



**UJAT**

UNIVERSIDAD JUÁREZ  
AUTÓNOMA DE TABASCO

“ESTUDIO EN LA DUDA. ACCIÓN EN LA FE”



**DACYTI**

División Académica  
de Ciencias y Tecnologías  
de la Información

# Universidad Juárez Autónoma de Tabasco

Tesis Doctoral

## Método para Detectar Violencia Verbal Mediante la Combinación de Reconocimiento de Emociones y Análisis de Palabras

Que presenta

**Obed Campos Solano**

Para obtener el grado de

**Doctor en Ciencias de la Computación**

Directores

**Dr. Pablo Pancardo García**

**Dr. José Adán Hernández Nolasco**

*Cunduacán, Tabasco, México*

*Julio 2022*



UNIVERSIDAD JUÁREZ  
AUTÓNOMA DE TABASCO

“ESTUDIO EN LA DUDA. ACCIÓN EN LA FE”



# UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

## DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



Cunduacán, Tabasco, a 05 de julio de 2022.

En la Universidad Juárez Autónoma de Tabasco, de acuerdo al Reglamento de Estudios de Posgrado vigente, se revisó el trabajo de investigación titulado "**Método para Detectar Violencia Verbal Mediante la Combinación de Reconocimiento de Emociones y Análisis de Palabras**", realizado por el **C. Obed Campos Solano**, estudiante del Doctorado en Ciencias de la Computación, para obtener el grado de Doctor en Ciencias de la Computación bajo la modalidad de Tesis.

Los integrantes del jurado, después de revisar el trabajo, y en virtud de que se han atendido satisfactoriamente las observaciones y recomendaciones, otorgamos nuestra aprobación para que continúe los trámites correspondientes a la obtención del grado

**Dr. Miguel Antonio Wister Ovando**  
Profesor Investigador

**Dr. José Hernández Torruco**  
Profesor Investigador

**Dr. Pablo Payró Campos**  
Profesor Investigador

**Dra. Cristina López Ramírez**  
Profesor Investigador

**Dr. Óscar Alberto Chávez Bosquez**  
Profesor Investigador

C.c.p. Dr. Eddy Arquímedes Gracia Alcocer. Encargado del despacho de la Coordinación de Posgrado.

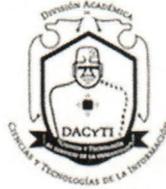
Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.  
Cunduacán, Tabasco, México.  
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870  
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx



**UNIVERSIDAD JUÁREZ  
AUTÓNOMA DE TABASCO**

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



**DIVISIÓN ACADÉMICA DE  
CIENCIAS Y TECNOLOGÍAS  
DE LA INFORMACIÓN**



2022 **Ricardo Flores Magón**  
Año de Magón  
PRECURSOR DE LA REVOLUCIÓN MEXICANA

"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 06 de julio de 2022

Oficio No. 0888/DACYTI/CP/2022

Asunto: Autorización de impresión de Tesis

**C. Obed Campos Solano**  
**Matricula: 191H18002**

En virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudios de Posgrado vigente en la Universidad, informo a Usted que se autoriza la impresión del trabajo recepcional "**Método para Detectar Violencia Verbal Mediante la Combinación de Reconocimiento de Emociones y Análisis de Palabras**", para presentar examen y obtener el Grado de Doctor en Ciencias de la Computación.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

**Atentamente**

**MTE. Oscar Alberto González González**  
**Director**



C.c.p. Dr. Eddy Arquímedes García Alcocer. - Encargado del Despacho de la Coordinación de Posgrado DACYTI  
Archivo.  
Consecutivo.  
MTE/OAGG/EAGA

*x*

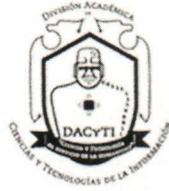
Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.  
Cunduacán, Tabasco, México.  
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870  
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx



UNIVERSIDAD JUÁREZ  
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE  
CIENCIAS Y TECNOLOGÍAS  
DE LA INFORMACIÓN



"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 06 de julio de 2022

Oficio No. 0888/DACYTI/CP/2022

Asunto: Autorización de Modalidad de examen

**C. Obed Campos Solano**

Doctorado en Ciencias de la Computación

En atención al oficio de fecha 05 de julio de 2022, en el cual solicita la autorización para titularse bajo la modalidad de Tesis, me permito informarle que se **AUTORIZA** el examen de grado, en virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudio de Posgrado vigente en la Universidad y los establecidos por el plan de estudios del Doctorado en Ciencias de la Computación.

Sin otro particular, aprovecho la ocasión para saludarle cordialmente.

Atentamente

**MTE. Oscar Alberto González González**

Director

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE  
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

C.c. p. Dr. Eddy Arquímedes García Alcocer. – Encargado del despacho de la Coordinación de Posgrado.  
Archivo.  
Consecutivo.

M.T.E. OAGG/EAGA

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.  
Cunduacán, Tabasco, México.  
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870  
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx

## Cesión de Derechos

Cunduacán, Tabasco. A 05 de julio de 2022

A quien corresponda

Los abajo firmantes, declaramos que el trabajo de tesis, titulado “**MÉTODO PARA DETECTAR VIOLENCIA VERBAL MEDIANTE LA COMBINACIÓN DE RECONOCIMIENTO DE EMOCIONES Y ANÁLISIS DE PALABRAS**” es de nuestra autoría intelectual y por lo tanto cedemos los derechos de comunicación pública, reproducción, distribución, difusión en general y puesta a disposición electrónica de la citada tesis doctoral, de forma gratuita y no exclusiva, a la Universidad Juárez Autónoma de Tabasco, a la cual relevamos de cualquier sanción y asumimos responder a cualquier reclamo de derechos de autor ante las autoridades competentes.

Atentamente

Autores

<b>M.T.I. Obed Campos Solano</b>	Calle Dr. Jaime Martínez Madrid 50. Col. Ejido Cunduacán. Cunduacán, Tabasco. CP. 86690	
<b>Dr. Pablo Pancardo García</b>	Universidad Juárez Autónoma de Tabasco. División Académica de Ciencias y Tecnologías de la Información. Carretera Cunduacán - Jalpa de Méndez Km. 1. Cunduacán, Tabasco. CP. 86690	
<b>Dr. José Adán Hernández Nolasco</b>	Universidad Juárez Autónoma de Tabasco. División Académica de Ciencias y Tecnologías de la Información. Carretera Cunduacán - Jalpa de Méndez Km. 1. Cunduacán, Tabasco. CP. 86690	

## CARTA DE AUTORIZACIÓN

El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la Tesis de grado denominada **“Método Para Detectar Violencia Verbal Mediante La Combinación De Reconocimiento De Emociones Y Análisis De Palabras”**, de la cual soy autor y titular de los Derechos de Autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa mas no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra Red Académica con las que Universidad tenga relación institucional.

Por lo antes mencionado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la Tesis mencionada y para los fines estipulados en este documento.

Se firma la presente autorización en la Ciudad de Cunduacán, Tabasco a los 5 días del mes de julio del año 2022

**AUTORIZO**



**OBED CAMPOS SOLANO**



**UJAT**

UNIVERSIDAD JUÁREZ  
AUTÓNOMA DE TABASCO

“ESTUDIO EN LA DUDA. ACCIÓN EN LA FE”



**DAGYTI**

División Académica  
de Ciencias y Tecnologías  
de la Información

# Universidad Juárez Autónoma de Tabasco

Tesis Doctoral

## Método para Detectar Violencia Verbal Mediante la Combinación de Reconocimiento de Emociones y Análisis de Palabras

Que presenta

**Obed Campos Solano**

Para obtener el grado de

**Doctor en Ciencias de la Computación**

Comité Tutorial: **Dr. Pablo Pancardo García**  
**Dr. José Adán Hernández Nolasco**  
**Dra. Betania Hernández Ocaña**  
**Dra. Juana Canul Reich**

Jurado: **Dr. Miguel Antonio Wister Ovando** Presidente  
**Dr. José Hernández Torruco** Secretario  
**Dr. Pablo Payró Campos** Vocal  
**Dra. Cristina López Ramírez** Vocal  
**Dr. Óscar Alberto Chávez Bosquez** Vocal

*Cunduacán, Tabasco, México*

*Julio 2022*



UNIVERSIDAD JUÁREZ  
AUTÓNOMA DE TABASCO

“ESTUDIO EN LA DUDA. ACCIÓN EN LA FE”

# Agradecimientos

A Dios primeramente; porque es el creador de todo y sin su ayuda no hubiera podido salir adelante.

A mi esposa Kassandra, que estuvo conmigo en cada momento de ésta aventura; y fué quien me motivó a dar inicio a pesar de todo.

A mis hijos Carlos Daniel y Salomón Alejandro, ya que ellos fueron el motor que inició éste proyecto; y por ellos decidí permanecer en el programa.

A mis padres Carlos y Rocío, quienes me apoyaron en todo momento y me brindaron ayuda económica cuando fue necesario.

A mi familia, quienes siempre han estado allí en cualquier circunstancia.

A mis asesores, Pablo y Adán; quienes me guiaron desde el inicio e incluso me ayudaron en muchas ocasiones, entregando el extra y demostraron que la docencia más que un arte es una pasión.

# Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Antecedentes . . . . .	5
1.3. Planteamiento del Problema . . . . .	8
1.4. Hipótesis y preguntas de investigación . . . . .	11
1.4.1. Hipótesis . . . . .	11
1.4.2. Pregunta de investigación . . . . .	11
1.5. Delimitación . . . . .	12
1.6. Objetivos . . . . .	13
1.6.1. Objetivo General . . . . .	13
1.6.2. Objetivos específicos . . . . .	13
1.7. Justificación . . . . .	14
<b>2. Marco Teórico y Estado del Arte</b>	<b>17</b>
2.1. Estado del Arte . . . . .	17
2.1.1. Marco Conceptual . . . . .	17
2.1.2. Computación Afectiva . . . . .	17
2.1.3. Inteligencia Artificial . . . . .	18
2.1.4. Lenguaje Natural . . . . .	23
2.1.5. Lógica Difusa . . . . .	24
2.2. Marco Tecnológico . . . . .	26
2.3. Marco Referencial . . . . .	28

---

2.4. Marco Legal . . . . .	30
<b>3. Modelo Dynamic Fuzzy Logic Propuesto</b>	<b>33</b>
3.1. Adquisición de Datos . . . . .	33
3.1.1. Procesamiento de los Datos . . . . .	36
3.1.2. Dataset . . . . .	37
3.2. Descripción del Modelo . . . . .	40
3.2.1. Preparación del Entorno . . . . .	40
3.2.2. Variables de Entrada . . . . .	41
3.2.3. Variables de Contexto . . . . .	48
3.2.4. Motor de Inferencia . . . . .	54
3.2.5. Variables de Salida . . . . .	56
<b>4. Pruebas y Resultados</b>	<b>61</b>
4.1. Pruebas . . . . .	61
4.2. Resultados . . . . .	65
<b>5. Conclusión</b>	<b>69</b>
Referencias . . . . .	72

# Lista de figuras

1.1. Estatura usando lógica convencional . . . . .	3
1.2. Estatura usando lógica difusa . . . . .	4
1.3. Botón de Pánico en semáforos . . . . .	8
2.1. Proceso de aprendizaje automático . . . . .	20
2.2. Perceptrón simple . . . . .	22
2.3. Ejemplo de una máquina de soporte vectorial . . . . .	22
2.4. Reglas en un modelo difuso . . . . .	25
2.5. Regla de gol en fútbol . . . . .	26
3.1. Item transaction . . . . .	38
3.2. Asalto combi . . . . .	39
3.3. Vector de datos . . . . .	39
3.4. Variable Palabras . . . . .	42
3.5. Parametros vector variable . . . . .	43
3.6. Parametros vector amplitud . . . . .	44
3.7. vector amplitud . . . . .	44
3.8. niveles amplitud . . . . .	45
3.9. Función de amplitud por medio de consola en MATLAB . . . . .	46
3.10. Ejemplo de contexto . . . . .	49
3.11. Ejemplo de contexto (Decisión tomada) . . . . .	50
3.12. Gráfica de incidencia delictiva 2018-2021 . . . . .	51
3.13. Mapa de zonas más inseguras CDMX . . . . .	53

---

3.14. Ejemplo de cambio de peso en las reglas . . . . .	55
3.15. Salida del Modelo . . . . .	56
3.16. Grupos de la variable salida . . . . .	57
3.17. Valor de salida con tres grupos . . . . .	58
3.18. Valor de salida con cinco grupos . . . . .	59
4.1. Nombre de archivo en dataset . . . . .	62
4.2. Ejemplo de escala NOVACO . . . . .	63
4.3. Vector construido a partir de los datos (color amarillo) . . . . .	64
4.4. Muestra de modelo entrenado a partir de vectores . . . . .	65
5.1. Proyecto general en el cual se está trabajando . . . . .	70

# Capítulo 1

## Introducción

### 1.1. Introducción

En muchas regiones del mundo, la sociedad actual depende completamente de la tecnología y el uso de dispositivos inteligentes como smartphones, Smart watches y otros, para realizar las actividades cotidianas de manera más fácil y rápida. Los asistentes virtuales han cobrado popularidad en éstos últimos años, permitiendo que las personas le hablen directamente a sus dispositivos y por medio de comandos de voz, ordenen al asistente (por medio de uso de algoritmos de inteligencia artificial que traduce la voz a señal digital) las tareas a realizar y se ejecuten automáticamente [19].

Gracias a la creciente demanda de parte de los usuarios por innovaciones en las tecnologías actuales, las grandes compañías líderes en el mercado de desarrollo de aplicaciones y sistemas operativos (Google, Apple, Samsung, etc.) han desarrollado aplicaciones y sistemas mucho más complejos que hacen uso de inteligencia artificial, cuyos algoritmos emplean el reconocimiento de patrones y realizan predicciones en tiempo real de las tareas y actividades que realiza el usuario. Un ejemplo es Google Assistant [9]; el asistente virtual capaz de realizar tareas simples como agendar citas, hacer llamadas, enviar mensajes o reproducir contenido multimedia. Lo único que necesita para operar es recibir comandos de voz por parte de los usuarios.

En éstos últimos años, los algoritmos de inteligencia artificial, máquinas de aprendizaje (*machine learning*), aprendizaje profundo (*Deep learning*) y otras técnicas de aprendizaje supervisado, han tenido un auge y desarrollo que ha ido creciendo a pasos agigantados, propiciando desarrollar tecnologías que hace algunos años el ser humano no imaginaba. Es por esto que muchos de éstos algoritmos emplean el uso de sensores para desempeñar el propósito por el cuál fueron creados. Los dispositivos inteligentes (*Smartphones*) que cargamos en nuestros bolsillos todos los días, están equipados con sensores que tienen diversas funciones. Dos de éstos sensores son muy útiles y con el enfoque adecuado se pueden realizar grandes tareas que pueden ayudar a las personas a comunicarse de manera más efectiva y sobre todo en tiempo real [38]. El micrófono y GPS, son sensores que son de gran utilidad y se pueden realizar tareas mucho más efectivas ya que mediante la ubicación y los comandos de voz, el usuario puede comunicarse con el dispositivo sin necesidad de sacarlo del bolsillo.

Otra tecnología que forma parte de la inteligencia artificial, aunque menos conocida por algunos, es la lógica difusa [54]. Ésta rama de la inteligencia artificial nos permite hacer uso de la lógica convencional de una manera más precisa. Por ejemplo, si usamos lógica convencional para determinar si la estatura de una persona es alta o baja, tendríamos únicamente dos valores.

Supongamos que para fines del ejemplo, la estatura promedio es 1.70m. Esto significa que una persona cuya estatura es 1.70m, está dentro del rango **normal**, mientras que una persona que mide 1.65m es una persona con estatura **baja**. Lo mismo pasa con una persona cuya estatura es 1.85m sería considerada **alta** (Ver Figura 1.1).

Pero dentro de la lógica difusa, los valores forman parte de ciertos grupos de pertenencia. Es decir, una persona con estatura normal, bien podría ir perteneciendo al grupo de las personas altas y en menor probabilidad al grupo de personas bajas.

La lógica difusa funciona con probabilidades. Esto significa que cierto valor pertenece muy

probablemente a cierto grupo y es poco probable que pertenezca a otro grupo. Otro ejemplo que podemos tomar es medir cierta temperatura. Mientras que la lógica convencional dicta que un valor mayor a  $36.5\text{ }^{\circ}\text{C}$  significa que una persona tiene fiebre, y un valor igual a  $36\text{ }^{\circ}\text{C}$  dicta que la persona tiene temperatura normal o promedio, la lógica difusa diría que, a manera de suposición, el valor de  $36.5\text{ }^{\circ}\text{C}$  significaría que hay un 80% de probabilidad que la persona tenga fiebre y un 20% que la persona tenga temperatura promedio.

Veamos una ilustración para que el concepto quede un poco más claro (Ver Figura 1.2):

Si tomamos el ejemplo de la estatura (Ver Figura 1.1), vemos que en la lógica difusa, los valores pertenecen en parte a cierto grupo. Éstos grupos son definidos por el usuario durante el diseño del modelo, de lo cuál hablaremos más adelante. Volviendo al ejemplo, observemos que la estatura de una persona no es absoluta de un grupo, es decir, la persona no es alta ni baja; sino es muy probable que la persona sea alta y poco probable que sea baja, o viceversa.

