del lenguaje de señas colombiano, para poder realizar una traducción adecuada entre los dos lenguajes.

## 1.15. REFERENCIAS BIBLIOGRÁFICAS

ALBERTO, P. y VALERO, T. Extracción de Información con Algoritmos de Clasificación. [en línea]. s.l: ALBERTO, P. y VALERO, T. [Citado el 12 mayo, 2018]. Disponible en internet: <[https://ccc.inaoep.mx/~mmontesg/tesis\\_estudiantes/TesisMaestria-AlbertoTellez.pdf](https://ccc.inaoep.mx/~mmontesg/tesis_estudiantes/TesisMaestria-AlbertoTellez.pdf)> p. 20.

ALEJANDRO OVIEDO. Colombia atlas sordo – Cultura Sorda. [en línea]. s.l: ----- . [Citado el 1 mayo, 2018]. Disponible en internet: <<http://www.cultura-sorda.org/colombia-atlas-sordo/>>

CASTRO, M.L., RUIZ, A.J., CÉSAR JIMÉNEZ, J., PATRICIA, N., ESPINOSA. Estadísticas e información para contribuir en el mejoramiento de la calidad de vida de la población sorda colombiana. [en línea]. s.l: CASTRO, M.L., RUIZ, A.J., CÉSAR JIMÉNEZ, J., PATRICIA, N., ESPINOSA [Citado el 7 junio 2018]. Disponible en internet: <[http://www.insor.gov.co/historico/images/boletín\\_observatorio.pdf](http://www.insor.gov.co/historico/images/boletín_observatorio.pdf)>

CLEMENTE, E., VARGAS, A., OLIVIER, A., KIRSCHNING, I. y CERVANTES, O. Entrenamiento y Evaluación de reconocedores de Voz de Propósito General basados en Redes Neuronales feed- forward y Modelos Ocultos de Markov Eduardo. 2018. p. 8.

CONGRESO DE COLOMBIA. E 1996 Ley 324 de 1996 - Normas a favor de la Población Sorda. Ley. [en línea]. Bogotá: CONGRESO DE COLOMBIA. [Citado el 1 mayo, 2018] Disponible en internet: <[https://puntodis.com/wp-content/uploads/2015/12/Ley\\_324\\_de\\_1996.pdf](https://puntodis.com/wp-content/uploads/2015/12/Ley_324_de_1996.pdf)> p. 1,2,3.

Centro de relevo. Relevo de llamadas. [en línea]. Bogotá: Centro de relevo. [Citado el 30 abril, 2018]. Disponible en internet: <<http://centroderelvo.gov.co/632/w3-channel.html>>

-----, [en línea]. s.l: Centro de relevo. [Citado el 28 marzo, 2018]. Disponible en internet: <<http://centroderelvo.gov.co/632/w3-channel.html>>

DEFINICION.DE, Definición de voz - Qué es, Significado y Concepto. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <https://definicion.de/voz/>.

DOCUMENTACION MATLAB, 2018. Hamming window - MATLAB hamming. 2006 [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://www.mathworks.com/help/signal/ref/hamming.html>.

-----, 2018. Discrete cosine transform - MATLAB dct. [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://www.mathworks.com/help/signal/ref/dct.html>.

DWI, M., NISHIZAKI, I., HAYASHIDA, T. y SEKIZAKI, S. Deep Belief Network Optimization in Speech Recognition. International Conference on Sustainable Information Engineering and Technology (SIET). 2017. p. 3,4,5.

Diego Calvo. Red neuronal Convolutacional CNN. [en línea]. s.l: Diego Calvo. [Citado el 4 junio 2018]. Disponible en internet: <<http://www.diegocalvo.es/red-neuronal-convolutacional-cnn/>>

DotCSV. ¿Qué es el Aprendizaje Supervisado y No Supervisado? [en línea]. s.l: DotCSV [Citado el 2 mayo, 2018]. Disponible en internet: <<https://www.youtube.com/watch?v=oT3arRRB2Cw>>

EDUARDO LLEIDA SOLANO, 2000. Definición de frecuencia - Qué es, Significado y Concepto. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <<http://physionet.cps.unizar.es/~eduardo/investigacion/voz/rahframe.html>>.

