

**ESTADO DEL ARTE DE LA APLICACIÓN DEL APRENDIZAJE DE MÁQUINA
NO SUPERVISADO EN EL ANÁLISIS DE PATRONES DE
COMPORTAMIENTO QUE PERMITAN IDENTIFICAR VARIABLES DE
INCIDENCIA EN FENÓMENOS SOCIALES**

**JUAN ANDRÉS GARCÍA MORENO
JUAN JOSÉ LONDOÑO RÍOS**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
PEREIRA
2020**

**ESTADO DEL ARTE DE LA APLICACIÓN DEL APRENDIZAJE DE MÁQUINA
NO SUPERVISADO EN EL ANÁLISIS DE PATRONES DE
COMPORTAMIENTO QUE PERMITAN IDENTIFICAR VARIABLES DE
INCIDENCIA EN FENÓMENOS SOCIALES**

**JUAN ANDRÉS GARCÍA MORENO
JUAN JOSÉ LONDOÑO RÍOS**

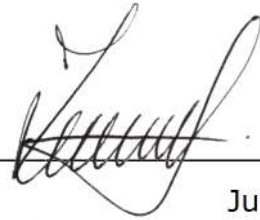
**Monografía presentada como requisito para optar al Título de
Ingeniero de Sistemas y Computación**

**Directora del Trabajo de Grado
Ingeniera Ivonne Castaño Osorio**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
PEREIRA
2020**

NOTA DE ACEPTACIÓN

APROBADO



Jurado

DEDICATORIA

Juan Andrés García Moreno

Quiero agradecer principalmente a mis padres, ustedes han sido mi motor para seguir adelante y no renunciar, gracias mil gracias por siempre creer en mi a pesar de mis errores, este proyecto es suyo, yo solo soy el interlocutor.

A mi familia por ser el apoyo que toda persona necesita para lograr sus metas, gracias por siempre estar ahí.

A mi esposa por su amor incondicional, su guía, su apoyo a cada instante, por enseñarme que nunca hay un imposible y que nunca es tarde para lograr un sueño. Gracias por compartir lo más importante en tu vida, los pequeños dragones. Eres lo mejor que me ha pasado, y lo mejor que me está pasando

A la ingeniera Ivonne quien, con sus enseñanzas, su guía, el amor y pasión por su profesión, hizo que la culminación de este proyecto fuera posible.

A mi hija Juanita por siempre estar ahí y nunca dejarme solo, sé que siempre contaré contigo al igual que siempre voy a estar para ti.

A mis amigos Julián y Felipe, los hermanos de la vida que Dios me dio, siempre he contado con su apoyo incondicional, a pesar de la distancia siempre están en mi corazón. Principalmente agradezco a Dios por colocarme tanto amor y apoyo, infinitas gracias te doy por todo lo que me has dado en mi vida.

Juan José Londoño Ríos

Dedicó este trabajo de grado a mis padres, a mis abuelos y a mi novia, personas que a lo largo de mi vida han depositado en mí grandes enseñanzas que hoy en día pongo en práctica con la esperanza de siempre crecer como persona. Así mismo quiero expresar total agradecimiento a la Profesora Ivonne Castaño quien me ha acompañado en este proceso de manera continua y extraordinaria.

CONTENIDO

INTRODUCCIÓN	7
1. PLANTEAMIENTO DEL PROBLEMA	8
2. OBJETIVOS	9
2.1. OBJETIVO GENERAL	9
2.2. OBJETIVOS ESPECÍFICOS.....	9
3. JUSTIFICACIÓN.....	10
4. METODOLOGÍA.....	11
5. TÉCNICAS DE APRENDIZAJE DE MÁQUINA NO SUPERVISADO MÁS UTILIZADAS EN EL ANÁLISIS DE PATRONES DE COMPORTAMIENTO EN FENÓMENOS SOCIALES	12
6. APLICACIÓN DE LAS TÉCNICAS DE APRENDIZAJE DE MÁQUINA NO SUPERVISADO	15
6.1. TÉCNICA DE AGRUPAMIENTO O CLÚSTER.....	15
6.1.1. CONDICIÓN DE PARADA	16
6.1.2. RESULTADOS.....	19
6.1.3. RESTRICCIÓN DE LA TÉCNICA DE AGRUPAMIENTO O CLÚSTER	19
6.1.4. CLASIFICADORES BASADOS EN REGLAS	20
6.1.5. ARQUITECTURA DEL CLASIFICADOR.....	20
6.2. TÉCNICA DE SEGUIMIENTO (TRACKING).....	21
6.2.1. CLASIFICACIÓN DE COMPORTAMIENTO NO SUPERVISADO.....	22