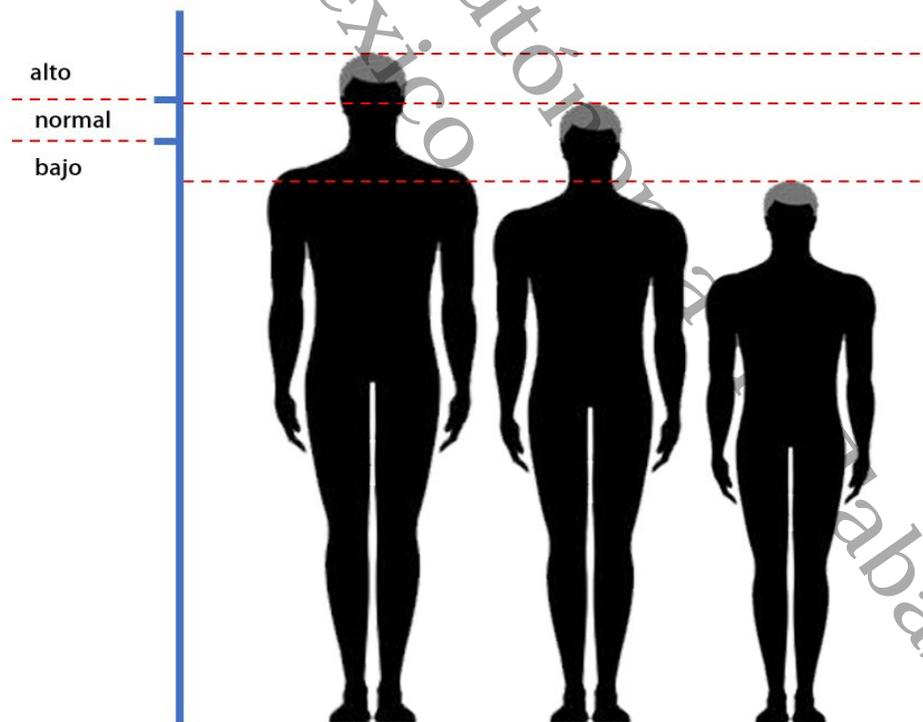


Figura 1.1: Estatura usando lógica convencional

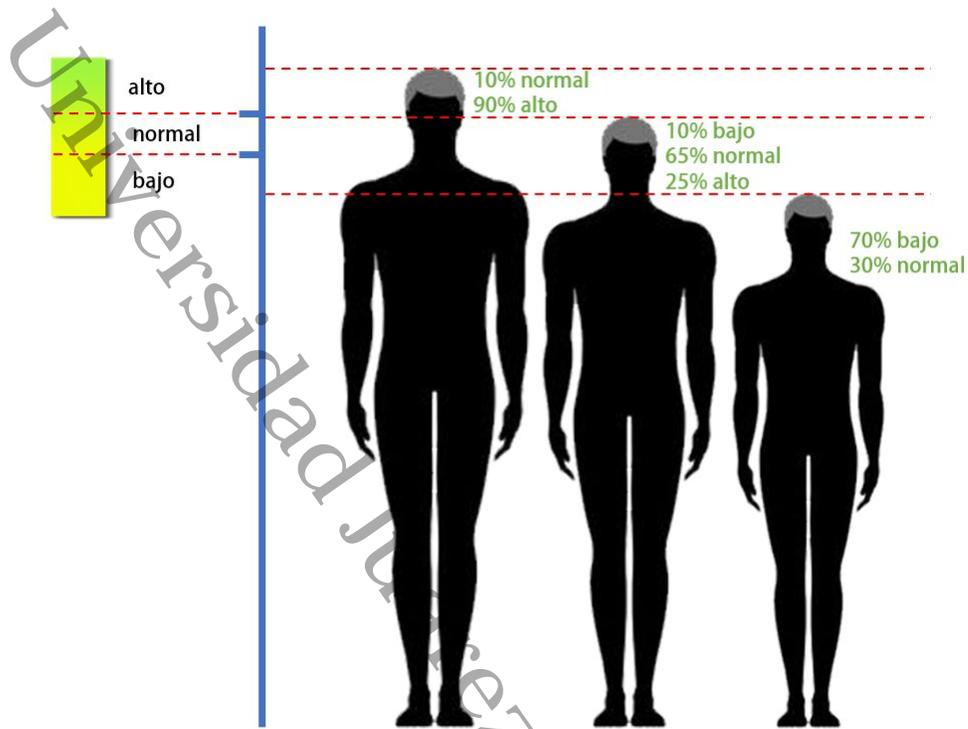


Figura 1.2: Estatura usando lógica difusa

Esto es solo una introducción al tema, mismo que se irá desarrollando en el transcurso del documento. Por lo pronto la idea está plasmada, para ser entendida de mejor manera más adelante.

## 1.2. Antecedentes

En la actualidad, la sociedad en la que vivimos está inmersa en un mundo tecnológico. Con la llegada del internet de las cosas (IoT), las compañías de telecomunicaciones desarrollan dispositivos más complejos; e integran a sus plataformas algoritmos de inteligencia artificial que van avanzando con el paso del tiempo. Llevamos en nuestros bolsillos, comunicación accesible desde cualquier parte del mundo gracias a los múltiples sensores integrados en dichos dispositivos, que hacen su uso más sencillo.

Éstos avances, nos permiten aportar conocimiento al desarrollo de nuevas tecnologías para ayudar a combatir una problemática presente en todo el mundo: el robo con violencia. Éste es uno de los muchos tipos de violencia que existen en México; los casos de robos o asaltos en la calle o transporte público por lo general presentan violencia verbal en las palabras [33] por parte del asaltante o agresor. Una forma de combatir éste tipo de robos, es identificar la violencia verbal en las palabras, empleando algoritmos de Inteligencia Artificial de tal manera que se puedan tomar las acciones necesarias para socorrer a las víctimas.

Éstos algoritmos mencionados anteriormente, como son: máquinas de soporte vectorial (*Support Vector Machines*), redes neuronales (*Neural Networks*) o técnicas de aprendizaje profundo (*Deep Learning*), han sido utilizados para detectar características presentes en la voz. Para identificar éstas características, se han hecho estudios para detectar agresiones verbales y así ayudar en la detección efectiva del crimen de una forma autónoma; es decir, que empleando estos métodos o algoritmos, el sistema sea capaz de autenticar la violencia sin que la víctima tenga que iniciarlo [7] [60].

Otro antecedente que se tiene en ésta investigación, es emplear computación afectiva (una rama de la inteligencia artificial) para analizar e identificar emociones tanto positivas (felicidad, neutra) como negativas (enojo, tristeza) [14]. Uno de los objetivos para el cuál ha sido utilizada ésta rama, es principalmente para identificar correctamente una emoción negativa

(en éste caso, como se mencionó anteriormente *ira o enojo*); ya que hay emociones que se pueden confundir, como es el caso de felicidad con enojo [13]. Para detectar éstas emociones, existen en el estado del arte trabajos que se han elaborado, y que toman en cuenta algunas de ellas incluyendo las negativas [46] [52] [39].

En [8] se analiza el audio mediante el uso de dos bases de datos, una de ellas contiene audios con contenido violento real y la otra contiene audios extraídos de escenas de películas. Todas las muestras tienen una duración de 5 segundos.

Otra investigación realizada por Bugueño Sáez [12], consiste en detectar agresiones verbales en archivos de audio extraídos de escenas de películas utilizando el modelo de Carletti y la herramienta OpenSMILE, así como las técnicas Bolsa de Palabras (BoW) y Máquinas de Soporte Vectorial (SVM); logrando una detección del 86 %.

Se han realizado estudios para detectar el disparo de armas de fuego, gritos, ruptura de vidrios, etc. [61][66][65][32][31]. Pero este tipo de sonidos son fáciles de identificar ya que su pitch es muy alto y sobresale entre los demás niveles.

En Londres, se hicieron pruebas en las unidades de transporte público [58] para detectar asaltos. Cabe mencionar que estas pruebas se hicieron instalando cámaras de vigilancia y analizando las imágenes en la central de policía.

En Brasil, un equipo de investigadores se dio a la tarea de registrar mediante una aplicación móvil, el audio del propietario con la finalidad de detectar violencia verbal [60], esto se hizo para ayudar a las mujeres, ya que en ese país es muy común que la pareja (por lo regular el hombre) golpee y maltrate a la mujer.

Después de haber analizado el estado del arte existente en la actualidad, no encontramos ningún trabajo que se haya realizado para detectar violencia verbal usando lógica difusa y

tampoco aplicando dinamismo al modelo, como es el caso de la propuesta de este trabajo. Por lo que tenemos grandes expectativas acerca de la funcionalidad del modelo.

Universidad Juárez Autónoma de Tabasco.  
México.

### 1.3. Planteamiento del Problema

Existe un interés por parte de las autoridades a nivel nacional, de erradicar en la medida de lo posible, el índice de criminalidad que es reportado por ciudadanos e instituciones como el INEGI (Instituto Nacional de Estadística y Geografía). Los estudios realizados muestran que los índices de asaltos en nuestro país son preocupantemente altos, ya que, a Diciembre de 2018, los registros muestran que la suma total del rubro “ *robos o asaltos en la calle o transporte público* ”, es de 10,775 casos por cada cien mil habitantes [30]; lo que significa que el 10.77% de las incidencias delictivas en México están conformadas por asaltos.

Dicho lo anterior, cabe mencionar que durante un evento con actividad criminal, como en el caso de un robo con violencia, los atacantes usan palabras y lenguaje violento; así como emociones negativas (ira o enojo), para intimidar a sus víctimas; lo que hace de la delincuencia un problema mundial que los gobiernos han tratado de erradicar mediante diversas propuestas a lo largo de los años. Éstas propuestas, en muchas ocasiones han incluido soluciones tecnológicas, que son creativas, e incluso inteligentes; como se demostró al instalar botones de pánico en semáforos [27] o cámaras de seguridad con audio en unidades del servicio público en municipios del Estado de México [42].



Figura 1.3: Botón de pánico instalado en semáforo de la ciudad de Medellín, Colombia

Pero lamentablemente éstas medidas implementadas carecen de efectividad, ya que las víctimas necesitan estar en la posibilidad de accionar dichos botones o dispositivos.

Sin embargo, existen trabajos de investigación que se han realizado para ayudar a solucionar éste problema en diversas partes del mundo, los cuales han intentado dar una solución más efectiva a éste problema [56]. En México, se han implementado varias medidas de seguridad para combatir dicho problema (como botones de pánico y aplicaciones móviles de seguridad personal); incluso se ha recopilado información de estadísticas delictivas para construir un sistema de información que sea fiable y contenga la mayor información posible [3]. Pero lamentablemente, las soluciones que existen actualmente para generar una alerta en una situación de violencia incrementan la probabilidad de que la víctima pierda la vida, dada la peligrosidad del escenario. Ésto es por que el uso de dichas soluciones, al ser detectadas por el agresor, incrementan el nivel de violencia del mismo; pudiendo provocar heridas graves, e incluso fatales en las víctimas [60]. Por tal motivo, es necesaria una propuesta que pueda reconocer la violencia verbal sin que el atacante se percate.

Debido a la efectividad de este botón, los delincuentes están al tanto de su existencia y amenazan al conductor para impedir que éste lo presione. Esto es un gran problema, ya que la efectividad ha disminuido con el paso del tiempo y además, algunos han dejado de funcionar debido a fallas físicas o electrónicas.

Cabe mencionar, que el material audiovisual no está a disposición del público en general, ya que es propiedad del C5 antes mencionado. El material al que se tiene acceso es aquel que se ha compartido mediante redes sociales como Facebook, Twitter, Instagram, YouTube y sitios web de periódicos populares en el país. Por tal motivo, no existe una amplia variedad de videos a los cuáles hacer pruebas y generar una base de datos (o Dataset) extensa.

Otra problemática es el reconocimiento de emociones para asegurar que la víctima se encuentra en una situación de violencia verbal agresiva. Trabajos de investigación se han llevado a

cabo para detectar emociones presentes en la voz humana [18] [49]; Incluso se han desarrollado herramientas web para analizar clips de audio y detectar emociones [23] [22], pero carecen de una detección efectiva en emociones negativas.

El problema por lo tanto, es que actualmente no existe una forma efectiva en la que se pueda detectar una situación de violencia verbal contenida en las palabras, que considere al mismo tiempo la emoción presente. Diversos autores han dado soluciones parciales a éste problema [26] [44] [50]; pero los resultados reportados no son muy acertados y la metodología empleada pudiera ser mejorada para obtener un nivel de efectividad mayor.

Lo anterior representa una oportunidad, ya que con ésta investigación se puede ayudar a dar solución a éste tipo de problemas y aportar conocimiento que ayude a fortalecer diversos aspectos que los autores antes mencionados presentan.

## 1.4. Hipótesis y preguntas de investigación

### 1.4.1. Hipótesis

el robo con violencia verbal se puede reconocer mediante un sistema difuso dinámico cuyas variables incluyan emociones y palabras agresivas

### 1.4.2. Pregunta de investigación

1. ¿Es el diseño de un sistema difuso dinámico una opción que da mejores resultados que los que existen actualmente para detectar violencia verbal?

Universidad Juárez Autónoma de Tabasco.  
México.

## 1.5. Delimitación

La presente investigación se limita a los siguientes enunciados:

- Se considerarán eventos en donde haya presencia de emociones negativas (*ira o enojo*) y palabras agresivas o frases identificadas como propias de una situación de robo con violencia. Las palabras comunes en una situación de violencia verbal no serán tomadas en consideración, así como emociones no negativas.
- Se identificarán eventos de violencia verbal que sean propios del léxico nacional mexicano. Las palabras que procedan de otras nacionalidades estarán descartadas.
- El método no será aplicable en procesos de prevención de la violencia, sólo su detección. Es decir, no se pretende predecir cuándo sucederá el evento; sino detectarlo en el menor tiempo posible, al momento que está sucediendo.
- Las fuentes de audio con niveles de ruido elevados no serán elegibles.

## 1.6. Objetivos

### 1.6.1. Objetivo General

Reconocer el robo con violencia verbal mediante algoritmos de aprendizaje automático que clasifiquen y combinen emociones y palabras agresivas

### 1.6.2. Objetivos específicos

1. Identificar las palabras que se usan cuando hay violencia.
2. Reconocer la emoción usando computación afectiva.
3. Comparar las palabras extraídas del audio con las palabras identificadas en violencia.
4. Clasificar y detectar la violencia verbal en escenarios de robo con violencia

## 1.7. Justificación

En éste contexto, cabe mencionar que el análisis de emociones no es nuevo, ya que existen trabajos de años anteriores cuyo objetivo principal es detectar la presencia de emociones en la voz[47]. De igual manera, la detección de eventos en el audio no es actual, ya que estudios demuestran que se han hecho análisis anteriormente para detectar emociones y características en la voz[51][41]. Y a pesar de que hay registros de esto, no existe un modelo o método que combine ambos factores para poder realizar una detección más efectiva de la violencia (al menos incluyendo los métodos propuestos).

Por lo tanto, la realización de éste proyecto es importante; ya que provee un método efectivo para detectar las agresiones verbales presentes por medio de las palabras y emociones presentes en un robo con violencia. Además, es una aportación actual y moderna, ya que se emplea el uso de algoritmos de inteligencia artificial para lograr un proceso automatizado, ubicuo y en tiempo real; de tal manera que la agresión se pueda detectar en el menor tiempo posible y tomar acciones supervisadas que serán programadas durante el entrenamiento del método.

La estructuración de éste método consiste en analizar el lenguaje natural en busca de dos variables principales: palabras agresivas, mismas que coincidan con el lenguaje violento utilizado por los asaltantes que practican el robo con violencia. La segunda variable se obtendrá mediante el análisis del lenguaje en busca de emociones negativas; en éste caso: ira o enojo. Una vez analizado éste lenguaje, se buscará la presencia de éstas dos variables y entonces, aplicar algoritmos de inteligencia artificial para determinar el porcentaje de violencia con la que cuenta ese lenguaje.

Por lo tanto, ésta investigación aporta conocimiento al área de las ciencias de la computación mediante una combinación de técnicas de inteligencia artificial para detectar en una forma más efectiva la violencia verbal agresiva, ayudando a proteger la integridad de los ciudadanos en la sociedades actuales; y es una contribución importante para aquellos que han

sido víctimas de la delincuencia que se vive en el país. Éste método antes mencionado, aporta una técnica novedosa en cuanto a la detección de robos con violencia, ya que de acuerdo con el estado del arte recopilado, no existen hasta el momento métodos que detecten dichos robos usando ésta combinación en específico; lo cuál es una oportunidad para implementar un nuevo algoritmo y el área de las ciencias de la computación se vea beneficiada gracias a ésta nueva técnica.

De igual manera, otro aporte de ésta investigación es que ayuda a los organismos de seguridad y autoridades encargadas de combatir la inseguridad, pues se propone una herramienta que ayude a combatir una problemática regional para disminuir lo niveles de robo con violencia y crimen en general.

Universidad Juárez Autónoma de Tabasco.  
México.

Universidad Juárez Autónoma de Tabasco.  
México.

## Capítulo 2

# Marco Teórico y Estado del Arte

### 2.1. Estado del Arte

#### 2.1.1. Marco Conceptual

Con el fin de lograr entender el propósito de ésta investigación, es necesario comprender algunos conceptos, mismos que conforman la base para poder realizar pruebas y experimentos, de tal manera que se pueda lograr un aporte preciso y más completo al área involucrada. En éste caso, las ciencias de la computación.

#### 2.1.2. Computación Afectiva

Las emociones juegan un rol importante en el rol de comunicación humano a humano. De hecho, la inteligencia emocional es más importante incluso que el valor de coeficiente intelectual para que se pueda llevar a cabo una interacción satisfactoria. La computación afectiva y el análisis de sentimientos son la clave para el avance de las áreas de la inteligencia artificial y todo lo que resulte de ella[14].

Las técnicas de minería de sentimientos que existen hoy en día pueden ser utilizadas para revelar quiénes son los clientes más importantes en un mercado de productos, mostrando así cuáles ofertas pueden generar una emoción en el comprador y aumentar la probabilidad

de que el cliente adquiriera el producto. Para llevar a cabo éstos resultados, la computación afectiva se basa en tres categorías principales: Técnicas basadas en conocimiento, métodos estadísticos y aproximación híbrida.

El uso de sistemas basados en conocimiento, aplicados en el análisis de emociones, permite explotar características implícitas asociadas con conceptos de lenguaje natural (por ejemplo, conocimiento del mundo real y el uso del sentido común) junto con la adición de información conceptual y afectiva a través del uso de redes semánticas. Éstas redes semánticas, analizan texto en busca de sentimientos y dicho texto es clasificado en categorías, teniendo como base la presencia de palabras sentimentales; tales como "felicidad", "tristeza", "temor", "aburrimiento", etc.

Existen sistemas diseñados para analizar y detectar éste tipo de emociones, como son: Affective Lexicon[48], linguistic annotation scheme[64], WordNet-Affect[5], SentiWordNet[21] y SenticNet[15] los cuáles pueden ser utilizados para obtener un análisis más exacto para explotar la detección de emociones.

### 2.1.3. Inteligencia Artificial

Coloquialmente, el término inteligencia artificial se aplica cuando una máquina imita las funciones cognitivas que los humanos asocian con otras mentes humanas, como por ejemplo: percibir, razonar, aprender y resolver problemas [59]

En 1980, Margaret Boden define el término inteligencia artificial como: "*the use of computer programs and programming techniques to cast light on the principles of intelligence in general and human thought in particular*" [11] que traducido y adaptado al español significa: "El uso de programas de computadora y técnicas de programación para dar luz a los principios de inteligencia en general y en particular al pensamiento humano". La definición apropiada de inteligencia artificial para tiempos modernos, a partir de la década de 2020 podría ser entonces:

*“Conjunto de técnicas y algoritmos computacionales cuyo objetivo es imitar el pensamiento humano para reconocer patrones y resolver problemas complejos”*

La inteligencia artificial ha tenido un auge estrepitoso en estos últimos años. Se han hecho avances en la ciencia gracias a sus técnicas y algoritmos, desde programas de cómputo (incluyendo aplicaciones móviles) hasta conducción vehicular autónoma [63].

Gracias a los avances en ésta rama de las ciencias de la computación, se han podido aportar soluciones parciales a problemas complejos que en épocas pasadas no se podían resolver sino en la mente humana. Los cálculos y procesos matemáticos que las computadoras pueden realizar sin duda superan los procesos cognitivos de la mente humana. Las máquinas son capaces de identificar patrones en forma precisa y aprender de ellos. Un ejemplo claro, son las redes neuronales, las cuáles son una técnica usada en inteligencia artificial y cuyo funcionamiento trata de imitar las neuronas del cerebro humano [35].