POLAVIDE.ES, el sonido. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <<http://www.polavide.es/energyluzsonido/sonido.html>>.

FUNDACIÓN GENERAL CSIC. Lychnos cuadernos de la Fundación General CSIC. [en línea]. s.l: Fundación General CSIC. [Citado el 1 mayo, 2018]. Disponible en internet: <[http://www.fgcsic.es/lychnos/es\\_es/articulos/inteligencia\\_artificial](http://www.fgcsic.es/lychnos/es_es/articulos/inteligencia_artificial)>

Fernando Sancho Caparrini. Aprendizaje por refuerzo: algoritmo Q Learning. [en línea]. s.l: Fernando Sancho Caparrini. [Citado el 4 mayo, 2018]. Disponible en internet: <<http://www.cs.us.es/~fsancho/?e=109>>

------. Redes Neuronales: una visión superficial. [en línea]. s.l: Fernando Sancho Caparrini. [Citado el 12 mayo, 2018]. Disponible en internet: <<http://www.cs.us.es/~fsancho/?e=72>>

Friedrich. ¿Qué es la validación cruzada en el aprendizaje automático? [en línea]. s.l: Friedrich [Citado el 10 mayo, 2018]. Disponible en internet: <<https://www.quora.com/What-is-cross-validation-in-machine-learning>>

GARCÍA, S., RA-MÍREZ-GALLEGO, S., LUENGO, J., HERRERA, F. y RAMÍREZ-GALLEGO, S. Big Data monografía Big Data: Preprocesamiento y calidad de datos. [en línea]. s.l: GARCÍA, S., RA-MÍREZ-GALLEGO, S., LUENGO, J., HERRERA, F. y RAMÍREZ-GALLEGO, S. [Citado el 4 mayo 2018]. Disponible en internet: <[http://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133\\_Nv237-Digital-sramirez.pdf](http://sci2s.ugr.es/sites/default/files/ficherosPublicaciones/2133_Nv237-Digital-sramirez.pdf)> p. 2.

GIL, L.J., CASTILLO, L.F. y FLÓREZ, R.D. Reconocimiento de comandos de voz en español orientado al control de una silla de ruedas. UIS Ingenierías, Revista de la facultad de ingeniería físico mecánicas [en línea]. S.L: GIL, L.J., CASTILLO, L.F. y FLÓREZ, R.D. Disponible en internet: <<http://search.ebscohost.com/login.aspx?direct=true&db=fua&AN=121143584&lang=es&site=ehost-live>> p. 10.

Google Play. Hablando con Julis. [en línea]. s.l: Hablando con Julis. [Citado el 30 abril, 2018]. Disponible en internet: <<https://play.google.com/store/apps/details?id=io.cordova.julistalkes&hl=es>>

GÓMEZ, J., SIMANCAS, J., ACOSTA, M., MELÉNDEZ, F. y VÉLEZ, J. Algoritmo de reconocimiento de comandos voz basado en técnicas no-lineales. [en línea]. S.L: GÓMEZ, J., SIMANCAS, J., ACOSTA, M., MELÉNDEZ, F. y VÉLEZ, J. Disponible en internet: <<http://repositorio.cuc.edu.co/xmlui/handle/11323/904>> p. 8,17.

INSOR. Contexto general de la población sorda en Colombia. [en línea]. Bogotá: INSOR. [Citado el 28 marzo, 2018]. Disponible en internet: <[http://www.insor.gov.co/observatorio/download/Infog\\_pan\\_sordos\\_Col\\_sept2016.pdf](http://www.insor.gov.co/observatorio/download/Infog_pan_sordos_Col_sept2016.pdf)>

----- . ¿Qué es la lengua de señas? Portal niños. [en línea]. s.l: INSOR. [Citado el 1 mayo, 2018]. Disponible en internet: <<http://insor.gov.co/ninos/que-es-la-lengua-de-senas/>>

IRCAM. Introducción - Codificación predictiva lineal. [en línea]. s.l: IRCAM [Citado el 10 mayo, 2018]. Disponible en internet: <<http://support.ircam.fr/docs/AudioSculpt/3.0/co/LPC.html>>

Ingsistemastelesup. Validación cruzada. [en línea]. s.l: Ingsistemastelesup. [Citado el 10 mayo, 2018]. Disponible en internet: <<https://ingsistemastelesup.files.wordpress.com/2017/03/validacion-cruzada.pdf>> p. 1,2.