6.2.2. EXTRACCIÓN DE COMPONENTES INDEPENDIENTES.....	23
6.2.2.1. MODELOS DE MEZCLA GAUSSIANA	25
7. ANÁLISIS COMPARATIVO DE LOS RESULTADOS OBTENIDOS CON LA APLICACIÓN DE LAS TÉCNICAS	26
8. CONCLUSIONES	28
BIBLIOGRAFÍA	30

INTRODUCCIÓN

Esta monografía pretende dar una visión amplia y completa acerca de la utilización del aprendizaje de máquina en contextos de tipo social, específicamente, el aprendizaje no supervisado, que es sin duda el que mejor se adapta a problemas de este tipo, pues el comportamiento de las comunidades o entes sociales cambia de acuerdo con las condiciones en las que se ve inmerso, por tanto, es difícil esperar un comportamiento específico que permita entrenar una máquina a partir de este.

Esta monografía se realizará como parte del marco teórico de la tesis de maestría "Construcción de un prototipo basado en un modelo analítico para el estudio de la violencia en poblaciones urbanas del departamento de Risaralda, utilizando Machine Learning", que propone utilizar el aprendizaje no supervisado para analizar las relaciones entre los diferentes factores o variables presentes en una población, permitiendo detectar y caracterizar variables de incidencia en la aparición del fenómeno de la violencia que no hayan sido identificadas en estudios previos sobre el tema, y facilitar encontrar sus patrones de comportamiento.

1. PLANTEAMIENTO DEL PROBLEMA

En el campo de la Ciencia de los Datos, el Aprendizaje Automático o Machine Learning es una disciplina científica del ámbito de la Inteligencia Artificial, que crea sistemas que aprenden automáticamente, es decir, que identifican patrones complejos de comportamiento entre millones de datos, y facilitan el análisis de las relaciones presentes entre diferentes conjuntos de variables objeto de estudio.

Dentro de esta disciplina se habla de dos tipos de aprendizajes: el supervisado, el cual, a partir de un resultado o patrón conocido, entrena la máquina para que sea capaz de reconocer ese mismo resultado o alguno muy similar entre un conjunto de datos que suele ser muy amplio. Y está el aprendizaje no supervisado, que busca reconocer patrones sin conocimiento previo de lo que se busca, de tal manera que se puedan detectar patrones que no se hayan previsto hasta el momento.

Sobre la base de las consideraciones anteriores surge la siguiente pregunta: ¿De qué manera el Machine Learning ha facilitado el análisis de patrones de comportamiento que permitan identificar la existencia de variables de incidencia en fenómenos sociales?

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Describir el estado del arte de la aplicación del Aprendizaje de Máquina No Supervisado en el análisis de patrones de comportamiento que permitan identificar variables de incidencia en fenómenos sociales.

2.2. OBJETIVOS ESPECÍFICOS

- Recopilar información acerca de la aplicación del Aprendizaje de Máquina en el análisis de patrones de comportamiento en fenómenos sociales.
- Identificar las técnicas de aprendizaje no supervisado utilizadas.
- Describir la manera en que son aplicadas las técnicas identificadas.
- Comparar los resultados obtenidos con las técnicas descritas.
- Concluir en qué casos es mejor la utilización de qué técnicas.

3. JUSTIFICACIÓN

La presente monografía se desarrolla con el propósito de dar claridad acerca de las técnicas de Machine Learning, específicamente del aprendizaje no supervisado, que pueden ser utilizadas en el estudio de los comportamientos de los grupos sociales, y de esta manera facilitar el trabajo de los científicos de datos que buscan analizar diferentes fenómenos. Con la ayuda de los análisis realizados en este documento, podrían hacer una selección adecuada de las herramientas a utilizar en sus investigaciones, de tal forma que les permita conocer a profundidad las problemáticas que se presentan, y así, poder aportar en la solución de las mismas para avanzar hacia una mejor sociedad.