Con el reconocimiento de patrones, se pueden llevar a cabo tareas como distinguir las características presentes en la voz humana; éstas características presentan patrones propios los cuales pueden ser identificados y clasificados para determinar situaciones reales. Por ejemplo: cuando la persona está gritando, llorando, conversando, riendo, etc. Con éstas características se puede determinar si la voz de una persona presenta comportamiento agresivo o violento.

### **Aprendizaje Automático**

El aprendizaje automático o machine learning es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. Un agente o máquina aprende cuando su desempeño mejora con la experiencia; es decir, cuando la habilidad no estaba presente en su genotipo o rasgos de nacimiento [59]



Figura 2.1: Proceso de aprendizaje automático

Los diferentes algoritmos de Aprendizaje Automático se agrupan en una taxonomía en función de la salida de los mismos. Algunos tipos de algoritmos son:

- **Aprendizaje supervisado:** El algoritmo produce una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema. Un ejemplo de este tipo de algoritmo es el problema de clasificación, donde el sistema de aprendizaje trata de etiquetar (clasificar) una serie de vectores utilizando una entre varias categorías (clases). La base de conocimiento del sistema está formada por ejemplos de etiquetados anteriores. Este tipo de aprendizaje puede llegar a ser muy útil en problemas de investigación biológica, biología computacional y bioinformática [34].
- **Aprendizaje no supervisado:** Todo el proceso de modelado se lleva a cabo sobre un conjunto de ejemplos formado tan sólo por entradas al sistema. No se tiene información sobre las categorías de esos ejemplos. Por lo tanto, en este caso, el sistema tiene que ser capaz de reconocer patrones para poder etiquetar las nuevas entradas [6].
- **Aprendizaje semisupervisado:** Este tipo de algoritmos combinan los dos algoritmos anteriores para poder clasificar de manera adecuada. Se tiene en cuenta los datos marcados y los no marcados [17].
- **Aprendizaje por refuerzo:** El algoritmo aprende observando el mundo que le rodea. Su información de entrada es el feedback o retroalimentación que obtiene del mundo

exterior como respuesta a sus acciones. Por lo tanto, el sistema aprende a base de ensayo-error. El aprendizaje por refuerzo es el más general entre las tres categorías. En vez de que un instructor indique al agente qué hacer, el agente inteligente debe aprender cómo se comporta el entorno mediante recompensas (refuerzos) o castigos, derivados del éxito o del fracaso respectivamente. El objetivo principal es aprender la función de valor que le ayude al agente inteligente a maximizar la señal de recompensa y así optimizar sus políticas de modo a comprender el comportamiento del entorno y a tomar buenas decisiones para el logro de sus objetivos formales [62].

- **Aprendizaje multi-tarea:** Métodos de aprendizaje que usan conocimiento previamente aprendido por el sistema de cara a enfrentarse a problemas parecidos a los ya vistos [16].

Éstos algoritmos, utilizan técnicas de clasificación; algunas de ellas se describen a continuación:

- **Árboles de decisión:** Este tipo de aprendizaje usa un árbol de decisiones como modelo predictivo. Se mapean observaciones sobre un objeto con conclusiones sobre el valor final de dicho objeto. Los árboles son estructuras básicas en la informática. Los árboles de atributos son la base de las decisiones. Una de las dos formas principales de árboles de decisiones es la desarrollada por Quinlan de medir la impureza de la entropía en cada rama, algo que primero desarrolló en el algoritmo ID3 y luego en el C4.5. Otra de las estrategias se basa en el índice GINI y fue desarrollada por Breiman, Friedman et alia. El algoritmo de CART es una implementación de esta estrategia [57].
- **Reglas de asociación:** Los algoritmos de reglas de asociación procuran descubrir relaciones interesantes entre variables. Entre los métodos más conocidos se hallan el algoritmo a priori, el algoritmo Eclat y el algoritmo de Patrón Frecuente [2].
- **Redes neuronales artificiales:** Las redes de neuronas artificiales (RNA) son un paradigma de aprendizaje automático inspirado en las neuronas de los sistemas nerviosos de los animales. Se trata de un sistema de enlaces de neuronas que colaboran entre

sí para producir un estímulo de salida. Las conexiones tienen pesos numéricos que se adaptan según la experiencia. De esta manera, las redes neurales se adaptan a un impulso y son capaces de aprender. La importancia de las redes neurales cayó durante un tiempo con el desarrollo de los vectores de soporte y clasificadores lineales, pero volvió a surgir a finales de la década de 2000 con la llegada del aprendizaje profundo [35].

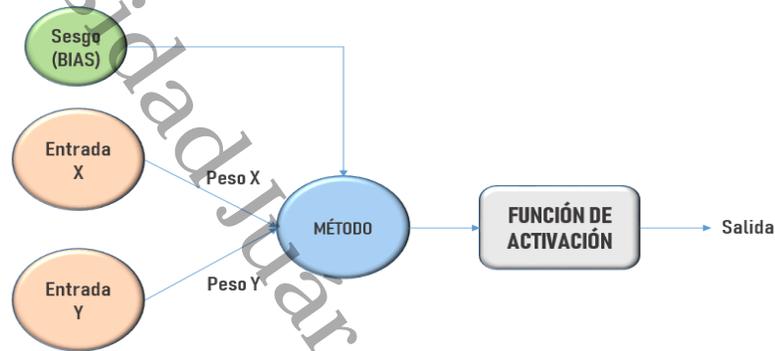


Figura 2.2: Perceptrón (el elemento más simple de una red neuronal)

- Máquinas de soporte vectorial:** Las SVM (*Support Vector Machine*) son una serie de métodos de aprendizaje supervisado usados para clasificación y regresión. Los algoritmos de MVS usan un conjunto de ejemplos de entrenamiento clasificado en dos categorías para construir un modelo que prediga si un nuevo ejemplo pertenece a una u otra de dichas categorías [40].

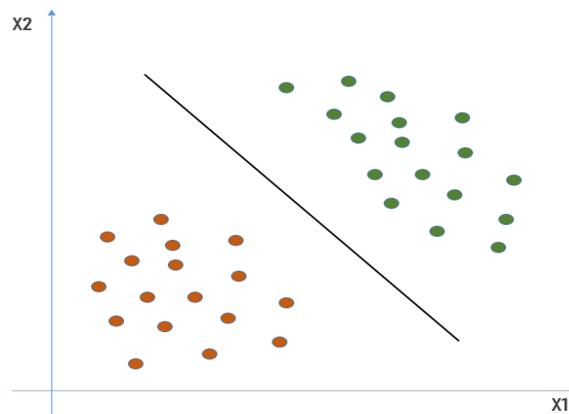


Figura 2.3: Ejemplo del resultado de una clasificación usando SVM

- **Redes bayesianas:** Una red bayesiana, red de creencia o modelo acíclico dirigido es un modelo probabilístico que representa una serie de variables de azar y sus independencias condicionales a través de un grafo acíclico dirigido. Una red bayesiana puede representar, por ejemplo, las relaciones probabilísticas entre enfermedades y síntomas. Dados ciertos síntomas, la red puede usarse para calcular las probabilidades de que ciertas enfermedades estén presentes en un organismo. Hay algoritmos eficientes que infieren y aprenden usando este tipo de representación [28].

### 2.1.4. Lenguaje Natural

En la filosofía del lenguaje, el lenguaje natural es la lengua o idioma hablado o escrito por humanos para propósitos generales de comunicación. En éste contexto, la comunicación verbal es el uso de sonidos y palabras para enviar un mensaje [20]. Éste tipo de mensajes pueden variar, dependiendo de la forma en la que se expresan y el escenario en el que se esté usando la expresión.

Una expresión utilizada en un ambiente laboral, seguramente constará de conversaciones con volumen moderado, sin gritos, sin palabras ofensivas, con un tema de conversación propio de un escenario de oficina; mientras que una expresión verbal agresiva o abusiva puede contener palabras antisonantes, insultos, tono de voz con decibeles altos (en comparación de la media establecida para niveles conversacionales) mismos que indican que puede preceder violencia, y aunque la expresión puede variar dependiendo del lenguaje, se requiere un análisis para entender el contexto de la situación [20].

El lenguaje natural entonces, es el lenguaje que hablamos todos los días, nuestra forma de comunicarnos por excelencia. Aunque para nosotros sea un gesto sencillo, casi inconsciente, el habla es un proceso que implica millones de conexiones neuronales y complejos procesos corporales de captación y comprensión. El lenguaje natural se diferencia de otros tipos de lenguaje porque surge de modo espontáneo entre la gente. Así, por ejemplo, aunque una computadora usa lenguajes propios para comunicarse con otros dispositivos, estos han sido

diseñados (impuestos) y los conocemos como algoritmos de computación. De igual forma, no sería un lenguaje natural, un idioma inventado y creado de modo artificial (como las señas que algunas personas crean para comunicarse con otra en específico).

En informática siempre se ha buscado que las computadoras sean capaces de comprender el lenguaje humano, de modo que obedezcan órdenes directas de las personas. Sin embargo, aunque la gramática es sencilla de asimilar para los ordenadores, no es así con el resto de variables. El timbre de cada persona (casi único), tono, los ademanes y gestos, etc. Todos ellos forman parte del lenguaje natural humano, y son mucho más difíciles de entender para una máquina. Sin embargo, los avances tecnológicos permiten que se hayan hecho ciertos experimentos con éxito, de captación de gestos, y reconocimiento de voz.

Estos avances [37] han permitido que haya surgido con fuerza una rama de la computación, llamada procesamiento del lenguaje natural, que busca precisamente permitir que los ordenadores sean capaces de captar la información transmitida por una persona y trasladarla a otra, interactuando con ella en su mismo lenguaje, o al menos, de un modo inteligible. El fin último sería un ordenador capaz de razonar y aprender como un ser humano; y para ello, el primer paso es que pueda recibir información y comunicarla como tal. Es por eso que esta rama es parte de la inteligencia artificial.

### 2.1.5. Lógica Difusa

Como se había mencionado en la parte introductoria del capítulo I, la lógica difusa es una rama de la inteligencia artificial cuya finalidad es otorgar cierto nivel de pertenencia a categorizar en cierta forma los valores en cuestión. Retomando el ejemplo de la estatura, el modelo propuesto se basa en reglas que determinarán a qué grupo o clase pertenece cierto valor.

En el caso específico de la detección de violencia verbal, el modelo servirá para determinar qué tan probable es que cierto audio contenga violencia verbal y en qué medida. El

diseño del modelo, descripción de variables, reglas y salida se detallará minuciosamente en la sección correspondiente.

### Sistemas Difusos Dinámicos

La lógica difusa por sí sola tiene un comportamiento estático. Esto significa que las entradas, reglas y salidas definidas no pueden cambiar. Para ejemplificarlo, imaginemos un partido de fútbol donde hay dos simples reglas:

1. **SI** el balón pasa la línea de porteria, **ENTONCES** gol.
2. **SI** el balón no pasa la línea de porteria, **ENTONCES** no gol.

Las palabras Si y ENTONCES están de esa manera, ya que las reglas se declaran de manera muy similar en lógica difusa. De hecho, utilizan la estructura IF-THEN.

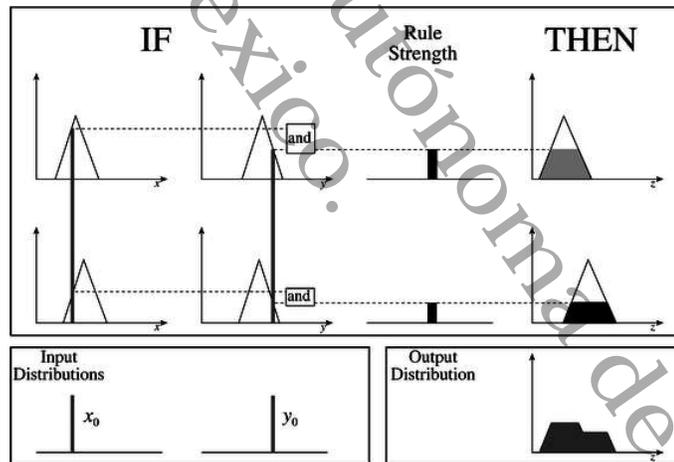


Figura 2.4: Ejemplo del uso de las reglas en un modelo.

Hasta este punto, el modelo está completo y cumple con la estructura propia de la lógica difusa. Pero, ¿Que pasaría si la jugada no es clara? hablando un poco de fútbol, las reglas dicen que el balón tiene que rebasar por completo la línea de gol para que pueda contar como auténtico. De lo contrario, no puede ser considerado como gol. Supongamos que el balón no rebasa por completo la línea como en la siguiente imagen:

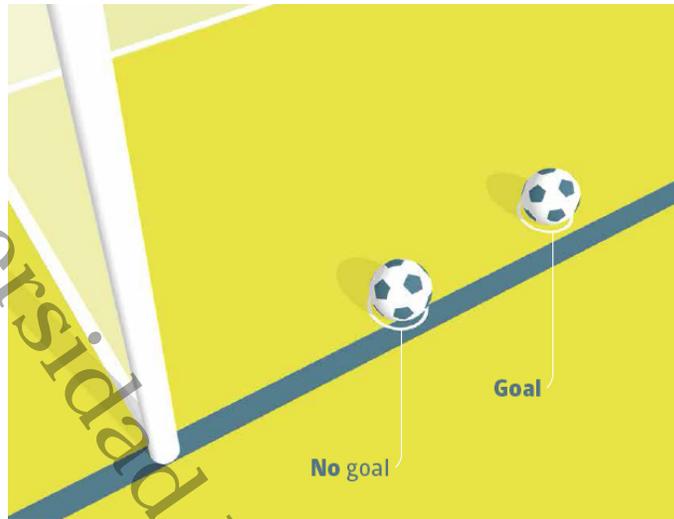


Figura 2.5: Ejemplo de gol legítimo.

Si nos basamos en la lógica convencional, la respuesta debería ser que sí hay gol, ya que el balón rebasó la línea. Pero es en este momento donde podemos comprender la función de la lógica difusa, ya que, en efecto, el balón rebasó la línea indicada, pero no cumple con todas las reglas para que el gol se cumpla. Por lo tanto, la respuesta es: **No Gol**.

## 2.2. Marco Tecnológico

Para llevar a cabo la investigación; dado que ésta se desarrolla en el campo de las ciencias de la computación, se requiere el uso de tecnología, misma que servirá para realizar las pruebas y experimentos para poder llegar a una conclusión. Dado que se trabajarán algoritmos especializados y avanzados, el uso de software y hardware especializado también es necesario.

Éste tipo de software que se utilizará necesita hacer uso de un licenciamiento por parte del fabricante o desarrollador, por lo que será necesario adquirir las licencias pertinentes para el software elegido y evitar así problemas legales. Dicho lo anterior, se enlista el software a utilizar:

- **Lenguaje R:** Es un lenguaje de programación con un enfoque al análisis estadístico. Es muy utilizado en investigación científica y es muy popular en el campo de aprendizaje

automático (*machine learning*). En éste lenguaje es posible cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y graficación.

- **Lenguaje Python:** Éste es un lenguaje de programación interpretado que hace énfasis en contar con un código legible. Como Java, es un lenguaje orientado a objetos y multiplataforma (Windows, MacOS, Linux). Actualmente, es utilizado para programar algoritmos de machine learning, como redes neuronales, bayesianas, SVM, entre otros.
- **Adobe Audition:** Con la finalidad de poder crear un dataset apropiado para realizar las pruebas y tratar de obtener mejores resultados, se pretende utilizar éste software para trabajar las pistas de audio y prepararlas para ser limpiadas de ruido y otras interferencias con que se pudiera contar. Éste software es un editor de audio profesional con filtros muy potentes y es por esa razón que fue elegido.
- **Matworks MatLab:** Es necesario el uso de éste software para realizar pruebas desde el punto de vista matemático de la investigación, ya que se requiere comprobar cada ecuación propuesta para determinar el comportamiento de ciertas características del sonido; fundamentando matemáticamente el ámbito de las ciencias de la computación.
- **Sistema Operativo:** Las pruebas se realizarán sobre dos sistemas operativos para poder comparar los resultados y detectar algunas diferencias que pudieran salir a la luz. Para ésta investigación se usará Microsoft Windows y MacOS. Los sistemas antes mencionados son compatibles con ambas plataformas, por lo que no debiera de haber problemas de compatibilidad de software.

Para instalar el software antes mencionado y realizar la fase de pruebas, es necesario contar con hardware que soporte los algoritmos de machine learning. Es por eso que se hará uso de dos equipos de cómputo con las siguientes características:

### 1. MacPro (Sistema Operativo MacOS)

- **Procesador:** Intel Xeon
- **Memoria RAM:** DDR3L 1666 Mega Hertz con capacidad de 64 Giga Bytes

- **Disco Duro:** 1 Tera Byte

## 2. PC de Escritorio (Sistema Operativo Windows)

- **Procesador:** Intel Core i3-3240 con velocidad de 3.40 Giga Hertz
- **Memoria RAM:** DDR3 1333 Mega Hertz con capacidad de 20 Giga Bytes
- **Disco Duro:** 1 Tera Byte
- **Unidad de Estado Sólido:** 120 Giga Bytes

Ambos equipos cuentan con elementos de altas prestaciones puesto que se requiere que los algoritmos sean compilados con velocidad y procesamiento robusto para que las pruebas no demoren mucho en obtener resultados.

## 2.3. Marco Referencial

Dentro de los marcos que conforman éste proyecto, existen referencias que están relacionadas de cerca con el proyecto a desarrollar. Éstas referencias son los trabajos de investigación que otros autores han hecho para atacar la misma problemática para la cual se llevará a cabo éste documento.

En éste artículo, Acar et al. [1] analiza clips de películas con contenido violento para determinar cuales títulos pueden ser consumidos por niños y jóvenes, ya que no existían muchos clasificadores de contenido para el periodo de tiempo en el cuál se desarrolló la investigación. La precisión promedio máxima obtenida fué de 93.8% usando *SVM con base radial* (algoritmo).

Otros trabajos que se encuentran relacionados fueron elaborados por Bautista-Durán et al. En donde, en uno de ellos [8], se hizo una clasificación de contenido violento en audio mediante el uso de redes neuronales, cuya finalidad es probar muestras de 5 segundos de duración obtenidas de dos diferentes datasets: el primero con audios extraídos de películas y

el segundo con audios recopilados de situaciones reales con violencia. Se utilizaron las redes neuronales como método principal, adicionando algoritmos de detectores cuadráticos (LSQD) en donde se obtuvo un factor de falsa alarma del 4 % lo que demuestra un 96 % de efectividad.

En otro paper de Bautista-Durán [7], la metodología consiste en instalar micrófonos en parques de una ciudad para monitorear el audio y detectar si una persona está siendo víctima de un asalto. Ésta es una solución que se trata de implementar para ciudades inteligentes, y la implementación del método consiste en analizar las características del audio (espectro y amplitud de la señal, frecuencia, energía entropía, entre otras) para diferenciar el sonido ambiental de la voz humana. Se evalúa también el costo computacional de la solución, tratando de determinar de qué forma se puede optimizar el algoritmo para no consumir mucho procesamiento de cómputo. Las pruebas se hicieron con archivos de audio de 5 segundos de duración usando el mismo dataset de su trabajo previo [8], obteniendo una probabilidad de detección del 78 % usando un clasificador cuadrático.