JANA ÁLVAREZ. Machine Learning y Support Vector Machines Analítica web. [en línea]. s.l: JANA ÁLVAREZ. [Citado el 12 mayo, 2018]. Disponible en internet: <<http://www.analiticaweb.es/machine-learning-y-support-vector-machines-porque-el-tiempo-es-dinero-2/>>

JASON BROWNLEE. Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning. [en línea]. s.l: JASON BROWNLEE [Citado el 12 mayo, 2018]. Disponible en internet: <<https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/>>

JOSÉ MUJICA. Transformada de Fourier. [en línea]. s.l: JOSÉ MUJICA. [Citado el 4 mayo, 2018]. Disponible en internet: <<http://www.escuelasuperiordeaudio.com.ve/articles/fourier-discrete.html>>

KLAYLAT, S., OSMAN, Z., HAMANDI, L. y ZANTOUT, R. Enhancement of an Arabic Speech Emotion Recognition System. International Journal of Applied Engineering Research. 2018. p. 4,7.

KOO PING SHUNG. Accuracy Precision, Recall or F1? – Towards Data Science. [en línea]. S.L: KOO PING SHUNG. [Citado el 17 mayo, 2018]. Disponible en internet: <<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>>

LLORENTE, C.R., ROBERTO, D., CHICOTE, B., JUAN, D., MONTERO MARTÍNEZ, M., JAVIER, D., GUARASA, M., RUBÉN, D., SEGUNDO, S., SECRETARIO, H., JUAN, D, MONTERO, M, SUPLENTE, M., FERNÁNDEZ, D.F. y CALIFICACIÓN, M. PROYECTO FIN DE CARRERA. [en línea]. S.L: LLORENTE, C.R., ROBERTO, D., CHICOTE, B., JUAN, D., MONTERO MARTÍNEZ, M., JAVIER, D., GUARASA, M., RUBÉN, D., SEGUNDO, S., SECRETARIO, H., JUAN, D, MONTERO, M, SUPLENTE, M., FERNÁNDEZ, D.F. y CALIFICACIÓN, M. PROYECTO FIN DE CARRERA. [Citado el 10 mayo, 2018]. Disponible en internet: <<http://lorien.die.upm.es/barra/pfcs/2007-carmenr/docs/proyecto.pdf>> p. 77.

MANUEL BENÍTEZ. emtic – Hetah traductor al lenguaje de signos y a Braille. [en línea]. s.l: MANUEL BENÍTEZ. [Citado el 30 abril, 2018]. Disponible en internet: <<https://enmarchaconlastic.educarex.es/244-emic/herramientas-2-0/1296-hetah-traductor-al-lenguaje-de-signos-y-a-braille>>

Maria Mercedes Gomez. ¿Sabes qué es una Machine Learning? - Comunidad e-Learning. [en línea], 2017. [Citado el 3 mayo, 2018]. Disponible en internet:

<<http://elearningmasters.galileo.edu/2017/09/21/sabes-que-es-una-machine-learning/>>

Mel Frequency Cepstral Coefficient (MFCC) tutorial. 2012 [en línea], [sin fecha]. [Consulta: 30 octubre 2018]. Disponible en: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>.

Mriquestions. Transformada de Fourier (FT): preguntas y respuestas en MRI. [en línea]. s.l: Mriquestions. [Citado el 4 mayo, 2018]. Disponible en internet: <<http://mriquestions.com/fourier-transform-ft.html>>

NAVARRO, B.G., [2015]. Trabajo de Fin de Grado Implementación de Técnicas de Deep Learning Implementation of Deep Learning Techniques. [en línea]. S.l.: [Consulta: 30 octubre 2018]. Disponible en: [https://riull.ull.es/xmlui/bitstream/handle/915/1409/Implementacion de Tecnicas de Deep Learning.pdf?sequence=1](https://riull.ull.es/xmlui/bitstream/handle/915/1409/Implementacion%20de%20Tecnicas%20de%20Deep%20Learning.pdf?sequence=1).