4. METODOLOGÍA

El estudio que se realizó es de tipo descriptivo, ya que busca señalar las características más importantes del aprendizaje de máquina no supervisado y sus aplicaciones en el campo de la investigación de fenómenos sociales. Para ello el método a utilizado fue el análisis, que permitió descomponer y entender de qué se trata y cómo se aplica, así como realizar las comparaciones entre técnicas y resultados.

Las fuentes que fueron utilizadas son artículos científicos de revistas indexadas y bases de datos especializadas que aborden aplicaciones del aprendizaje de máquina no supervisado en problemas de ámbito social y similares.

La herramienta principal de búsqueda fue el buscador especializado Web of Science, que es una plataforma de investigación que ayuda a buscar, analizar y compartir rápidamente información del campo de las ciencias, las ciencias sociales, las artes y las humanidades, garantizando acceso integrado a literatura de alta calidad a través de una plataforma unificada que recopila una amplia variedad de contenido y términos de búsqueda.

5. TÉCNICAS DE APRENDIZAJE DE MÁQUINA NO SUPERVISADO MÁS UTILIZADAS EN EL ANÁLISIS DE PATRONES DE COMPORTAMIENTO EN FENÓMENOS SOCIALES

Debido a la complejidad y variabilidad del comportamiento humano, es una tarea difícil identificar técnicas y componentes para el aprendizaje de máquina no supervisado que ayuden a analizar patrones de comportamientos humanos de manera objetiva. Para esta monografía se realizó una investigación en diferentes fuentes de información donde se encontró información sobre técnicas utilizadas para analizar comportamientos en fenómenos sociales que utilizan aprendizaje de máquina no supervisado. Las técnicas encontradas de forma recurrente fueron: Agrupamiento, clasificadores basados en reglas y rastreo.

AGRUPAMIENTO (CLUSTERING)

El agrupamiento tiene como objetivo identificar grupos de datos similares. La intención es identificar grupos de valores con características similares utilizando procedimientos de agrupamiento.

Los científicos conductuales a menudo definen categorías simplemente por inspección visual, y es aquí donde el correcto empleo de esta técnica define el éxito o fracaso del análisis. En todo caso para poder definir estos grupos de datos se debe contar con la suficiente experiencia en pruebas y recolección de información para poder definir los grupos o categorías. Por ejemplo, Fentress y Stiwell [1] identificaron siete categorías en el análisis de comportamiento de ratones mediante inspección visual las cuales son:

- Grupo de extremidades anteriores
- Grandes grupos de movimientos síncronos pero asimétricos
- Interrupción momentánea del movimiento activo

Sin embargo, esta manera de clasificación visual deja a quien observa decidir el tipo de clasificación a los grupos de datos, y esto no es lo que se busca en esta técnica. Para obtener resultados reproducibles, es necesario que las decisiones sean cuantificables determinando de forma apropiada de características como por ejemplo posiciones y velocidad de movimiento.

CLASIFICADORES BASADOS EN REGLAS (RULE-BASED CLASSIFIERS)

Braun [2] analizó el comportamiento de una mosca, y para clasificar sus comportamientos utilizaron un clúster global considerando todas las características, las cuales son menores que las características del comportamiento humano.

Como inconveniente, este análisis es limitado debido a que los comportamientos son analizados de manera lineal, es decir, como sucesión de eventos suponiendo que solo exista un evento en un momento dado, lo cual no se asemeja a la complejidad del comportamiento social, por ejemplo, caminar y oler. Estos eventos han sido capaces de ser descritos por clasificadores basados en reglas, que son los únicos capaces de describir eventos simultáneas ya que filtran los datos con parámetros geométricos extraídos de las trayectorias, posición relativa, distancias y velocidad.

RASTREO TÉCNICA DE SEGUIMIENTO (TRACKING)

Esta técnica utiliza primitivas geométricas (círculos, cuadrados, triángulos, entre otras) que permiten modelar y rastrear un objeto, sin requerir ningún etiquetado específico.