Es interesante mencionar que las emociones se ha tratado de clasificar transcribiendo las palabras en texto plano y así combinar el análisis textual y verbal, como lo propone Bhaskar et al. [10]. De ésta manera se lograron identificar 6 emociones (incluidas miedo y enojo) empleando el clasificador SVM y una herramienta desarrollada para el análisis de emociones conocida como *WordNetAffect* y empleando un clasificador (*SentiWordNet*) que funciona en conjunto con la herramienta antes mencionada. El dataset empleado fue *SemEval-2007* proporcionado por MultiMedia Eval, una base de datos multimedia muy conocida en el mundo de la investigación. Por separado, la clasificación de emociones verbales obtuvo un 57.1 % de efectividad, la clasificación en texto , un 76 % y la combinación híbrida de ambos métodos obtuvo un 90 %.

Por su parte, Giannakopoulos et al. [32] pretende identificar violencia en audio, ya que muchos trabajos proponen detectar violencia en video, y no ven los beneficios de analizar el audio; puesto que es mas rápido realizar el análisis y no necesita mucho recurso computacional.

Usando el clasificador SVM, los resultados muestran que, en promedio, 85.5 % de los archivos de audio fueron clasificados de forma correcta, mientras que el porcentaje de los segmentos que fueron correctamente identificados fue de 90.5 %.

Shulby et al. (2018): Solución de Inteligencia Artificial para detectar violencia verbal abusiva.

## 2.4. Marco Legal

Esta investigación se sustentará dentro del marco que corresponde a las leyes y ámbitos legales que rodean el problema en general. El Código Penal Federal de los Estados Unidos Mexicanos [24], establece en los artículos 372 y 374 el marco legal homologado por la Cámara de Diputados del H. Congreso de la Unión, en los cuales se describe la penalización para los delincuentes que practiquen el *robo con violencia*, el cual es el core de la investigación.

Adicionalmente, se establece en el artículo 95 de la Constitución Política de los Estados Unidos Mexicanos [53], que la reputación social de las personas que deseen ostentar un cargo político debe ser limpia, tal que quien quiera ejercer un cargo público deberá contar con antecedentes fuera de los registros de las Casas de Justicia de la nación como robo, fraude, falsificación, entre otros.

Además, el Código Civil Federal [25] establece en el capítulo VII (*“De las Personas Inhabiles para el Desempeño de la Tutela y de las que deben ser Separadas de ella”*), artículo 503.- *“No pueden ser tutores, aunque estén anuentes en recibir el cargo: ... V. El que haya sido condenado por robo, abuso de confianza, estafa, fraude o por delitos contra la honestidad;”*.

Las licencias de software que se usarán para llevar a cabo el proyecto también son parte del marco legal de la investigación. Estas licencias son para tres sistemas de software: **Adobe Audition** ya que éste software no es libre y por lo tanto requiere la adquisición de los derechos de licenciamiento por parte del desarrollador. **Mathworks Matlab** es otro siste-

ma que no es de uso libre puesto que es un sistema muy complejo y completo, y no es de sorprender que su licenciamiento requiera pago. De igual manera utilizamos el software de edición de video **Wondershare Filmora X** para poder hacer recortes necesarios, mismos que se utilizarán para el análisis de los segmentos de video. Éste software utiliza una licencia comercial, misma que se tendrá que adquirir para poder hacer uso de la herramienta. Y **Microsoft Windows**, uno de los sistemas operativos con los cuales se estará trabajando requiere de un licenciamiento para poder utilizar todas las funciones, ya que si se usa sin licencia: probablemente no se obtenga el desempeño deseado por parte de los programas y sistemas de software (además del hecho que el periodo de prueba tiene un tiempo corto de caducidad).

Universidad Juárez Autónoma de Tabasco.  
México.

Universidad Juárez Autónoma de Tabasco.  
México.

## Capítulo 3

# Modelo Dynamic Fuzzy Logic Propuesto

### 3.1. Adquisición de Datos

Como ya se había mencionado anteriormente, las cifras oficiales en cuanto a incidencias delictivas a lo largo del territorio nacional son preocupantes. Es por eso que el Gobierno del Estado de México decidió instalar cámaras de seguridad en las unidades de transporte público y a partir de su instalación, han salido a la luz en medios oficiales y redes sociales, videos que se han filtrado al público, donde se aprecian diversas situaciones en cuanto a asaltos.

A pesar de que hay algo de material en la red, no es posible obtener todo el conjunto ni siquiera actualizado, ya que el C5 resguarda esa información como provada o de uso por personal autorizado, razón por la cuál nuestro trabajo principal para obtener los datos y generar el dataset correspondiente tuvo que ser realizado por medio de una recopilación en plataformas como Youtube y Facebook, así como para obtener los metadatos (fecha, contexto, información adicional) de los mismos en páginas oficiales de noticieros y periódicos importantes del país.

Para obtener los datos, se realizó una búsqueda en la plataforma YouTube usando las siguientes cadenas de texto:

- “ASALTO COMBI ESTADO DE MÉXICO”

- “ASALTO ESTADO DE MÉXICO”
- “ASALTO CIUDAD DE MÉXICO”
- “ROBO TRANSPORTE PÚBLICO CDMX”
- “ROBO VIOLENTO ESTADO DE MÉXICO”

Después de obtener resultados, el siguiente paso consistió en observar el video en busca de elementos clave como lo son:

- Claridad en las palabras y el audio en general del video
- Palabras más frecuentes en un asalto (ejemplo): Cartera, bolsa, balazo, iris, dinero, teléfono, celular, entre otras.
- Que el tono de la voz de los asaltantes fuera violento o enojado (alterado).

Una vez que el video cumplió con los requisitos, cada uno de ellos se editó y separó en partes de 5 segundos que llamamos ventanas. Éstas ventanas son periodos de tiempo en los cuáles se divide cada video. Por ejemplo, si un video tiene una duración total de 2 minutos, se esperarían 24 ventanas:

$$v = \frac{n}{5} \quad (3.1)$$

Si queremos obtener el total de ventanas de tiempo en un video ( $v$ ), simplemente basta con dividir el número total de la duración del archivo en segundos ( $n$ ), en éste caso un video de 2 minutos es igual a 120 segundos, éste valor lo dividimos por 5, que es la duración de una ventana de tiempo en segundos. Obtenemos un valor de 24:

$$\frac{120}{5} = 24 \quad (3.2)$$

Pero no todo es tan sencillo, ya que esta fórmula aplica en un escenario ideal, sin elementos que obstruyan el análisis. Lamentablemente el material obtenido son fragmentos recopilados de noticieros y adaptaciones que hacen los usuarios en internet. Es por eso que el segmento

para analizar debe ser aislado primero, para ello usaremos la siguiente fórmula:

Sea  $v$  el total de ventanas a obtener,  $nf$  el tiempo en segundos donde termina el incidente,  $ni$  el tiempo en segundos donde inicia el incidente, esto dividido por 5, la fórmula queda de la siguiente manera:

$$v = \frac{nf - ni}{5} \quad (3.3)$$

Lo anterior aplica para archivos de video que en su inicio cuentan con un relato por parte de algún comentarista. Todo ese tiempo necesita excluirse de ser considerado en las ventanas de tiempo, ya que no son relevantes para el análisis.

Lo anterior significa que únicamente se debe registrar a partir del tiempo en que inicia el evento y el tiempo donde termina. Una vez terminado el registro, los archivos se fueron almacenando en una carpeta digital para su posterior análisis.

El análisis consta en identificar dos variables sumamente importantes, y adicionalmente, otras dos que sirven para determinar el resultado final.

Para cada archivo, se identificó la marca de tiempo en el que inicia el evento y la marca dónde termina. Posteriormente, el tiempo total es dividido en segmentos de 5 segundos, basándonos en un artículo que arroja muy buenos resultados [32].

Cada segmento de tiempo se analiza en busca de los siguientes parámetros: Variables con más peso

- Tono de voz del asaltante (Emoción). Para categorizar el nivel de la emoción, nos basamos en la escala NOVACO Anger Scale (NAS) [45] y a partir de ahí, generamos nuestras propias categorías.
- Número de palabras mencionadas en el segmento que corresponden con el análisis previo

de frases o palabras más frecuentes.

- Variables de contexto (lo explicaremos más a fondo)
  - Hora del asalto.
  - Ubicación geográfica del evento.

Con estos parámetros, se fue construyendo el Dataset para entrenar el modelo difuso desarrollado en MATLAB, usando el paquete de herramientas que incluye para lógica difusa.

### 3.1.1. Procesamiento de los Datos

Las ventanas de tiempo que se obtuvieron después de analizar los archivos de video, solamente son una mera referencia si los datos no son procesados. Y esta es una tarea difícil si los archivos de video se usan en su forma original, ya que al ser resubidos múltiples veces a diversas plataformas, es incierto el nivel de calidad de audio en dichos archivos, ya que en muchos de ellos no se aprecia de forma precisa lo que los asaltantes están diciendo.

Para realizar éste procesamiento, es necesario que el medio digital cumpla con ciertos requisitos, pero antes de explicar en detalle cada uno de estos, fue necesario usar un editor de video para poder realizar el recorte de cada ventana de tiempo para su posterior análisis. Para esta tarea usamos el software con licencia **Wondershare Filmora X**.

Con éste software, podemos separar cada uno de los segmentos de video, con el objetivo de analizar cada uno de ellos por separado. Una vez obtenidos todos los segmentos, entonces procedemos con los demás puntos para el correcto procesamiento de los segmentos:

- **Claridad en las palabras y el audio en general del video:**

Para esta tarea, entra en función el uso del software **Adobe Audition** con el cual aplicaremos filtros al audio de los archivos de video, tratando de hacer que las voces presentes se acentúen, de manera que el ruido de vehiculos, tráfico y otras frecuencias que no sean las de voz humana, disminuyan su volumen en la medida de lo posible.

- Palabras más frecuentes en un asalto (ejemplo): Cartera, bolsa, balazo, iris, dinero, teléfono, celular, entre otras.
- Que el tono de la voz de los asaltantes fuera violento o enojado (alterado).

Fue necesario hacer el procesamiento adecuado de los datos ya que gracias a esto se logro fabricar el dataset que alimentaria el modelo difuso. La edición de los videos resultó de gran ayuda y el modelado de los datos se llevó a cabo de forma satisfactoria.

De igual manera, se utilizó el software **Adobe Audition** para aplicar filtros a cada uno de los segmentos, y quitar ruidos, aclarar voces; de tal manera que el sonido fuera un poco más limpio y las palabras se pudieran apreciar con más claridad.

Éste software fue creado con la finalidad de mejorar las producciones musicales, por lo que tiene herramientas muy útiles e interesantes.

### 3.1.2. Dataset

El set de datos, o dataset en inglés, es un elemento de uso principal en el desarrollo de proyectos de análisis de datos; en especial cuando cualquier algoritmo de inteligencia artificial se ve involucrado.

Un set de datos es una recopilación de valores importantes que tienen en común un mismo escenario o situación. Por ejemplo, si se quisiera determinar cuales son los factores mas importantes que intervienen para que las ventas de un negocio puedan dar mayores ganancias, es necesario recopilar la información de las compras de todos los clientes.

Como ejemplo práctico, tomemos el caso actual de una de las cadenas comerciales más importantes en la actualidad [36]. Ésta tienda ubicada en Estados Unidos, hace uso de inteligencia artificial para determinar qué productos son los que mayormente compran las personas.

Después de realizar un estudio, los resultados arrojaron que en casi todos los tickets que contenían compra de pañales, también se compraba leche y cervezas. De tal modo que los pañales, leche y cervezas están acomodados en el mismo pasillo, para que el cliente no tenga que buscar en otros pasillos los productos que necesita.

Pero también, cuando el padre de familia llega a la tienda por la despensa, sienta la necesidad de adquirir los productos, ya que están a la mano en el mismo pasillo.

TID	Market	Items
1	A	Bread, Milk, Peanut
2	A	Bread, Diaper, Beer, Eggs, Peanut
3	B	Milk, Diaper, Beer, Cola
4	B	Bread, Milk, Diapers, Beer
5	C	Diaper, Beer, Eggs
6	D	Bread, Beer, Peanut, Eggs
7	D	Beer, Milk, Peanut
8	E	Beer, Peanut, Diaper
9	E	Bread, Milk, Cola, Eggs
10	F	Beer, Milk, Peanut, Cola, Eggs

Figura 3.1: Ejemplo del resultado del análisis realizado

El ejemplo anterior, sirve para entender lo siguiente: Un dataset siempre tendrá en común que todos los datos compilados forman parte de una misma situación o escenario.

Por ejemplo, el estudio de la tienda se hizo en los tickets de compra. Pudieron ser tiendas de marcas diferentes, incluso ciudades o países diferentes, pero el dataset está compuesto por compras en tiendas comerciales.

De igual manera, la recopilación de datos que se hizo para construir el set de datos del actual estudio fue sobre videos existentes en internet, específicamente donde se aprecian asaltos con violencia en unidades del transporte público del estado de México.



Figura 3.2: Imagen tomada por cámaras de seguridad en transporte público del estado de México

Los datos contenidos en el set de datos no tienen que seguir una estructura en específico, pero si deben considerar tener una clase. Esta clase nos indica el resultado de los datos. Por ejemplo, una vez registrados los componentes de datos en el segmento de 5 segundos, como lo son: número de palabras mencionadas, hora, fecha, emoción registrada, ubicación dentro de la ruta; se sabe que ese segmento de 5 segundos contiene violencia verbal, o no. Entonces, se especifica la clase 01 como Violento, y clase 02 como No Violento.

File	Emotion	Novaco (NAS)	# P	Words	Location	Start	End	Time	Date	Vector	Class	
12	No_Enojado	-1	1			30	00:03	00:08	00:05	27-dic-18	[1,30,-1]	02
12	No_Enojado	-1	0			30	00:08	00:13	00:05	27-dic-18	[0,30,-1]	02
12	Poco_Enojado	1	1			26	00:13	00:18	00:05	27-dic-18	[1,26,1]	02
12	Poco_Enojado	1	3	VERI	EFONO	22	00:18	00:23	00:05	27-dic-18	[3,22,1]	02
12	Muy_Poco_Enojado	0	2			20	00:23	00:28	00:05	27-dic-18	[2,20,0]	02
12	No_Enojado	-1	4			26	00:28	00:33	00:05	27-dic-18	[4,26,-1]	02
12	Muy_Enojado	4	3	TE	PUTA	28	00:33	00:38	00:05	27-dic-18	[3,28,4]	01
12	Poco_Enojado	1	2			17	00:38	00:43	00:05	27-dic-18	[2,17,1]	02
12	No_Enojado	-1	0			16	00:43	00:48	00:05	27-dic-18	[0,16,-1]	02
12	Muy_Poco_Enojado	0	2			30	00:48	00:53	00:05	27-dic-18	[2,30,0]	02
12	No_Enojado	-1	0			23	00:53	00:58	00:05	27-dic-18	[0,23,-1]	02
12	Moderadamente_Enojado	2	1			19	00:58	01:03	00:05	27-dic-18	[1,19,2]	02
12	Muy_Poco_Enojado	0	2			22	01:03	01:08	00:05	27-dic-18	[2,22,0]	02
12	No_Enojado	-1	1			19	01:08	01:13	00:05	27-dic-18	[0,19,-1]	02
12	No_Enojado	-1	1			29	01:13	01:18	00:05	27-dic-18	[1,29,-1]	02
13	Poco_Enojado	1	4	CARTE	LULAR (2)	22	00:25	00:30	00:05	26-ene-20	[4,22,1]	02
13	Poco_Enojado	1	4			18	00:30	00:35	00:05	26-ene-20	[4,18,1]	02
13	Enojado	3	2			30	00:35	00:40	00:05	26-ene-20	[2,30,8]	01
13	Enojado	3	1			21	00:40	00:45	00:05	26-ene-20	[1,21,3]	02
13	Enojado	3	1			16	00:45	00:50	00:05	26-ene-20	[1,16,3]	02
13	Muy_Enojado	4	4	F	(2)	16	00:50	00:55	00:05	26-ene-20	[4,16,4]	01
13	Enojado	3	0			28	00:55	01:00	00:05	26-ene-20	[0,28,3]	02
13	Poco_Enojado	1	0			24	01:00	01:05	00:05	26-ene-20	[0,24,1]	02
13	Enojado	3	0			30	01:05	01:10	00:05	26-ene-20	[0,30,3]	02
13	No_Enojado	-1	0			18	01:10	01:15	00:05	26-ene-20	[0,18,-1]	02
13	Moderadamente_Enojado	2	2			16	01:15	01:20	00:05	26-ene-20	[2,16,2]	02
13	No_Enojado	-1	1			17	01:20	01:25	00:05	26-ene-20	[1,17,-1]	02
14	Moderadamente_Enojado	2	4	CHIN	NO (2)	29	00:10	00:15	00:05	05-ago-19	[4,29,2]	01
14	Moderadamente_Enojado	2	4		(3)	22	00:15	00:20	00:05	05-ago-19	[4,22,2]	01
14	Poco_Enojado	1	2			29	00:20	00:25	00:05	05-ago-19	[2,29,1]	02
14	Moderadamente_Enojado	2	2			29	00:25	00:30	00:05	05-ago-19	[2,29,2]	02
14	Moderadamente_Enojado	2	2			16	00:30	00:35	00:05	05-ago-19	[2,16,2]	02

Figura 3.3: Fragmento del dataset generado (se censuran las palabras ofensivas)

Como se puede observar en la imagen anterior, el último campo del dataset tiene la etiqueta **Class**, esto es debido a que el modelo de aprendizaje que se está utilizando es supervisado, lo que significa que el resultado de cada segmento se sabe con anterioridad. A diferencia del aprendizaje no supervisado, en el cual no se tiene conocimiento del resultado, sino que el algoritmo lo aprenderá conforme la lectura de los datos. El total de segmentos del dataset fue de 548 segmentos de 5 segundos cada uno, provenientes de 38 videos compilados.

## 3.2. Descripción del Modelo

Como ya se había mencionado en la parte de la Introducción, ésta investigación consta de un proyecto que hace uso de la lógica difusa, proponiendo un modelo que aparte de ser difuso, es dinámico. Pero la parte dinámica del modelo es algo diferente a otras investigaciones que se han realizado anteriormente.

En los trabajos que existen en el estado del arte, la parte dinámica del modelo es implementada en una sola etapa del proceso. Es decir, en la entrada, el motor de inferencia (reglas) o la salida. Pero el presente proyecto propone un modelo difuso con la parte dinámica en las tres etapas. Esto lo hace un modelo único, con un porcentaje de efectividad mayor.

### 3.2.1. Preparación del Entorno

El modelo difuso fue desarrollado e implementado en el software MATLAB, uno de los sistemas que mejor se adapta a este tipo de proyectos. Pero antes de comenzar a describir cada una de las etapas, fue necesario crear un entorno personalizado que se adaptara a las necesidades propias del proyecto.

Para esto, vamos a explicar el funcionamiento de algunas funciones que intervienen en las variables de entrada.