NICOLÁS SÁNCHEZ ANZOLA. Vista de Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD/COP spot intradiario ODEON. [en línea]. s.l: NICOLÁS SÁNCHEZ ANZOLA. [Citado el 12 mayo, 2018]. Disponible en internet: <<https://revistas.uexternado.edu.co/index.php/odeon/article/view/4414/5256>>

Noticias RCN. Colombia y el reto que tiene pendiente con los sordos del país. [en línea]. Bogotá: Noticias RCN. [Citado el 28 marzo, 2018]. Disponible en internet: <<https://www.noticiasrcn.com/nacional-pais/colombia-y-el-reto-tiene-pendiente-los-sordos-del-pais>>

PELEG, N. Linear Prediction Coding. [en línea]. s.l: [Citado el 10 mayo, 2018]. Disponible en internet: <<http://cs.haifa.ac.il/~nimrod/Compression/Speech/S4LinearPredictionCoding2009.pdf>> p. 7.

RENUKA JOSHI. Precisión recuperación y puntaje de F1: interpretación de las medidas de rendimiento - Exsilio Blog. [en línea]. S.L: RENUKA JOSHI. [Citado el 17 mayo, 2018]. Disponible en internet: <<http://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>>

Ruye Wang. Discrete Cosine Transform. [en línea]. s.l: Ruye Wang. [Citado el 9 mayo, 2018]. Disponible en internet: <[http://fourier.eng.hmc.edu/e101/lectures/Image\\_Processing/node13.html](http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node13.html)>

SIGNAL PROCESSING, Q., 2018. Why is each window/frame overlapping? 2016-12-28 [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://dsp.stackexchange.com/questions/36509/why-is-each-window-frame-overlapping>.

-----, Q.,2018. Why is each window/frame overlapping? 2016-12-28 [en línea]. [Consulta: 27 octubre 2018]. Disponible en: <https://dsp.stackexchange.com/questions/36509/why-is-each-window-frame-overlapping>.

SISTEMAS.COM, 2016. Definición de Hertz - Significado y definición de Hertz. [en línea]. [Citado el 1 mayo, 2018]. Disponible en: <https://sistemas.com/hertz.php>.

Slide Player. La Transformada Rápida de Fourier. [en línea]. s.l: Slide Player [Citado el 8 mayo, 2018]. Disponible en internet: <http://slideplayer.es/slide/1715431/>

Speech processing - The origin of constants in mel-scale formula - Signal Processing Stack Exchange. 2018 [en línea], [sin fecha]. [Consulta: 30 octubre 2018]. Disponible en: <https://dsp.stackexchange.com/questions/46209/the-origin-of-constants-in-mel-scale-formula>.

TECHcetera. ¿Qué es el Test de Turing (y, Google: qué has hecho)? [en línea]. s.l: TECHcetera [Citado el 2 mayo, 2018]. Disponible en internet: <http://techcetera.co/que-es-el-test-de-turing/>

Voz y Señas - Traductor LSM. [en línea]. s.l: Voz y Señas - Traductor LSM. [Citado el 28 marzo, 2018]. Disponible en internet: <http://www.vozysenas.com/>.

## 1.16. ANEXOS

### **Anexo A: Conjunto de datos.**

Se entrega el conjunto de datos de pronunciación de vocales en español en intervalos de tiempo de 1, 2 y 4 segundos, se grabaron 1675 audios para la construcción del conjunto de datos.

URL: [https://drive.google.com/drive/folders/1VYJSuZpsOirBqhXyjoRvLI\\_MvrBY0pJZ?usp=sharing](https://drive.google.com/drive/folders/1VYJSuZpsOirBqhXyjoRvLI_MvrBY0pJZ?usp=sharing)

### **Anexo B: Repositorio.**

Se entrega el código utilizado del experimento realizado, para fines educativos y investigativos, utilizando el Reconocimiento 4.0 Internacional (CC BY 4.0).

URL: <https://github.com/darubiano/SPEECH-RECOGNITION-OF-VOWELS>

### **Anexo C: Artículo científico.**

Se entrega un artículo científico, donde se expone el experimento de forma resumida y concreta realizado durante los 6 meses del trabajo de grado.

URL: <https://www.overleaf.com/read/rbvtkxztvdqy>