Debido a que las interacciones sociales y cambios de comportamiento exigen una información integrada proveniente de las emociones, no es posible realizar un análisis simultáneo mediante inspección visual, debido a esto, se utilizan técnicas de análisis de video los cuales aplican métodos de visión por computadora que rastrean objetos (moscas, hormigas, manzanas, ratones, entre muchos más), esto permite el estudio de interacciones y posiciones entre ellos [3]. Sin embargo, para el caso de estudio los métodos existentes no utilizan el conocimiento a priori al rastrear las partes del objeto y, por lo tanto, no pueden lidiar con situaciones como cuando están en contacto cercano, por ejemplo, el tacto.

Con la información de orientación, la distancia y la velocidad se pueden mapear los objetos de estudio, en este caso, ratones que interactúan con estos estados geométricos y de comportamiento donde se monitorea esta información, construyendo representaciones jerárquicas en forma de datos de comportamiento sincrónicos y combinados.

6. APLICACIÓN DE LAS TÉCNICAS DE APRENDIZAJE DE MÁQUINA NO SUPERVISADO

Se revisará en el análisis imparcial del comportamiento social del ratón usando aprendizaje de máquina no supervisado, debido a que los modelos de ratón son ampliamente utilizados para estudiar los mecanismos de los trastornos neuropsiquiátricos y para probar posibles tratamientos en humanos. Sin embargo, las técnicas de aprendizaje de máquina no supervisado utilizadas en este análisis también son utilizadas para predecir comportamientos del mercado, fidelización de clientes, fraudes, preferencias de compra de una persona, estrategias de venta, entre otras.

Para este análisis se requiere que haya nuevas mediciones cuantitativas con menos sesgos que en los experimentos previos hechos mediante inspección visual, y esto se puede lograr debido a que los nuevos estudios de comportamiento están cambiando hacia un análisis computacional. Las nuevas técnicas de visión por computadora permiten rastrear a los animales, transformando los datos de video en trayectorias. Para interpretar el comportamiento, se necesita una descripción de sus partes relevantes de manera objetiva y cuantitativa. Por lo tanto, el objetivo es desarrollar una clasificación de comportamiento automática.

6.1. TÉCNICA DE AGRUPAMIENTO O CLÚSTER

Para extraer automáticamente los comportamientos se han desarrollado varias técnicas no supervisadas, para este propósito se utilizó la técnica de clúster.

La técnica de clúster se utiliza cuando se tiene datos no etiquetados, es decir, datos sin categorías o grupos definidos en los datos. Los datos se agrupan según la similitud o características [2]. En esta técnica la variable independiente (características) es la que

se manipula para determinar grupos en los datos. El desarrollo de esta técnica se define mediante los siguientes pasos:

Primer paso: teniendo un grupo de datos en el plano donde se tendrá de manera aleatoria un número que defina la cantidad de clústeres o agrupaciones, esto es definido como una constante K .

Segundo paso: después de haber definido la constante K , de la misma manera, es decir, aleatoriamente, se selecciona K datos del conjunto total de datos a los cuales se les llamará centroides.

Tercer paso: debido a que cada centroide conforma un grupo, se debe medir la distancia de cada uno de los datos a cada uno de los centroides.

Cuarto paso: se verifican las distancias resultantes del dato seleccionado a cada uno de los centroides, asignando al grupo del centroide, los datos cuya distancia sea la menor, es decir, los datos que están más cerca al centroide.

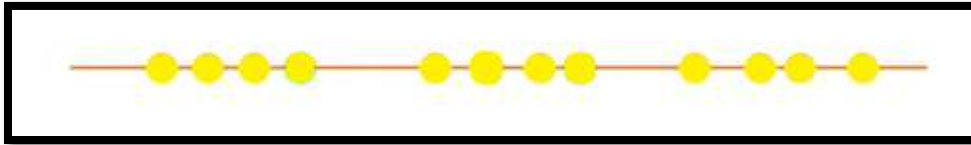
Quinto paso: Se repite el proceso seleccionando los K puntos (centroides) de manera aleatoria generando nuevas posiciones de estos centroides y repitiendo todos los pasos anteriores, generando de esta forma diversos análisis de selección de grupos.