El modelo consta de tres variables de entrada, las cuáles se denominan variables lingüís-

ticas:

- **Palabras**
- **Ubicación dentro de la ruta**
- **Emoción**

De igual manera, para un entorno ideal, fue necesario modificar ciertas funciones de MATLAB para el correcto funcionamiento del modelo.

Como preparación para las variables de entrada y de salida, se tuvo que llevar a cabo un proceso que consistió en generar una función por medio de código para que los parámetros de entrada de cierta función admitieran un parámetro nuevo. Los detalles de la implementación son profundizados en cada una de las variables.

### 3.2.2. Variables de Entrada

Las variables de entrada es la primer etapa del modelo, son los datos crudos que serán procesados más adelante en la etapa de las reglas o motor de inferencia, en esta etapa se determina una parte crucial en el diseño del modelo, ya que dependiendo del comportamiento de los datos de entrada, es el comportamiento de los datos de salida.

Así que esta es una etapa que requiere de mucho análisis y recopilación de datos. Las variables de entrada son las siguientes:

- **Palabras:** la variable lingüística palabras contiene tres valores lingüísticos: Poca Coincidencia, Mediana Coincidencia y Alta Coincidencia. El rango de estos valores va de 0 a 7. Al analizar los segmentos de video, la presencia de 0 a 2 palabras significa Baja Coincidencia, 1 a 3 palabras significa Mediana Coincidencia y 3 o más palabras significa Alta Coincidencia.

Ahora bien, surge una pregunta interesante: ¿Por qué motivo el rango 0-2 interseca con 1-3? es decir, la presencia de 1 palabra pertenece a dos grupos, Baja y Mediana coincidencia.

Esto es debido a que en lógica difusa, es necesario que exista una intersección entre dos o más grupos, ya que de otro modo no podría ser llamada lógica difusa. La intersección entre grupos es la parte difusa.

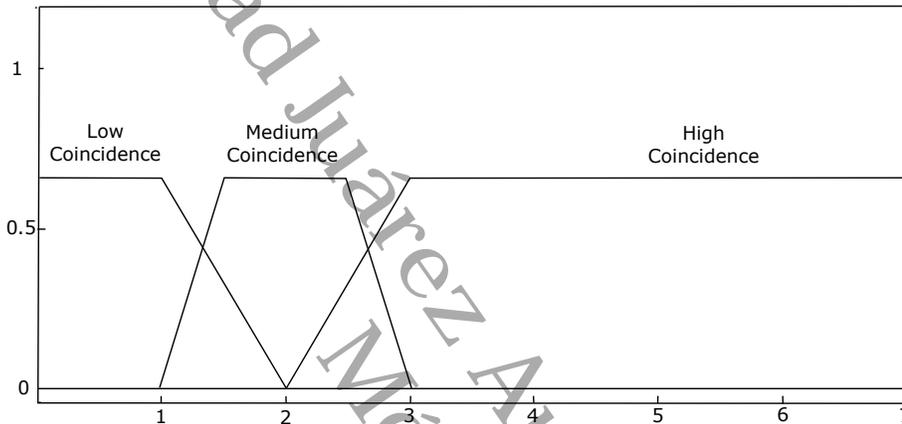


Figura 3.4: Diseño de la variable Palabras

El rango de 0 a 7 fué establecido de esa forma debido a que producto del análisis realizado a cada uno de los videos, se detectó que en promedio, son mencionadas al redor de 7 palabras durante un asalto con violencia en unidades de transporte público.

La variable está conformada por tres funciones de membresía, las cuales son baja, media y alta coincidencia de tipo trapezoidal. Cada función de membresía consta de un vector de entrada, el cual tiene el siguiente formato: [START, MAX, DOWN, FINAL]

- **START:** Es el punto de inicio de la función de membresía. Es el punto donde comienza la configuración del vector, dependiendo del diseño establecido.
- **MAX:** Es el punto máximo que puede alcanzar la función. A partir de ese punto, se genera una línea recta con el punto declive (DOWN).
- **DOWN:** Punto declive. Es el punto donde termina la constante y empieza el

declive hacia el punto final. Generalmente es donde se hace la intersección entre funciones de membresía.

- **FINAL:** Es el punto donde termina el rango de la función de membresía. Por lo general, se forma una pendiente entre el punto declive y el punto final.

El formato se observa de la siguiente forma: [ 1, 1.5, 2.5, 3 ]

Si nos ubicamos dentro de la variable, cada uno de los parámetros del vector toman una forma trapezoidal:

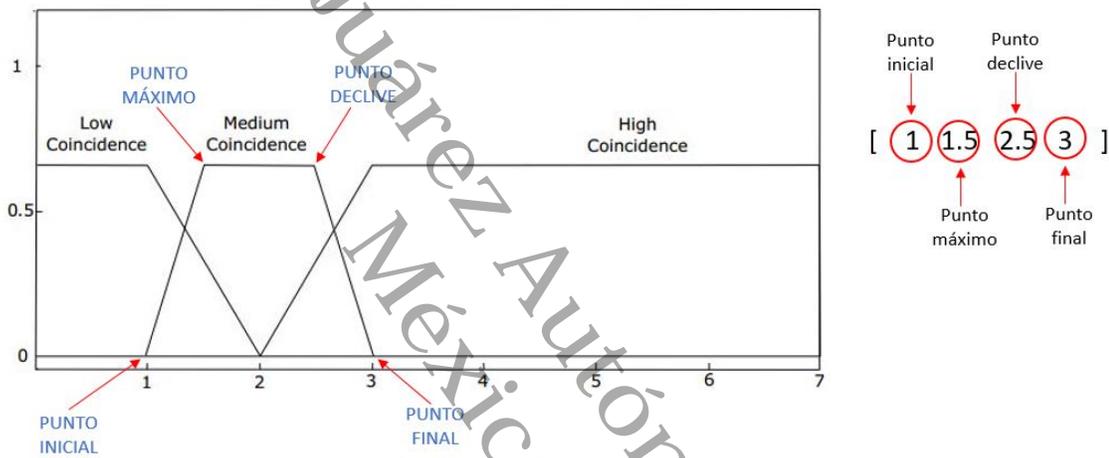


Figura 3.5: Parámetros de configuración del vector en la variable

Adicionalmente, era necesario añadir la parte dinámica al vector dentro de la misma configuración del vector; por lo que fue conveniente modificar ciertas funciones programadas dentro del software de MATLAB, con el propósito de agregar un quinto parámetro al vector de entrada.

Este quinto parámetro es el valor de la amplitud de la variable, mediante el cual se controla en cierta medida el impacto que tendrá el valor de salida, desde el valor de entrada. Esto no significa que se tiene un control a voluntad de la salida del modelo. Lo que significa es que este parámetro está disponible para ser modificado por un controlador, es decir, un agente externo; el cual tomará la decisión del valor de la amplitud

de la variable, una vez que el modelo se implemente en un dispositivo que tenga esa capacidad. Tomando en cuenta lo anterior, se logró obtener un vector como el siguiente:

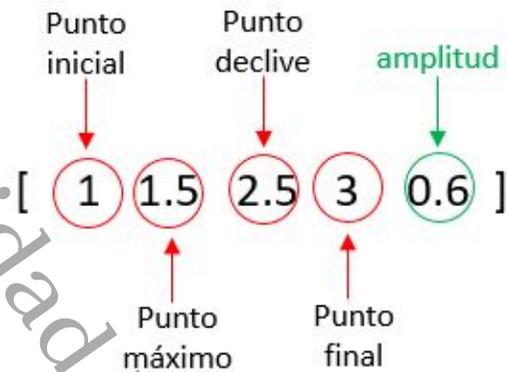


Figura 3.6: Parámetro de amplitud en vector de entrada

Este último valor, denota el valor de la amplitud. Con este parámetro, se puede controlar en qué nivel se necesita la amplitud de la variable. Por ejemplo, si es necesario que el valor esté en 0.3, simplemente se agrega. Con esto podemos controlar los valores necesarios configurados en el modelo:

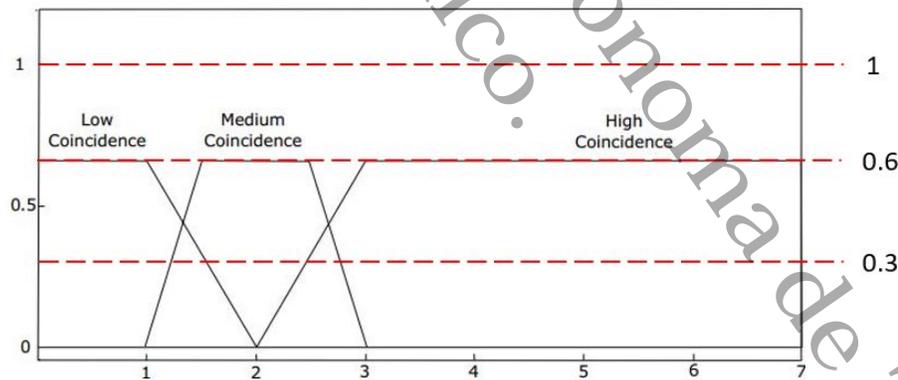


Figura 3.7: Niveles de amplitud configurables para el modelo

Así, obtenemos los tres valores de configuración posibles para la amplitud de la variable de entrada. Si observamos cada uno de los vectores, los niveles pueden visualizarse de la siguiente forma:

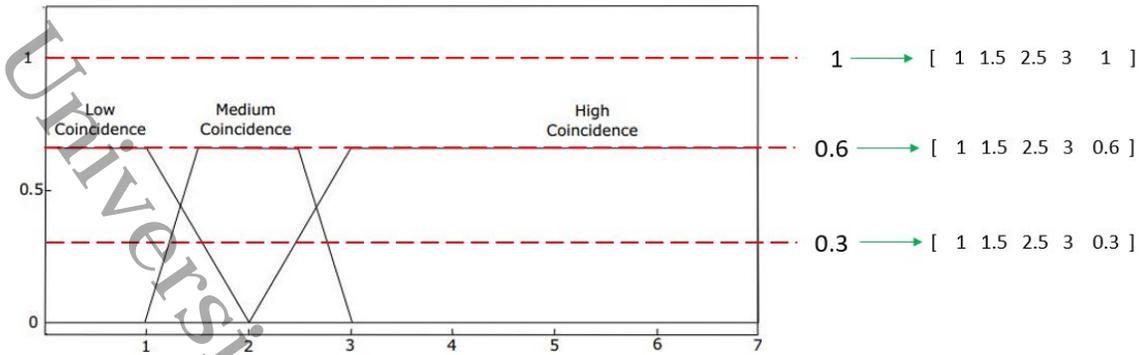


Figura 3.8: Niveles de amplitud acordes al vector

De manera predeterminada, la interfaz de MATLAB (Fuzzy Logic Toolbox) no admite este último valor que en este caso, es el valor de la amplitud. Pero sí es posible agregar una función de membresía customizada en las opciones del menú.

Estas funciones de membresía customizadas o personalizadas, son funciones específicas que el usuario necesita programar para poder hacer uso de ellas. Ya que por lo general, son funciones que no se ajusta a las convencionales, ya que necesitan un comportamiento poco común. Para el desarrollo del modelo, fue necesario programar una de éstas funciones personalizadas.

Para hacer ésto posible, se tomó el diseño de la función de membresía trapezoidal, que es el único grupo presente en esta variable.

De tal manera que se programó una función en la consola de MATLAB que mandara a llamar el algoritmo del trapecio para multiplicarlo por el número de la variable de la amplitud que el usuario escriba en el vector de configuración. La función queda de la siguiente forma:

```

% Función para generar función de membresía Trapezoidal
% Usando 5 parámetros:
% Parámetros 1 al 4 determinan la forma del trapecio
% 1: valor de inicio, 2: valor de ascenso,
% 3: valor de descenso, 4: valor de término
% El parámetro 5 determina la amplitud de la función
function out = custmfTrap(x,params)
out = params(5)*trapmf(x,[params(1) params(2) params(3) params(4)]);
end

```

Figura 3.9: Función de amplitud por medio de consola en MATLAB

Ésta función es llamada en el momento de configurar la variable de entrada. De igual manera como se hizo con la función trapezoidal, se hizo con la función triangular, misma que se especificará más adelante, cuando sea el momento de detallar la variable pertinente.

- **Ubicación dentro de la ruta:** Otra de las variables de entrada del modelo propuesto es la ubicación dentro de la ruta.

Esta variable fue diseñada de acuerdo a reportes presentados por la secretaría de seguridad pública del estado de México y de acuerdo a la **Encuesta Nacional de Seguridad Pública Urbana (ENSU)** del cuarto trimestre del año 2021 [29], las personas encuestadas reportan sentirse inseguras en el transporte público de su ciudad (70.2%), lo cual es alarmante; ya que a pesar de los esfuerzos por parte de las autoridades por disminuir los índices de violencia, tal pareciera que son en vano, ya que la delincuencia sigue operando de manera normal e incluso los índices han ido en aumento sin importar las medidas que se tomen.

Después de observar los videos recopilados, se determinó que los asaltantes tienen zonas de trabajo establecidas, y que los lugares en donde se llevan a cabo los incidentes de inseguridad son zonas ya identificadas por los organismos de seguridad y por los usuarios del transporte público.

En 2020 se hizo un estudio en la ciudad de bogotá, Colombia. En donde se analiza el acoso sexual en el transporte público, como una forma de crimen.

En el artículo [55] se establece que hay ciertos lugares en la ciudad en donde los acosos son más frecuentes; como por ejemplo, sitios de construcción, bares, discotecas, e incluso los autores detallan que hay ciertos asientos dentro la unidad de transporte público que los acosadores tienen establecidas como lugares estratégicos para hacer sus movimientos.

Adicionalmente, los noticieros que reportan éste tipo de noticias mencionan el lugar donde sucedió el asalto, para que la población evite estas rutas o use vías alternas. Esto corrobora que las zonas de asaltos están identificadas por los medios y que forman parte de un tramo de la ruta.

Otro estudio que podemos encontrar es un reportaje realizado por Carlos Millares [43] en donde se detallan cuáles son las carreteras más peligrosas de México, por tramo (es decir, de ciudad a ciudad) y en algunos casos, en qué kilómetro se han reportado incidencias.

- **Emoción** Otra de las variables de este modelo es la variable Emoción. Consiste una parte extremadamente importante de este modelo, ya que es una de las variables que definirán en gran medida el resultado final en la salida.

Imaginemos por un momento que estamos sentados frente a una pantalla con los ojos cerrados y escuchamos el audio del video de un asalto. Inmediatamente nos daríamos cuenta que lo que estamos escuchando es lo que es.

Pero, ¿cómo nos pudimos dar cuenta de esa situación? ¿cómo supimos de que se trataba? La respuesta es relativamente sencilla. El ser humano transmite sus emociones por medio de la voz. Para ser más específico, esas emociones se expresan por medio del tono medio de la voz. Es fácil distinguir cuando una persona siente preocupación

o estrés. También cuando una persona está feliz o enojada. Son emociones muy fáciles de detectar por otro ser humano, ya que los tonos de la voz son muy característicos [47].

Es por eso que la razón principal para desarrollar la presente tesis consiste en identificar esas características en la voz, ya que en el ejemplo anterior, en donde imaginamos escuchar la voz con los ojos vendados o cerrados, representa una gran oportunidad para detectar el evento de un asalto de manera muy característica y fácil para un ser humano, ya que los tonos de voz que utilizan los asaltantes mientras están amedrentando a sus víctimas, la mayoría de las veces son con tonos de violencia, o lo que es lo mismo, la emoción enojo, la cuál es la emoción que estaremos analizando más adelante.

En el análisis previo realizado para obtener las características mas importantes a evaluar en situaciones de un asalto con violencia, se encontró que en el 99% de los casos, las cámaras de seguridad registraron asaltos en donde el tono de voz de los asaltantes tiene las características de una voz violenta, dando órdenes a sus víctimas para causar pánico y de esa manera generar tensión y hacer que los pasajeros de la unidad de transporte público obedezcan.

### 3.2.3. Variables de Contexto

El contexto de una situación nos ayuda a entender mejor una escena de lo que tal vez parece. Por ejemplo, si empezamos a escuchar una conversacion entre dos personas cuando la misma ya esta avanzada, es decir, no la escuchamos desde el principio, tal vez entendamos algo totalmente diferente de lo que realmente se está hablando.

A manera de ejemplo, tomemos como referencia la siguiente imagen en donde podemos ver que una niña con bata de hospital está observando por una cámara de cine profesional. En la imagen podemos observar dos diferentes tipos de escenario.



Figura 3.10: Ejemplo de contexto

1. La joven con ropa de hospital está pasando por una etapa de una enfermedad terminal y su sueño era conocer cómo se graba una escena en la pantalla grande, así que sus padres lograron cumplir su sueño llevándola a un set de grabación.
2. La joven es una actriz de una película y está corroborando la toma de la escena para saber donde situarse exactamente y la toma salga acorde con lo planeado.

Realmente, si lo pensamos por un momento, no sabríamos qué decisión tomar; pareciera que cualquiera de las dos situaciones es la correcta y no sería fácil la decisión, dado que nos falta algo de información para poder saber a ciencia cierta cual de las dos opciones tomar.

Ahora, como información adicional, tenemos que la joven que se muestra en la imagen es una actriz muy famosa de los últimos tres años, su rostro se nos hace familiar, ya que recordando un poco, la hemos visto participando en comerciales en televisión, y la hemos visto en alguna portada de revista, por lo que su cara nos resulta familiar.

Con ésta información relevante, es mucho más fácil tomar una decisión final. Dadas las circunstancias, decidimos irnos por la segunda opción. La imagen es una fotografía tomada dentro de un set de televisión y la joven es realmente una actriz, llevando a cabo una escena

en donde se encuentra recuperándose en el hospital.



Figura 3.11: Ejemplo de contexto (Decisión tomada)

La situación anterior fue solo un ejemplo para clarificar cómo funciona el contexto en las variables diseñadas para nuestro modelo difuso, ya que éstas variables juegan un papel importante en la toma de decisiones y en el comportamiento de las reglas del modelo, las cuáles forman parte del motor de inferencia.

Información recabada por parte de las autoridades de la Ciudad de México, arrojan que, en un periodo de tres años, hay dos épocas en el año en donde hay más denuncias de asaltos con violencia por parte de los ciudadanos.

Estas épocas son Durante los meses de Mayo y Noviembre. Tal vez suene raro, ¿porqué

esos meses son los que más registros presentan? ¿cuál es el patrón que siguen éstas denuncias? Las autoridades suponen que durante el mes de Noviembre, muchas compañías pagan aguinaldos a sus empleados, ya que así lo establece la ley federal del trabajo. Y en cuanto al mes de Mayo aún no se logra determinar la causa más probable, ya que no hay ningún pago o festividad importante durante ese mes.