6.1.1. CONDICIÓN DE PARADA

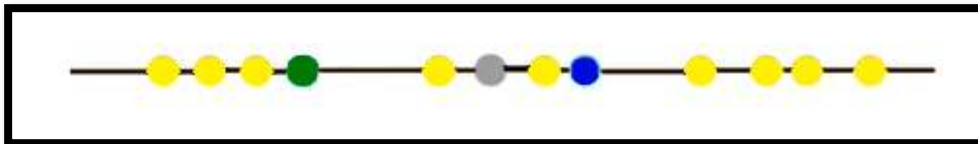
Esta técnica seguirá generando análisis hasta que por sí misma determine que varios de estos se repiten, luego se busca el análisis que más se repite, el cual es el mejor de los análisis de distribución.

Para poder comprender mejor este concepto se presenta un ejemplo el cual permita entender la asignación de los datos a cada grupo:

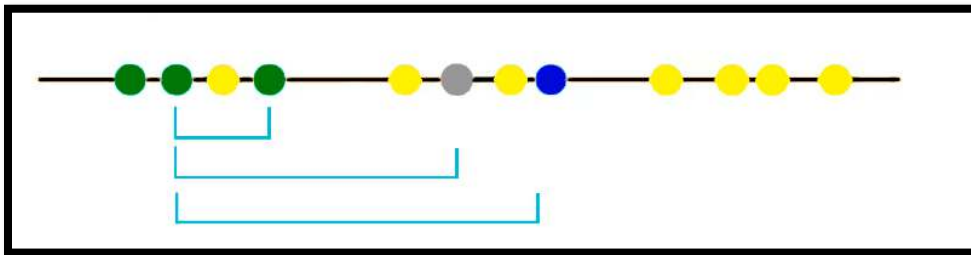
Paso 1: $K=3$



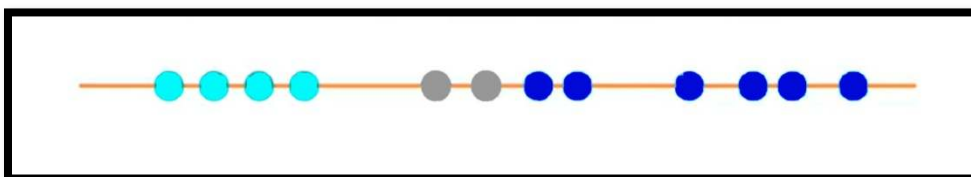
Paso 2: Selección de los K puntos (centroides)



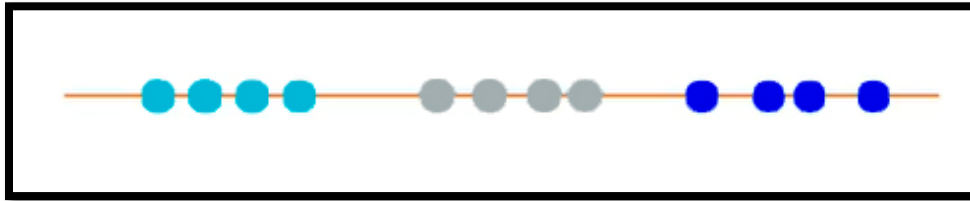
Paso 3: Medición de la distancia entre un dato y cada uno de los centroides, donde la distancia menor define el grupo al que pertenece el dato.



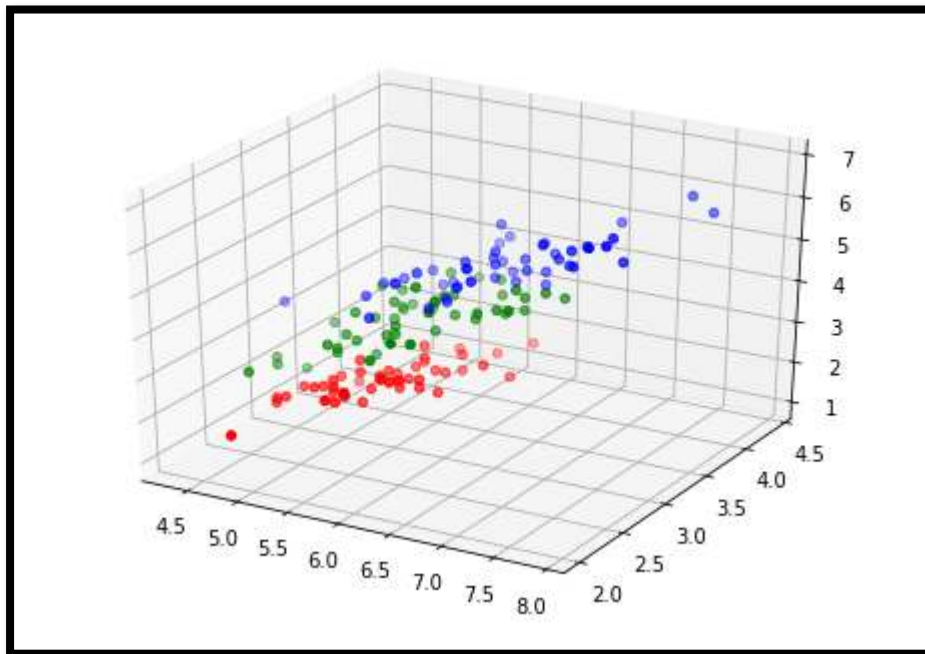
Paso 4: asignación de grupo al dato con respecto a la distancia medida a los centroides fue la menor.



Paso 5: determinación del mejor análisis (condición de parada)



Los datos recolectados por las técnicas de visión por computadora se capturan en un espacio tridimensional, donde aplicando la técnica de agrupamiento realizando los mismos pasos descritos anteriormente. Se debe recordar que la fórmula matemática de la distancia en un espacio tridimensional es $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$, de esta manera la clasificación de los datos en los respectivos grupos se podría visualizar de la siguiente manera:



A diferencia de la clasificación supervisada, se aplican enfoques de agrupamiento si los grupos no se conocen de antemano y, por lo tanto, pertenecen al campo del aprendizaje no supervisado en informática. Sin aplicar conocimientos previos sobre las clases, el proceso de agrupación tiene como objetivo descubrir la estructura interna de los datos,

lo que resulta en clasificaciones objetivas y estables. "Objetivo" significa que los mismos datos procesados con el mismo método conducen a los mismos resultados. "Estable" significa que los resultados son invariables frente a las variaciones de los datos especiales, es decir, si se utiliza otro conjunto de datos debe obtenerse el mismo tipo de experimento.

6.1.2. RESULTADOS

Como resultado tendremos grupos definidos y estables de datos que comparten características similares, por ejemplo, algunos grupos de datos son los que describen características físicas cabeza, cola, cuerpo, orejas, como también grupos que describen situaciones, tipos de movimiento, entre otros. A esto se les llama repertorio de eventos.

6.1.3. RESTRICCIÓN DE LA TÉCNICA DE AGRUPAMIENTO O CLÚSTER

La técnica de clúster es robusta y simple, y proporciona resultados confiables cuando los conjuntos de datos son distintos o bien separados entre sí de forma lineal [4], esto quiere decir, detenerse, avanzar, retroceder, ascender, es decir una sucesión de eventos. Sin embargo, de los datos obtenidos con la visión por computadora se tienen grupos de datos que no fue posible etiquetar mediante la técnica de clúster debido a que no son lineales, es decir, comportamientos simultáneos como por ejemplo avanzar y oler, saltar hacia adelante, retroceder y girar, debido a que este tipo de dato no es lineal. La clasificación de este tipo de datos fue posible por la implementación de otra técnica llamada clasificadores basados en reglas.

6.1.4. CLASIFICADORES BASADOS EN REGLAS

Este tipo de técnica es capaz de identificar eventos simultáneos [4], pero para hacer esto posible se necesita inicialmente un repertorio de eventos, el cual puede generarse de forma automática.

Para tener un mejor uso de este repertorio y mejorar el costo computacional que se requiere al utilizar técnicas de agrupamiento y clasificación, se realiza una vectorización de características de estos grupos donde se considera patrones de comportamiento dinámicos y características fisiológicas generalmente más permanentes, tales como rasgos, movimientos, entre otros.