La gráfica presentada por la secretaría de seguridad y protección ciudadana es la siguiente:

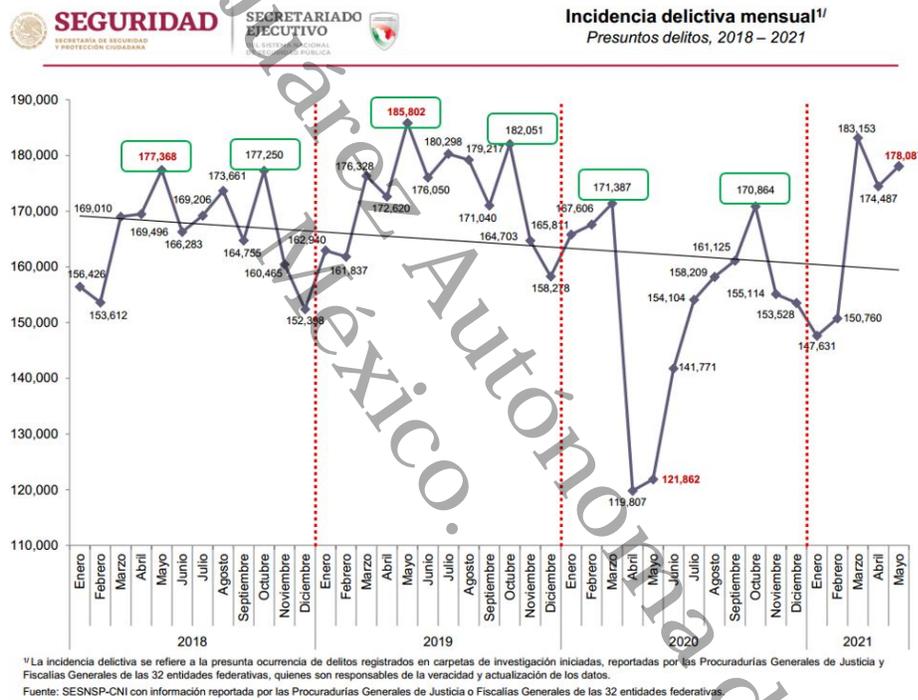


Figura 3.12: Gráfica de incidencia delictiva 2018-2021

Otras fuentes extraoficiales afirman que todos los meses hay altos índices de incidencias delictivas, en específico, cada 15 días, ya que en México, el esquema de pagos es quincenal y no mensual como en otros países y en diversas organizaciones. Pero los datos oficiales arrojados por la secretaría de seguridad pública son los que se muestran en la imagen.

Las variables de contexto entonces, serán aquellas que no forman parte del modelo difuso,

pero que influyen directamente a la toma de decisiones en la salida del mismo.

- **Hora del Asalto** La hora del asalto, es una variable que nos permitirá definir el peso de la variable de salida mediante la aplicación de las reglas. Ésto significa que si las palabras del asaltante son 3 groserías, la emoción es moderadamente enojado y el kilómetro de la ruta de la unidad pública no está dentro del rango definido como peligroso, la salida será la siguiente: **Violento**.

Ésto quiere decir que el audio contiene violencia pero no lo suficiente como para afirmar que en efecto, se trata de un asalto, pudiera ser que se trata de una pareja en una discusión, por lo que los niveles de emoción y palabras pudieran ser algo altos.

Sin embargo, si tomamos en cuenta que este escenario se está desarrollando a las 10:00 pm y la ubicación geográfica es una de las zonas más peligrosas que se muestran en el mapa de incidencias, entonces las posibilidades de que el evento sea un asalto con violencia aumentan significativamente, cambiando el estado de salida de Violento a **Extremadamente Violento**.

- **Ubicación Geográfica** Otra de las variables de entorno que influyen en la toma de decisiones en la variable de salida es la ubicación geográfica.

La secretaría de seguridad pública de la Ciudad de México recopila información de manera constante de acuerdo a los reportes producto de las denuncias que la ciudadanía levante en las diversas delegaciones.

Éstos reportes se concentran y se envían a diversas instituciones y dependencias de gobierno, para generar concentrados estadísticos que posteriormente se analizan y sirven para generar gráficas y tener una imagen más clara de lo que está pasando en la ciudad.

Gracias a este registro de datos, las autoridades tienen un panorama completo muy

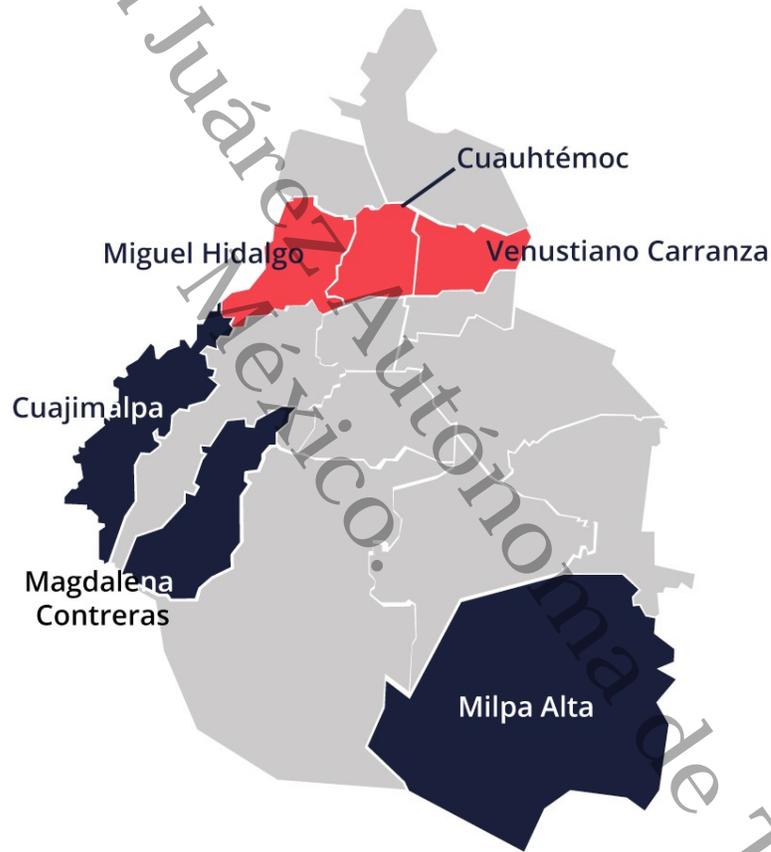
aproximado acerca de las zonas con más actividad delictiva en la ciudad de México y sus alrededores. Éstas son las zonas con más actividad identificadas por las autoridades correspondientes:

**DELEGACIONES MÁS Y MENOS INSEGURAS**

nación ③②①

Estas fueron las delegaciones con mayor y menor número de delitos por cada 100 mil habitantes, de enero a marzo de 2017

■ mayor incidencia ■ menor incidencia



Fuente: Dirección General de Política y Estadística Criminal de la PGJ CDMX

Figura 3.13: Mapa de zonas más inseguras CDMX

Con estos datos, pudimos diseñar el modelo difuso tomando en cuenta las variables de entorno.

### 3.2.4. Motor de Inferencia

Todo modelo difuso necesita tres elementos para poder funcionar de manera correcta:

1. Variables de Entrada
2. Reglas (Motor de Inferencia)
3. Variables de Salida

Las reglas del modelo son las directivas de procesamiento que determinarán el procesamiento del mismo. Las reglas se componen de una serie de condiciones IF-THEN y conectores. En éste caso utilizaremos conectores AND. Una regla puede verse de la siguiente manera:

**IF**

palabras=3 **AND** ubicacion=PELIGROSO **AND** emocion=MUY\_ENOJADO

**THEN**

salida=EXTREMADAMENTE\_VIOLENTO

Para el uso del motor de inferencia nos centraremos en el contenido del condicional IF:

{ palabras=3 **AND** ubicacion=PELIGROSO **AND** emocion=MUY\_ENOJADO }

Las reglas se comportan de una manera muy sencilla: si se cumplen las condiciones, entonces la salida muestra el valor configurado en el diseño de las reglas del modelo.

Cada una de las reglas fué diseñada y programada de acuerdo a un entrenamiento previo del modelo, en donde tomando en cuenta las variables de entorno, se obtuvo el valor de salida correspondiente a cada una de las combinaciones de las entradas del modelo.

Cada una de las reglas se comporta de manera dinámica, ya que es posible cambiar el valor

que determina el peso de una regla. Este peso es un valor entre 0.1 y 1. Tomando en cuenta el valor del peso de la regla, podemos una regla con un peso modificado en 0.5 de la siguiente manera:

**IF**

palabras=3 **AND** ubicación=PELIGROSO **AND** emocion=MUY\_ENOJADO ( 0.5 )

**THEN**

salida=EXTREMADAMENTE\_VIOLENTO

En la estructura y declaración de las reglas y variables de salida, los valores no cambian. Sin embargo, podemos ver el cambio en el valor de salida del modelo.

En la imagen podemos observar el comportamiento en el valor de salida al cambiar el peso de una regla de 1 a 0.3

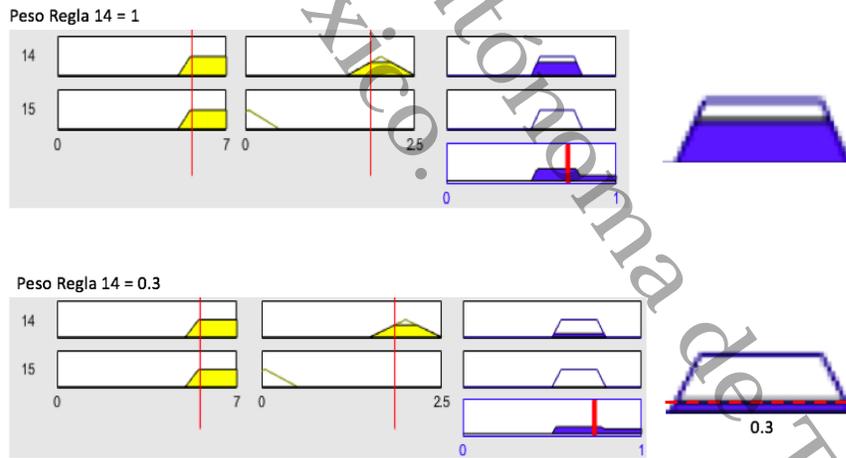


Figura 3.14: Ejemplo de cambio de peso en las reglas

al cambiar el valor, la salida cambia su comportamiento, ya que el área remarcada con azul, que simboliza el valor de salida, resultado de la combinación de todas las variables que intervienen en la regla, es significativamente menor, por lo que el impacto de la regla en el modelo no es tanto como cuando tiene peso en valor 1.

### 3.2.5. Variables de Salida

La última parada en un modelo difuso son las variables de salida. Es el resultado final después de haber hecho el análisis y ecuaciones correspondientes en el motor de inferencia utilizando las variables de entrada.

Para ésta investigación únicamente se necesitó una única variable de salida.

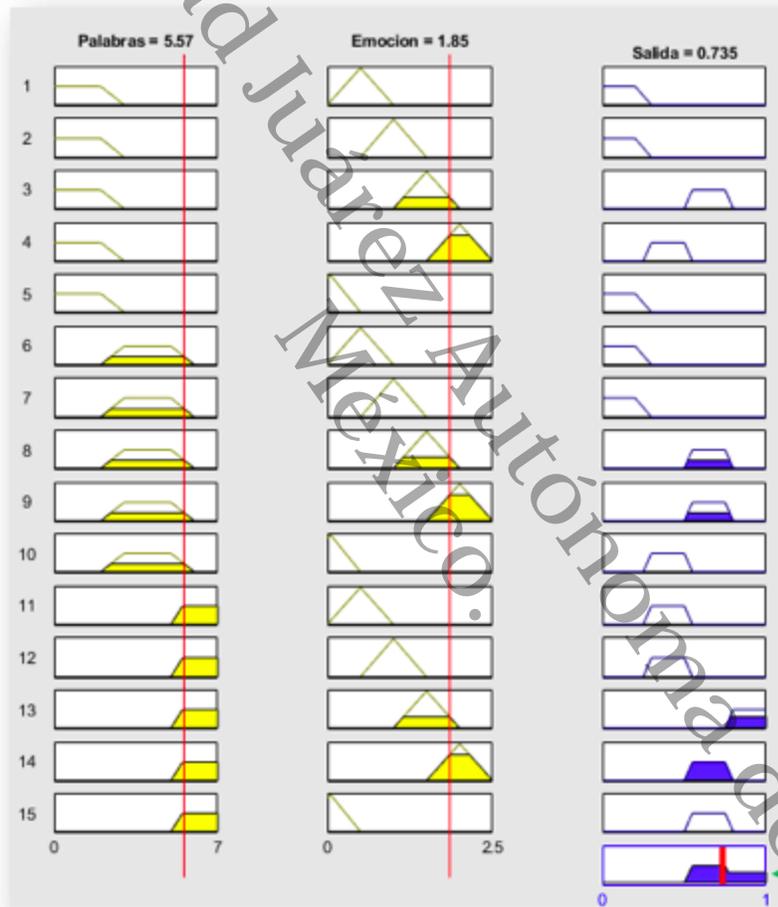


Figura 3.15: variable de salida en el modelo

Ésta variable de salida representa un valor, mismo que va cambiando mediante la parte dinámica del modelo. Es decir, si se ejecuta el modelo con los valores estáticos, el valor es diferente al que se presenta cuando los valores son dinámicos, no importa si el cambio diná-

mico es en una sola etapa del modelo, como en la entrada o las reglas; incluso en la salida. El valor siempre será variable aunque sea un valor decimal, ya que se está aplicando dinamismo al modelo en cualquiera de sus partes.

El diseño de esta variable se basó en el resultado que se quería obtener al final del análisis. Puesto que la necesidad para la investigación era saber qué tanta probabilidad o porcentaje de violencia contenía el material a analizar, únicamente se necesitó un único valor de salida, a diferencia de otros modelos que trabajan con dimensiones y se necesitan más variables de salida dependiendo de las necesidades de cada proyecto.

Pero la parte dinámica de esta variable consiste en cambiar el número de grupos de pertenencia.

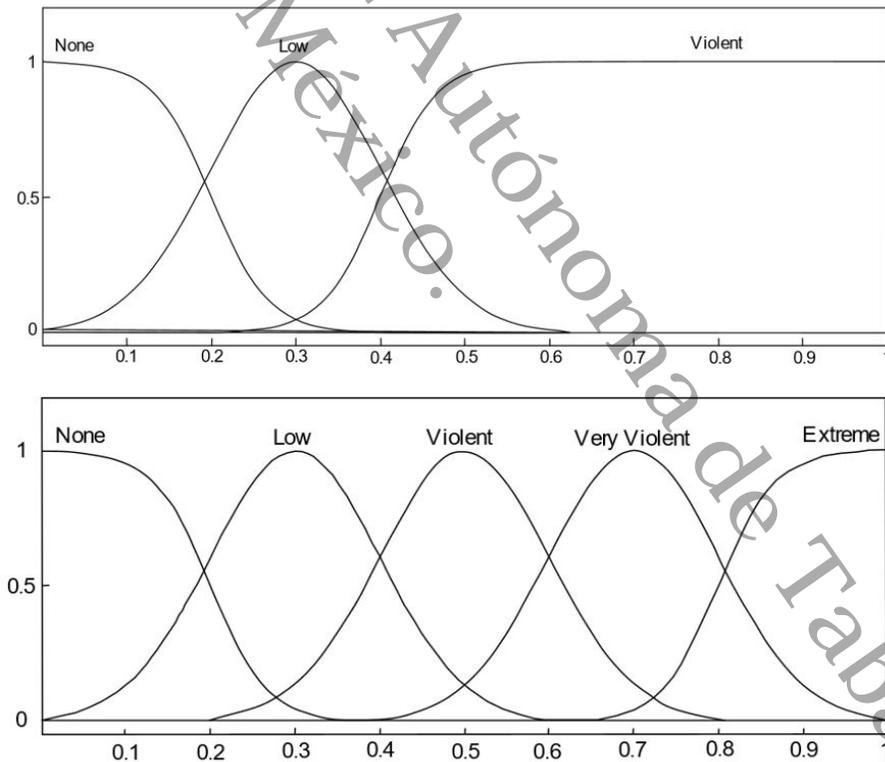


Figura 3.16: Grupos de la variable salida

En la imagen anterior, podemos comparar que existen aparentemente dos variables de

salida. Pero la realidad es que es la misma variable, con diferentes grupos de pertenencia. Esto es porque, dependiendo de la variable de contexto **Ubicación**, la variable tendrá tres o cinco grupos.

La fórmula es sencilla. Si la ubicación coincide con una zona peligrosa dentro de la ciudad, entonces se considerarán cinco grupos de pertenencia, mientras que si el usuario se encuentra dentro de una zona no tan peligrosa, entonces serán considerados tres grupos.

Tal vez parezca que el razonamiento no es lógico de momento, pero al haber más grupos, hay más posibilidades de que el valor se acerque más límite superior. Aunque hay que considerar que la variable de contexto ubicación, regula la existencia de tres o cinco grupos, dependiendo del comportamiento de la misma.

Por ejemplo, consideremos un usuario que se encuentra dentro de una zona categorizada con zona peligros dentro de la ciudad, y el modelo presenta una salida con valor de 0.7 representado en la siguiente imagen.

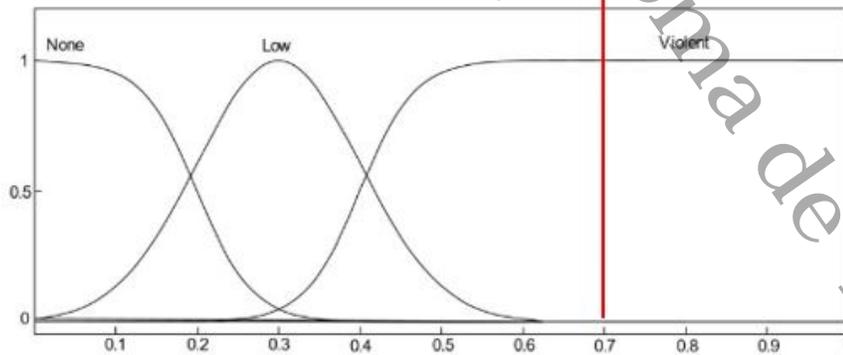


Figura 3.17: Valor de salida con tres grupos

El valor de salida es situado dentro del grupo de pertenencia **Violento**. Ahora bien, consideremos exactamente el mismo ejemplo con las mismas condiciones, mismas reglas, pero la diferencia será que en vez de tomar la variable de contexto ubicación como zona menos

peligrosa, ahora se tomará como zona muy peligrosa.

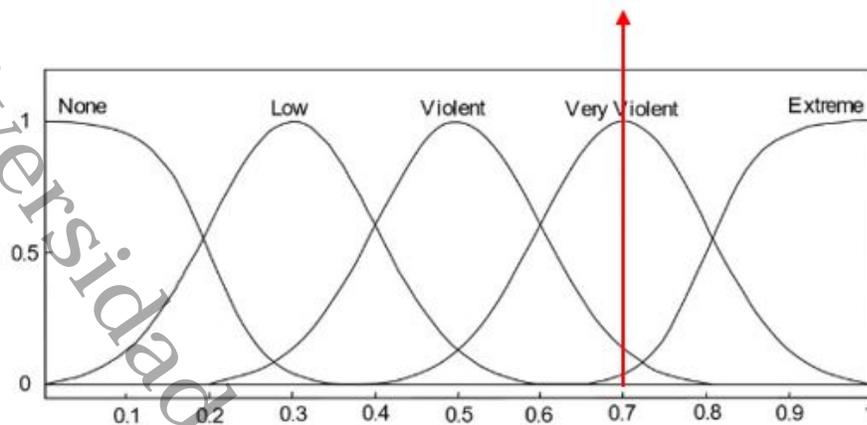


Figura 3.18: Valor de salida con cinco grupos

Como podemos observar, el valor de la variable ahora se sitúa en el grupo de pertenencia **Muy Violento**, y esto es debido a la existencia de cinco grupos, en donde existe más variedad de pertenencia, ya que el diseño de la misma variable así lo demanda, para cubrir las necesidades del modelo.

Ahora que ya se cubrió toda la metodología y diseño del modelo, es momento de la fase de pruebas y experimentación, misma que se discutirá en el siguiente capítulo.

Universidad Juárez Autónoma de Tabasco.  
México.

# Capítulo 4

## Pruebas y Resultados

### 4.1. Pruebas

Después de haber recopilado el material audiovisual que conformó el corpus de la presente investigación, fue necesario llevar a cabo pruebas y experimentos que arrojaran resultados acerca de qué tan eficaz pudiera llegar a ser el modelo a implementar.

Para el caso del modelo en discusión, se elaboró un set de datos donde se fueron registrando todos los eventos presentes en cada uno de los videos analizados. Cada video fue dividido en segmentos de 5 segundos y cada segmento fué analizado en busca de las tres variables mencionadas en secciones anteriores. Y así, de esa forma ir completando el archivo final.

La siguiente imagen muestra una tabla que es el concentrado final obtenido despues de acumular los datos de cada uno de los videos antes mencionados. Cabe mencionar que por respeto a los lectores, la columna denominada como **Texto** fué censurada, ya que los datos que originalmente se muestran contienen palabras altisonantes, de forma que esos datos fueron suprimidos para no herir las susceptibilidad de algunos lectores.

La primer columna, denominada **File**, muestra el nombre del archivo que fué analizado. Como podemos observar, el nombre se repite, ya que, como se mencionó anteriormente, cada

uno de los videos fué segmentado en fragmentos de 5 segundos, por lo que si un video dura 1 minuto, quiere decir que se segmentará en 12 ventanas de tiempo.

File	Emo	Novaco (NAS)	#	Texto	Ruta	Vinicio	Vfinal	Tiempo	Fecha	Vector	Clase
2	Enojado	3	2		25	01:19	01:24	00:05	08-oct-16	[2,25,3]	01
2	Moderadamente_Enojado	2	0		21	01:24	01:29	00:05	08-oct-16	[0,21,2]	02
2	Enojado	3	2		14	01:19	01:24	00:05	08-oct-16	[2,14,3]	01
2	Moderadamente_Enojado	2	0		9	01:24	01:29	00:05	08-oct-16	[0,9,2]	02
15	No_Enojado	-1	2		17	00:15	00:20	00:05	21-ago-18	[2,17,-1]	02
15	No_Enojado	-1	0		21	00:20	00:25	00:05	21-ago-18	[0,21,-1]	02
15	No_Enojado	-1	0		27	00:25	00:30	00:05	21-ago-18	[0,27,-1]	02
15	No_Enojado	-1	0		19	00:30	00:35	00:05	21-ago-18	[0,19,-1]	02
15	No_Enojado	-1	0		22	00:35	00:40	00:05	21-ago-18	[0,22,-1]	02
15	No_Enojado	-1	1		19	00:40	00:45	00:05	21-ago-18	[1,19,-1]	02
15	No_Enojado	-1	0		23	00:45	00:50	00:05	21-ago-18	[0,23,-1]	02
15	No_Enojado	-1	2		3	00:15	00:20	00:05	21-ago-18	[2,3,-1]	02
15	No_Enojado	-1	0		5	00:20	00:25	00:05	21-ago-18	[0,5,-1]	02
15	No_Enojado	-1	0		6	00:25	00:30	00:05	21-ago-18	[0,6,-1]	02
15	No_Enojado	-1	0		12	00:30	00:35	00:05	21-ago-18	[0,12,-1]	02
15	No_Enojado	-1	0		3	00:35	00:40	00:05	21-ago-18	[0,3,-1]	02
15	No_Enojado	-1	1		9	00:40	00:45	00:05	21-ago-18	[1,9,-1]	02
15	No_Enojado	-1	0		9	00:45	00:50	00:05	21-ago-18	[0,9,-1]	02
1	Moderadamente_Enojado	2	7		22	00:32	00:37	00:05	17-oct-18	[7,22,2]	01
1	Enojado	3	5		26	00:37	00:42	00:05	17-oct-18	[5,26,3]	01
1	Enojado	3	5		22	00:42	00:47	00:05	17-oct-18	[5,22,3]	01
1	Enojado	3	5		27	00:47	00:52	00:05	17-oct-18	[5,27,3]	01
1	Muy_Enojado	4	4		19	00:52	00:57	00:05	17-oct-18	[4,19,4]	01
1	Poco_Enojado	1	0		19	00:57	01:02	00:05	17-oct-18	[0,19,1]	02
1	Enojado	3	4		22	01:02	01:07	00:05	17-oct-18	[4,22,3]	01
1	Enojado	3	3		27	01:07	01:12	00:05	17-oct-18	[3,27,3]	01
1	Enojado	3	1		17	01:12	01:17	00:05	17-oct-18	[1,17,3]	02
1	Muy_Enojado	4	2		18	01:17	01:22	00:05	17-oct-18	[2,18,4]	01
1	Moderadamente_Enojado	2	7		8	00:32	00:37	00:05	17-oct-18	[7,8,2]	01
1	Enojado	3	5		9	00:37	00:42	00:05	17-oct-18	[5,9,3]	01
1	Enojado	3	5		6	00:42	00:47	00:05	17-oct-18	[5,6,3]	01
1	Enojado	3	5		9	00:47	00:52	00:05	17-oct-18	[5,9,3]	01
1	Muy_Enojado	4	4		3	00:52	00:57	00:05	17-oct-18	[4,3,4]	01
1	Poco_Enojado	1	0		11	00:57	01:02	00:05	17-oct-18	[0,11,1]	02

Figura 4.1: Nombre de archivo en dataset

La segunda columna, denominada **Emo**, hace referencia a la emoción que se encuentra en esa ventana de tiempo. Dado que el enfoque del proyecto era usar lógica difusa para probar el modelo, todavía no se ha hecho uso de la inteligencia artificial para extraer las emociones de manera automática mediante el análisis de sonido, sino que las pruebas se hicieron de forma manual. Es decir, se escuchó el segmento y se determinó cual era la emoción predominante en el segmento.

Así mismo, la tercer columna del dataset muestra la leyenda **Novaco (NAS)**. Esta columna muestra el valor en la escala NOVACO, una escala para determinar emociones [4] realizada a convictos en una prisión de Estados Unidos.

En dicha escala, se categorizan las emociones por medio de una simple tabla, indicando un valor numérico de acuerdo a la emoción predominante. En este caso, la emoción No enojado obtiene el valor -1 mientras que la emoción Extremadamente enojado obtiene el valor 5.

# Anger Inventory

## Novaco Anger Scale

0	-	you would feel very little or no annoyance
1	-	you would feel a little irritated
2	-	you would feel moderately upset
3	-	you would feel quite angry
4	-	you would feel very angry

Figura 4.2: Ejemplo de escala NOVACO

La columna denominada con el símbolo numeral ( $\#$ ), determina el número de palabras altisonantes encontradas en el segmento de tiempo, y son éstos los que intervienen directamente en la variable de entrada **Palabras**.

La columna con la palabra **Texto**, contiene las palabras altisonantes mencionadas, convertidas a texto. Estas palabras no se envían al modelo, simplemente son una referencia para llevar un control acerca de qué palabras fueron mencionadas por si se quisiera realizar una validación posteriormente.

La **Ruta** es la columna que registra el valor numérico en kilómetros de la ruta de la unidad de transporte público. Significa que si el valor es 25, la unidad se encuentra en el kilómetro 25 de su ruta. Cabe mencionar que cada unidad tiene sus propios valores en cuanto a las rutas, y el kilómetro 25 denominado como peligros en una ruta, no es peligroso dentro de otra ruta.

La columna **Vector** es un dato de suma importancia dentro del set de datos, ya que es el valor que será la entrada del modelo en MATLAB. Éste valor vectorial es construido a partir de los diversos valores encontrados en diversas columnas. En específico intervienen las siguientes columnas: [  $\#$ , Ruta, Novaco(NAS) ].

File	Emo	Novaco (NAS)	#	Texto	Ruta	Vinicio	Vfinal	Tiempo	Fecha	Vector	Clase
2	Enojado	3	2		25	01:19	01:24	00:05	06-oct-16	[2,25,3]	01
2	Moderadamente_Enojado	2	0		21	01:24	01:29	00:05	06-oct-16	[0,21,2]	02
2	Enojado	3	2		14	01:19	01:24	00:05	06-oct-16	[2,14,3]	01
2	Moderadamente_Enojado	2	0		9	01:24	01:29	00:05	06-oct-16	[0,9,2]	02
15	No_Enojado	-1	2		17	00:15	00:20	00:05	21-ago-18	[2,17,-1]	02
15	No_Enojado	-1	0		21	00:20	00:25	00:05	21-ago-18	[0,21,-1]	02
15	No_Enojado	-1	0		27	00:25	00:30	00:05	21-ago-18	[0,27,-1]	02
15	No_Enojado	-1	0		19	00:30	00:35	00:05	21-ago-18	[0,19,-1]	02
15	No_Enojado	-1	0		22	00:35	00:40	00:05	21-ago-18	[0,22,-1]	02
15	No_Enojado	-1	1		19	00:40	00:45	00:05	21-ago-18	[1,19,-1]	02
15	No_Enojado	-1	0		23	00:45	00:50	00:05	21-ago-18	[0,23,-1]	02
15	No_Enojado	-1	2		3	00:15	00:20	00:05	21-ago-18	[2,3,-1]	02
15	No_Enojado	-1	0		5	00:20	00:25	00:05	21-ago-18	[0,5,-1]	02
15	No_Enojado	-1	0		6	00:25	00:30	00:05	21-ago-18	[0,6,-1]	02
15	No_Enojado	-1	0		12	00:30	00:35	00:05	21-ago-18	[0,12,-1]	02
15	No_Enojado	-1	0		3	00:35	00:40	00:05	21-ago-18	[0,3,-1]	02
15	No_Enojado	-1	1		9	00:40	00:45	00:05	21-ago-18	[1,9,-1]	02
15	No_Enojado	-1	0		9	00:45	00:50	00:05	21-ago-18	[0,9,-1]	02
1	Moderadamente_Enojado	2	7		22	00:32	00:37	00:05	17-oct-18	[7,22,2]	01
1	Enojado	3	5		26	00:37	00:42	00:05	17-oct-18	[5,26,3]	01
1	Enojado	3	5		22	00:42	00:47	00:05	17-oct-18	[5,22,3]	01
1	Enojado	3	5		27	00:47	00:52	00:05	17-oct-18	[5,27,3]	01
1	Muy_Enojado	4	4		19	00:52	00:57	00:05	17-oct-18	[4,19,4]	01
1	Poco_Enojado	1	0		19	00:57	01:02	00:05	17-oct-18	[0,19,1]	02
1	Enojado	3	4		22	01:02	01:07	00:05	17-oct-18	[4,22,3]	01
1	Enojado	3	3		27	01:07	01:12	00:05	17-oct-18	[3,27,3]	01
1	Enojado	3	1		17	01:12	01:17	00:05	17-oct-18	[1,17,3]	02
1	Muy_Enojado	4	2		18	01:17	01:22	00:05	17-oct-18	[2,18,4]	01
1	Moderadamente_Enojado	2	7		8	00:32	00:37	00:05	17-oct-18	[7,8,2]	01
1	Enojado	3	5		9	00:37	00:42	00:05	17-oct-18	[5,9,3]	01
1	Enojado	3	5		6	00:42	00:47	00:05	17-oct-18	[5,6,3]	01
1	Enojado	3	5		9	00:47	00:52	00:05	17-oct-18	[5,9,3]	01
1	Muy_Enojado	4	4		3	00:52	00:57	00:05	17-oct-18	[4,3,4]	01
1	Poco_Enojado	1	0		11	00:57	01:02	00:05	17-oct-18	[0,11,1]	02

Figura 4.3: Vector construido a partir de los datos (color amarillo)

Una vez los datos estuvieron completos, se obtuvo un total de 383 registros, los cuales se dividieron en dos partes, para formar un 70% para entrenamiento (268 registros) y un 30% para pruebas (115 registros).

Esta división se hizo siguiendo las prácticas de otros autores, ya que es necesario tomar todo el conjunto de datos y reservar una parte para realizar los experimentos, de manera que los datos reservados para entrenamiento no sean los mismos donde se hagan las pruebas, porque podría haber cabida a error, ya que un mismo dato usado para entrenar dará siempre positivo en etapa de pruebas.

Para el entrenamiento, se fue probando fila por fila en el modelo difuso, registrando el valor de salida obtenido, comparándolo con la clase seleccionada para validar su coincidencia. En caso de no coincidir con la clase, se reajustó la regla involucrada para que la clase coincidiera, y de esa manera se iba entrenando el modelo poco a poco hasta terminar de introducir todos los valores y obtener un modelo completamente entrenado.

VECTOR	SALIDA SIGMOIDE		5 GRUPOS	3 GRUPOS
	5 GRUPOS	3 GRUPOS		
[2;14;20.6;1.5]	0.3	0.3	POCO	POCO
[3;14;20.6;0.1]	0.301	0.301	POCO	POCO
[2;14;20.6;0.1]	0.127	0.127	NADA	NADA
[1;22;20.6;0.1]	0.3	0.3	POCO	POCO
[2;22;20.6;0.1]	0.5	0.677	VIOLENTO	VIOLENTO
[2;22;20.6;1.5]	0.7	0.677	VIOLENTO	VIOLENTO
[1;5;20.6;0.1]	0.117	0.117	NADA	NADA
[3;5;20.6;0.1]	0.7	0.677	VIOLENTO	VIOLENTO
[8;5;20.6;0.1]	0.7	0.677	VIOLENTO	VIOLENTO
[2;5;20.6;0.1]	0.3	0.3	POCO	POCO
[2;19;20.6;0.5]	0.5	0.677	VIOLENTO	VIOLENTO
[4;19;20.6;1.5]	0.744	0.677	VIOLENTO	VIOLENTO
[3;19;20.6;0.1]	0.7	0.677	VIOLENTO	VIOLENTO
[3;19;20.6;1.5]	0.744	0.677	VIOLENTO	VIOLENTO

Figura 4.4: Muestra de modelo entrenado a partir de vectores

Una vez terminado el entrenamiento, la parte restante que fué reservada para realizar las pruebas, se introdujo al modelo y se fué capturando cada una de las salidas del modelo, aplicando todas las combinaciones posibles, con los grupos dinámicos, y observando las matrices de confusión para corroborar la veracidad de los mismos.

## 4.2. Resultados

Una vez registrados todos los datos, se procedió a crear las matrices de confusión con los resultados obtenidos, los cuales fueron:

Tabla 4.1: Resultados del modelo

Salida	Amplitud	Efectividad
3 grupos	0.3	86.21 %
	0.6	<b>90.73 %</b>
	1	85.42 %
5 grupos	0.3	<b>94.23 %</b>
	0.6	93.13 %
	1	85.42 %

Si observamos la tabla anterior, podemos ver que los resultados con mayor efectividad son aquellos cuyas configuraciones corresponden a la salida con tres grupos y amplitud media, mientras que para la salida de cinco grupos, los resultados más efectivos se dan con amplitud en nivel bajo.

cabe mencionar que los resultados efectivamente presentarán variaciones dependiendo del escenario. Eso es porque si se implementa el modelo en un ambiente controlado, los resultados probablemente serán más efectivos en cada una de sus configuraciones. Adicionalmente, si el escenario presenta menor ruido, la voces se aprecian mejor y con mayor calidad, y tanto las palabras como las emociones son claramente identificadas, los resultados serán excepcionales.

A continuación se presentan los resultados correspondientes de acuerdo a tres grupos, con todas sus pruebas en variaciones en amplitud (Para las siguientes tablas, FN=Falsos Negativos y FP=Falsos Positivos).

Tabla 4.2: Resultados para 3 grupos con amplitud = 0.3

	<b>Positivos</b>	<b>Negativos</b>	<b>Efectividad</b>
<b>Positivos</b>	85	15 (FN)	86.21 %
<b>Negativos</b>	5 (FP)	60	

Tabla 4.3: Resultados para 3 grupos con amplitud = 0.6

	<b>Positivos</b>	<b>Negativos</b>	<b>Efectividad</b>
<b>Positivos</b>	85	14 (FN)	90.73 %
<b>Negativos</b>	0 (FP)	66	

Tabla 4.4: Resultados para 3 grupos con amplitud = 1

	<b>Positivos</b>	<b>Negativos</b>	<b>Efectividad</b>
<b>Positivos</b>	74	15 (FN)	85.42 %
<b>Negativos</b>	6 (FP)	70	

A continuación, se presentan las tablas de efectividad para las pruebas con tres grupos y las variaciones de todas sus amplitudes:

Tabla 4.5: Resultados de efectividad para la tabla 4.2

<b>Correctos</b>	145
<b>Incorrectos</b>	20
<b>Efectividad</b>	86.21 %

Tabla 4.6: Resultados de efectividad para la tabla 4.3

<b>Correctos</b>	151
<b>Incorrectos</b>	14
<b>Efectividad</b>	90.73 %

Tabla 4.7: Resultados de efectividad para la tabla 4.4

<b>Correctos</b>	144
<b>Incorrectos</b>	21
<b>Efectividad</b>	85.42 %

Éstos son los resultados correspondientes de acuerdo a cinco grupos, con todas las pruebas en variaciones de amplitud. Cabe mencionar que las pruebas fueron realizadas en igualdad de condiciones, únicamente cambiando la variación de grupos. Esto con la finalidad de que las pruebas dieran los resultados más exactos posibles, sin tener que reentrenar el modelo ni hacer adaptaciones a las reglas (Para las siguientes tablas, FN=Falsos Negativos y FP=Falsos Positivos):

Tabla 4.8: Resultados para 5 grupos con amplitud = 0.3

	<b>Positivos</b>	<b>Negativos</b>	<b>Efectividad</b>
<b>Positivos</b>	85	4 (FN)	94.23 %
<b>Negativos</b>	5 (FP)	71	

Tabla 4.9: Resultados para 5 grupos con amplitud = 0.6

	Positivos	Negativos	Efectividad
Positivos	87	4 (FN)	93.13 %
Negativos	7 (FP)	73	

Tabla 4.10: Resultados para 5 grupos con amplitud = 1

	Positivos	Negativos	Efectividad
Positivos	74	16 (FN)	85.42 %
Negativos	5 (FP)	70	

A continuación, ya que los resultados de eficiencia para tres grupos fueron presentados, estas son las tablas de eficiencia para las pruebas con cinco grupos, con todas las variantes de amplitud:

Tabla 4.11: Resultados de efectividad para la tabla 4.8

<b>Correctos</b>	156
<b>Incorrectos</b>	9
<b>Efectividad</b>	94.23 %

Tabla 4.12: Resultados de efectividad para la tabla 4.9

<b>Correctos</b>	160
<b>Incorrectos</b>	11
<b>Efectividad</b>	93.13 %

Tabla 4.13: Resultados de efectividad para la tabla 4.10

<b>Correctos</b>	144
<b>Incorrectos</b>	21
<b>Efectividad</b>	85.42 %

## Capítulo 5

### Conclusión

A lo largo de este trabajo hemos afrontado un gran reto, ya que hay muy pocos trabajos realizados en los cuales se haya empleado un sistema difuso que sea dinámico y, además, que tome en cuenta el contexto para ajustar los valores acordes a las necesidades del usuario. Y a pesar de que ya existen aportes al respecto, ninguno de ellos es completamente dinámico, como lo es el objetivo de nuestro trabajo.