Un clasificador basado en reglas se aplica a datos resultado del reconocimiento de imágenes obtenidas por visión computarizada, este enfoque es genérico y puede aplicarse a diversos problemas sociales. El núcleo clasificador se basa en arboles de decisión utilizando la combinación IF/THEN [5] donde teniendo definido su repertorio de eventos y el árbol de reglas, el clasificador puede evolucionar por si solo y es capaz de dar una precisión de clasificación muy alta y computacionalmente eficiente [6]. Teniendo en cuenta la vectorización previa de los grupos de datos obtenidos, podemos mejorar la construcción de las reglas IF/THEN para tener un mejor resultado. Adicionalmente puede comenzar la clasificación desde la primera muestra de datos.

6.1.5. ARQUITECTURA DEL CLASIFICADOR

La arquitectura general de un clasificador se representa con las siguientes capas:

1. Bloque de preprocesamiento.
2. Capa de extracción de características.
3. Conjunto masivamente paralelo de reglas altamente interpretables IF/THEN.
4. Tomador de decisiones.

6.1.6. RESTRICCIÓN

Los datos extraídos por esta técnica como por ejemplo trayectorias, posición previa, distancias y velocidades permiten tratar eventos simultáneos, pero solo extraen comportamientos definidos por expertos, por lo tanto, también depende de los umbrales definidos por estos mismos, lo que genera un gran inconveniente para el aprendizaje de máquina no supervisado debido a la intervención, sin embargo, esta restricción se ve superada por la combinación de ambas técnicas: Clúster y tracking.

6.2. TÉCNICA DE SEGUIMIENTO (TRACKING)

El estudio de interacciones sociales en ratones se utiliza como modelo para procesos cognitivos y emocionales normales y patológicos en seres humanos, sin embargo, también se puede utilizar para la detección de problemas socioeconómicos, comportamientos de vida social en todos los estratos sociales y poder verificar qué puede ser el detonante para una acción social sea violenta o cultural. Pero extraer información exhaustiva sobre el comportamiento de los videos de ratones que interactúan sigue siendo un desafío, es aquí donde se debe utilizar técnicas de tracking o seguimiento donde su implementación de forma inicial utilizando la visión por computadora permite recolección de datos, la cual emplea primitivas geométricas para rastrear los ratones del experimento, sin requerir ningún etiquetado específico [7].

Existen variaciones asociadas con la precisión en las que este tipo de técnicas se aplican para llevar a cabo el rastreo de objetos en movimiento, lo cual puede influir de manera significativa sobre la calidad en la captura, así como el rendimiento de procesamiento utilizado por dispositivos físicos contenedores. Actualmente el rastreo de imágenes es de gran importancia para diversos campos debido a que al analizar una interfaz desde el conocimiento teórico sobre cómo las personas percibimos visualmente, podemos predecir en gran medida cuál será el comportamiento visual de los usuarios, detectando qué elementos atraerán su atención visual con más fuerza. Por ejemplo, en función del tipo de elementos que esté buscando visualmente el usuario en cada instante

(contenidos, navegación, mapa del sitio web, contacto...), será mayor la probabilidad de que atienda automáticamente a diferentes zonas de la página; un comportamiento que habrá interiorizado a partir de su experiencia previa navegando por otros sitios web. Algunos de los campos de uso es el de la biología, estudios sociales, educación, seguridad, entre otros, así que desarrollar una herramienta que ayude a solucionar esto es de gran importancia, por ejemplo, en biología se desea estudiar un organismo en particular, de ahí es necesario seguirlo para ver su comportamiento. Se observan distorsiones en las métricas psicológicas que influyen en las elecciones en los trastornos psiquiátricos, como los comportamientos adictivos, la depresión, los trastornos impulsivos y de personalidad. [8]

6.2.1. CLASIFICACIÓN DE COMPORTAMIENTO NO SUPERVISADO

Utilizamos un método de seguimiento que proporciona la ubicación de la cabeza y la base de la cola de los animales. Utilizamos esos datos como un legado de nuestro rastreador ya que fue diseñado para rastrear específicamente aquellas partes del animal que se conocen como elementos clave en la interacción de los animales (la base de la cola es el área anogenital). No obstante, diseñamos el método para que pueda incluir características más detalladas.