Esta investigación forma parte de un proyecto que estamos desarrollando actualmente. Nos encontramos en la primera de tres partes, ya que nuestra intención es ofrecer un modelo que se pueda conectar a diversos dispositivos, especialmente cámaras de seguridad ubicadas en unidades de transporte público de las ciudades con mayor incidencia delictiva del país. Pretendemos que este dispositivo sea desarrollado en Arduino para efectuar las primeras pruebas de manera real, para posteriormente desarrollar un dispositivo con diseño propio, pero antes se debe terminar el proyecto por completo.

El proyecto final consta de tres etapas:

- **Modelo Difuso Dinámico** (La presente investigación).
- **Procesamiento del Lenguaje Natural**  
(pretendemos hacer uso de inteligencia artificial para analizar palabras y transcribirlas

en texto para detectar violencia verbal).

- **Análisis de Sentimientos**

(con el uso de inteligencia artificial trataremos de analizar audio para detectar emociones negativas tales como enojo o furia.).

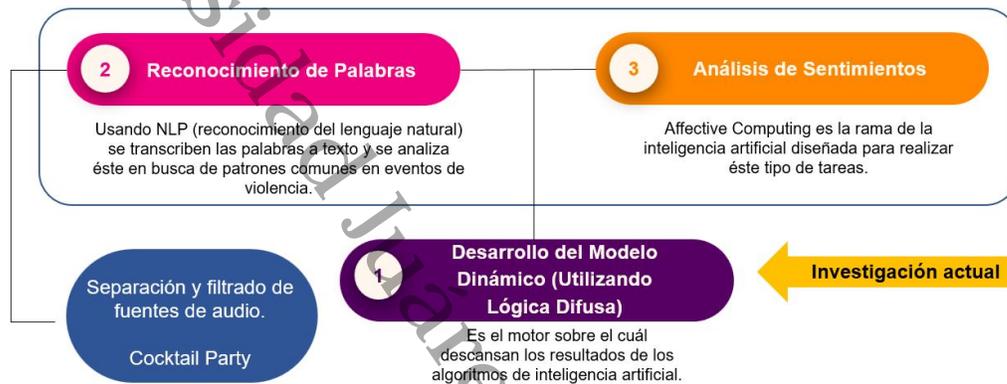


Figura 5.1: Proyecto general en el cual se está trabajando

Con nuestro aporte pretendemos ayudar a los organismos de seguridad pública y de esa forma disminuir el porcentaje delictivo del país, para que las personas y familias puedan tener una vida con mayor seguridad y dejen de ser víctimas de crímenes y asaltos. El INEGI ha liberado un documento en donde se muestran estadísticas acerca de la percepción de inseguridad en cada uno de los estados de la república.

En este reporte se muestra, a manera de porcentaje, que tan seguras se sienten las personas en su casa, calle, transporte público, lugar de trabajo y en la colonia en donde viven. Este trabajo surge como necesidad a combatir la violencia que se vive en casi todo el territorio mexicano y que es un problema que muchas personas y organizaciones públicas han querido resolver, pero nadie ha podido hasta el momento.

Con nuestro aporte no pretendemos solucionar el problema, ya que para eso necesitamos la colaboración de muchas organizaciones y elementos de seguridad. Nuestro principal interés es aportar una herramienta que ayude a los sistemas de seguridad en detectar este tipo de

situaciones y alertar a las autoridades de manera automática y de esa forma ellos evalúen la situación y determinen que acciones tomar.

Este trabajo es un gran aporte ya que hemos obtenido muy buenos resultados, a pesar de que no se ha llegado a implementar al cien por ciento. Pero de acuerdo al modelo que hemos diseñado y a los datos reales que hemos logrado recopilar gracias a las publicaciones que se han hecho por medio de redes sociales y plataformas confiables como YouTube, los buenos resultados son posibles.

Como ya mencionamos anteriormente, este trabajo forma parte de un proyecto que tiene tres etapas. Pero relativo al proyecto actual, sabemos que podemos obtener grandes resultados en la implementación. El modelo dinámico que proponemos es un modelo novedoso ya que implementa la parte dinámica en cada una de sus partes y con eso obtenemos un modelo versátil que se adapte a los cambios variables de las necesidades de un escenario.

Dados los buenos resultados que obtuvimos y las variables del modelo diseñadas de acuerdo a datos de asaltos reales, podemos decir que el modelo es confiable y dará resultados acertados cuando el proyecto esté terminado. El dinamismo del modelo es un aporte a la ciencia que a nuestro conocimiento no se ha hecho todavía. Sin embargo, todavía falta desarrollar las etapas restantes para que el aporte sea completo.

---

Universidad Juárez Autónoma de Tabasco.  
México.

# Referencias

- [1] Acar E., Hopfgartner F., Albayrak S.: Detecting violent content in Hollywood movies by mid-level audio representations. In: *2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 73–78. IEEE, 2013.
- [2] Agrawal R., Mannila H., Srikant R., Toivonen H., Verkamo A.I., et al.: Fast discovery of association rules. In: *Advances in knowledge discovery and data mining*, vol. 12(1), pp. 307–328, 1996.
- [3] Arango A.: Indicadores de seguridad pública en México: la construcción de un sistema de estadísticas delictivas. In: *Federacion*, 2003.
- [4] Baker M.T., Van Hasselt V.B., Sellers A.H.: Validation of the Novaco Anger Scale in an incarcerated offender population. In: *Criminal Justice and behavior*, vol. 35(6), pp. 741–754, 2008.
- [5] Bandhakavi A., Wiratunga N., Massie S., Padmanabhan D.: Lexicon Generation for Emotion Detection from Text. In: *IEEE Intelligent Systems*, vol. 32(01), pp. 102–108, 2017. ISSN 1941-1294. URL <http://dx.doi.org/10.1109/MIS.2017.22>.
- [6] Barlow H.B.: Unsupervised learning. In: *Neural computation*, vol. 1(3), pp. 295–311, 1989.
- [7] Bautista-Durán M., García-Gómez J., Gil-Pita R., Mohino-Herranz I., Rosa-Zurera M.: Energy-Efficient Acoustic Violence Detector for Smart Cities. In: *International Journal of Computational Intelligence Systems*, vol. 10(1), pp. 1298–1305, 2017.

- [8] Bautista-Durán M., García-Gómez J., Gil-Pita R., Sánchez-Hevia H.A., Mohino-Herranz I., Rosa-Zurera M.: Acoustic Detection of Violence in Real and Fictional Environments. In: *ICPRAM*, pp. 456–462. 2017.
- [9] Bhanuabhiram K., Tharun Kumar L., Srinivasan N.: Google Assistant Controlled Home Automation. In: *Journal of Computational and Theoretical Nanoscience*, vol. 16(8), pp. 3259–3264, 2019.
- [10] Bhaskar J., Sruthi K., Nedungadi P.: Hybrid approach for emotion classification of audio conversation based on text and speech mining. In: *Procedia Computer Science*, vol. 46, pp. 635–643, 2015.
- [11] Boden M.: Artificial intelligence and natural man. In: *Elsevier*, 1980.
- [12] Bugueño Sáez V.G.: *Modelo de detección de agresiones verbales, por medio de algoritmos de Machine Learning*. Universidad de Chile, 2017.
- [13] Bustamante P., López N., Pérez M., Quintero O.: Reconocimiento y Regionalización de las Emociones en el Plano Excitación-Valencia. In: *XX Congreso Argentino de Bioingeniería*. 2015. URL <http://dx.doi.org/10.13140/RG.2.1.3188.6164>.
- [14] Cambria E.: Affective computing and sentiment analysis. In: *IEEE Intelligent Systems*, vol. 31(2), pp. 102–107, 2016.
- [15] Cambria E., Olsher D., Rajagopal D.: SenticNet 3: A Common and Common-sense Knowledge Base for Cognition-driven Sentiment Analysis. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, pp. 1515–1521. AAAI Press, 2014. URL <http://dl.acm.org/citation.cfm?id=2892753.2892763>.
- [16] Caruana R.: Multitask learning. In: *Machine learning*, vol. 28(1), pp. 41–75, 1997.
- [17] Chapelle O., Scholkopf B., Zien A.: Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. In: *IEEE Transactions on Neural Networks*, vol. 20(3), pp. 542–542, 2009.

- [18] Close J.G., Licona M.F.M.: Reconocimiento de Emociones usando Técnicas de Aprendizaje Maquinal. In: *Universidad Nacional Autónoma de México*, 2012.
- [19] Dascano M.: *Samsung Galaxy Home: An Easy Guide to Learning the Baics*. Independently published, 2018. ISBN 1728677394, 9781728677392.
- [20] Dorogy Y., Kolisnichenko V., Levchenko K.: Violent Crime Detection System. In: *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, vol. 1, pp. 352–355. IEEE, 2018.
- [21] Esuli A., Sebastiani F.: SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 06)*. European Language Resources Association (ELRA), Genoa, Italy, 2006. URL [http://www.lrec-conf.org/proceedings/lrec2006/pdf/384\\_pdf.pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/384_pdf.pdf).
- [22] Eyben F., Schuller B.: openSMILE): the Munich open-source large-scale multimedia feature extractor. In: *ACM SIGMultimedia Records*, vol. 6(4), pp. 4–13, 2015.
- [23] Eyben F., Wöllmer M., Schuller B.: OpenEAR—introducing the Munich open-source emotion and affect recognition toolkit. In: *2009 3rd international conference on affective computing and intelligent interaction and workshops*, pp. 1–6. IEEE, 2009.
- [24] Federal D.: *Código Penal Federal*. Ediciones Lazcano Lozano, 2001.
- [25] Federal D., Fuentevilla J.G.: *Código Civil Federal*. Ediciones Lazcano Lozano, 2001.
- [26] Ferdinando H., Ye L., Han T., Zhang Z., Sun G., Huuki T., Seppänen T., Alasaarela E.: Violence detection from ECG signals: a preliminary study. In: *J. Pattern Recognit. Res.*, vol. 12(1), pp. 7–18, 2017.
- [27] Flórez D.A.: Botones de Pánico instalados en Medellín Colombia, 2016. URL <https://publications.iadb.org/publications/spanish/document/Estudios-de-casos-internacionales-de-ciudades-inteligentes-Medellin-Colombia.pdf>. Last accessed 20 December 2021.

- [28] Friedman N., Nachman I., Peér D.: Learning bayesian network structure from massive datasets: the «sparse candidate «algorithm. In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 206–215. Morgan Kaufmann Publishers Inc., 1999.
- [29] de Estadística y Geografía I.N.: Encuesta Nacional del Seguridad Pública Urbana (ENSU), 2021. URL [https://inegi.org.mx/contenidos/programas/ensu/doc/ensu2021\\_diciembre\\_presentacion\\_ejecutiva.pdf](https://inegi.org.mx/contenidos/programas/ensu/doc/ensu2021_diciembre_presentacion_ejecutiva.pdf). Last accessed 03 February 2022.
- [30] de Estadística y Geografía (INEGI) I.N.: Tasa de incidencia delictiva por entidad federativa de ocurrencia por cada cien mil habitantes. <https://www.inegi.org.mx/temas/incidencia/>, 2017.
- [31] Giannakopoulos T.: Study and application of acoustic information for the detection of harmful content, and fusion with visual information. In: *Department of Informatics and Telecommunications, vol. PhD. University of Athens, Greece*, 2009.
- [32] Giannakopoulos T., Kosmopoulos D., Aristidou A., Theodoridis S.: Violence content classification using audio features. In: *Hellenic Conference on Artificial Intelligence*, pp. 502–507. Springer, 2006.
- [33] Goodwill A.M., Stephens S., Oziel S., Yapp J., Bowes N.: Multidimensional latent classification of ‘street robbery’ offences. In: *Journal of Investigative Psychology and Offender Profiling*, vol. 9(1), pp. 93–109, 2012.
- [34] Hardt M., Price E., Srebro N., et al.: Equality of opportunity in supervised learning. In: *Advances in neural information processing systems*, pp. 3315–3323. 2016.
- [35] Jain A.K., Mao J., Mohiuddin K.M.: Artificial neural networks: A tutorial. In: *Computer*, vol. 29(3), pp. 31–44, 1996.
- [36] Jang H.J., Yang Y., Park J.S., Kim B.: FP-Growth Algorithm for Discovering Region-Based Association Rule in the IoT Environment. In: *Electronics*, vol. 10(24), p. 3091, 2021.

- [37] Kalyanathaya K.P., Akila D., Rajesh P.: Advances in Natural Language Processing—A Survey of Current Research Trends, Development Tools and Industry Applications. In: *International Journal of Recent Technology and Engineering (IJRTE)*, 2019.
- [38] Kang S., Jung J.: Mobile communication for human needs: A comparison of smartphone use between the US and Korea. In: *Computers in Human Behavior*, vol. 35, pp. 376–387, 2014.
- [39] Kaushik L., Sangwan A., Hansen J.H.: Sentiment extraction from natural audio streams. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8485–8489. IEEE, 2013.
- [40] Lin C.F., Wang S.D.: Fuzzy support vector machines. In: *IEEE transactions on neural networks*, vol. 13(2), pp. 464–471, 2002.
- [41] Lu L., Cai R., Hanjalic A.: Towards a unified framework for content-based audio analysis. In: *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 2, pp. ii–1069. IEEE, 2005.
- [42] del Estado Libre y Soberano de México G.: Periodico Oficial, 2018. URL <http://legislacion.edomex.gob.mx/sites/legislacion.edomex.gob.mx/files/files/pdf/gct/2018/mar204.pdf>. Last accessed 20 December 2021.
- [43] Millares C.: ¡Cuidado! Estas son las carreteras más inseguras de México, 2021. URL <https://motormania.com.mx/noticias/cuidado-estas-son-las-carreteras-mas-inseguras-de-mexico/>. Last accessed 19 February 2022.
- [44] Miller P., Liu W., Fowler C., Zhou H., Shen J., Ma J., Zhang J., Yan W., McLaughlin K., Sezer S.: Intelligent Sensor Information System For Public Transport—To Safely Go.... In: *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 533–538. IEEE, 2010.

- [45] Mills J.F., Kroner D.G., Forth A.E.: Novaco Anger Scale: Reliability and validity within an adult criminal sample. In: *Assessment*, vol. 5(3), pp. 237–248, 1998.
- [46] Mirsamadi S., Barsoum E., Zhang C.: Automatic speech emotion recognition using recurrent neural networks with local attention. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2227–2231. IEEE, 2017.
- [47] Ortego Resa C., et al.: *Detección de emociones en voz espontánea*. B.S. thesis, Universidad Autónoma de Madrid, 2009.
- [48] Ortony A., Clore G.L., Collins A.: *Frontmatter*, pp. i–vi. Cambridge University Press, 1988.
- [49] Palacios Alonso D.: *Contribución al estudio de selección de parámetros para identificación de estrés en la voz*. Ph.D. thesis, ETSI\_Informatica, 2017.
- [50] Patwardhan A., Knapp G.: Aggressive actions and anger detection from multiple modalities using Kinect. In: *arXiv preprint arXiv:1607.01076*, 2016.
- [51] Penet C., Demarty C.H., Gravier G., Gros P.: Audio event detection in movies using multiple audio words and contextual Bayesian networks. In: *2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 17–22. IEEE, 2013.
- [52] Pokorny F.B., Graf F., Pernkopf F., Schuller B.W.: Detection of negative emotions in speech signals using bags-of-audio-words. In: *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 879–884. IEEE, 2015.
- [53] Política C., LA DE Q.R., FEBRERO D.: Constitución Política de los Estados Unidos Mexicanos. In: *Extraído de <http://www.diputados.gob.mx/LeyesBiblio/pdf/1.pdf> Última reforma publicada en el DOF el*, vol. 9, 1917.
- [54] Putri S.N., Saputro D.: Construction fuzzy logic with curve shoulder in inference system mamdani. In: *Journal of Physics: Conference Series*, vol. 1776, p. 012060. IOP Publishing, 2021.

- [55] Quinones L.M.: Sexual harassment in public transport in Bogotá. In: *Transportation Research Part A: Policy and Practice*, vol. 139, pp. 54–69, 2020.
- [56] Radhakrishnan R., Divakaran A., Smaragdis A.: Audio analysis for surveillance applications. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pp. 158–161. IEEE, 2005.
- [57] Rokach L., Maimon O.Z.: *Data mining with decision trees: theory and applications*, vol. 69. World scientific, 2008.
- [58] Rouas J.L., Louradour J., Ambellouis S.: Audio events detection in public transport vehicle. In: *2006 IEEE Intelligent Transportation Systems Conference*, pp. 733–738. IEEE, 2006.
- [59] Russell S.J., Norvig P.: *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.
- [60] Shulby C., Pombal L., Jordão V., Ziolle G., Martho B., Postal A., Prochnow T.: Proactive Security: Embedded AI Solution for Violent and Abusive Speech Recognition. In: *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 248–253. IEEE, 2018.
- [61] Siantikos G., Sgouropoulos D., Giannakopoulos T., Spyrou E.: Fusing multiple audio sensors for acoustic event detection. In: *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 265–269. IEEE, 2015.
- [62] Van Ments M.: *The effective use of role-play: Practical techniques for improving learning*. Kogan Page Publishers, 1999.
- [63] VIOLANTE M., SINI J., MARCEDDU A.C.: Automatic Recognition And Classification Of Passengers' Emotions In Autonomous Driving Vehicles. In: *Springer*, 2019.
- [64] Wiebe J., Wilson T., Cardie C.: Annotating Expressions of Opinions and Emotions in Language. In: *Language Resources and Evaluation*, vol. 39(2), pp. 165–210, 2005. ISSN 1572-0218. URL <http://dx.doi.org/10.1007/s10579-005-7880-9>.

- 
- [65] Zhuang X., Zhou X., Hasegawa-Johnson M.A., Huang T.S.: Real-world acoustic event detection. In: *Pattern Recognition Letters*, vol. 31(12), pp. 1543–1551, 2010.
- [66] Zieger C., Brutti A., Svaizer P.: Acoustic based surveillance system for intrusion detection. In: *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 314–319. IEEE, 2009.

Universidad Juárez Autónoma de Tabasco.  
México.