Primero se detalla cómo se crean las características. Estas son luego procesados por un análisis de componentes independientes (ICA) que extrae descriptores independientes. Cada componente independiente (IC) muestra una combinación lineal de las características. Luego, se utiliza un modelo de mezcla Gaussiana unidimensional (GMM) para generar en cada CI una agrupación diferente de comportamientos exclusivos, como animales con o sin contacto.

Durante las interacciones sociales se pueden obtener múltiples y rápidos cambios de comportamiento y exhiben una información integrada proveniente de la emoción, así como información proveniente del comportamiento común.

Medir el comportamiento animal es difícil. Tanto la agresión como el cortejo, por ejemplo, consisten en conjuntos ricos de comportamientos estereotipados, que a menudo se desarrollan en una secuencia característica. Actualmente, estos comportamientos se miden manualmente, lo cual es lento y laborioso. Las decisiones subjetivas del observador pueden llevar a dificultades para reproducir los experimentos. Además, los observadores humanos pueden fallar al detectar eventos de comportamiento que son demasiado rápidos o también demasiado lentos, y pueden perder eventos debido a la falta de atención. Estos constituyen obstáculos sustanciales para llevar a cabo demostraciones de comportamiento a gran escala. Con base en lo anterior se propone la implementación de un análisis de algoritmos de seguimiento basado en técnicas de visión por computador y machine learning, para identificar, rastrear y clasificar diferentes elementos y patrones presentes en un video, además se implementará un modelo de filtros gaussianos y después de ser implementado es donde se empieza a buscar una forma de selección óptima, en este caso, la técnica de clúster que define los grupos de características comunes.

6.2.2. EXTRACCIÓN DE COMPONENTES INDEPENDIENTES

Las salidas de la técnica de seguimiento son la ubicación dimensional del inicio y del final de la longitud del cuerpo relacionado en la imagen por computadora en cada punto de tiempo, en este caso, los ratones estudiados en el modelo de ratón para la comprensión del comportamiento humano en estudios de comportamiento.

Los componentes son extraídos por el analizador de componentes independientes y de este modo se maximiza la información extraída. Estas variables consisten en distancias y ángulos, medidos en cada cuadro o entre cuadros consecutivos para características dinámicas. En la siguiente tabla se muestra las características extraídas de la ubicación de la cabeza y la cola de ambos animales. Las características son individuales (1 a 4, es decir, longitud del animal, velocidad) o sociales (5 a 13, es decir, distancias, ángulos entre animales). Luego alimentamos un ICA con el conjunto completo de características medidas. [4]

- 1 length of the A animal vector $\|\overrightarrow{T_t^A H_t^A}\|$
- 2 speed of the head of the A animal $\|\overrightarrow{H_t^A H_{t+1}^A}\|$
- 3 speed of the tail base of the A animal $\|\overrightarrow{T_t^A T_{t+1}^A}\|$
- 4 directional change of the A animal vector $\overrightarrow{T_t^A H_t^A}, \overrightarrow{T_{t+1}^A H_{t+1}^A}$
- 5 distance between A and B heads $\|\overrightarrow{H_t^A H_t^B}\|$
- 6 distance between A head and B tail $\|\overrightarrow{H_t^A T_t^B}\|$
- 7 distance between A tail and B head $\|\overrightarrow{T_t^A H_t^B}\|$
- 8 distance between A and B tail $\|\overrightarrow{T_t^A T_t^B}\|$
- 9 speed of the A toward the current position of the B

$$\left\| \frac{\overrightarrow{T_t^A + H_t^A}}{2} - \frac{\overrightarrow{T_t^B + H_t^B}}{2} \right\| - \left\| \frac{\overrightarrow{T_{t+1}^A + H_{t+1}^A}}{2} - \frac{\overrightarrow{T_{t+1}^B + H_{t+1}^B}}{2} \right\|$$
- 10 speed of the A toward the next position of the B

$$\left\| \frac{\overrightarrow{T_t^A + H_t^A}}{2} - \frac{\overrightarrow{T_{t+1}^B + H_{t+1}^B}}{2} \right\| - \left\| \frac{\overrightarrow{T_{t+1}^A + H_{t+1}^A}}{2} - \frac{\overrightarrow{T_{t+1}^B + H_{t+1}^B}}{2} \right\|$$
- 11 angle between the A vector and the vector from A middle to B head

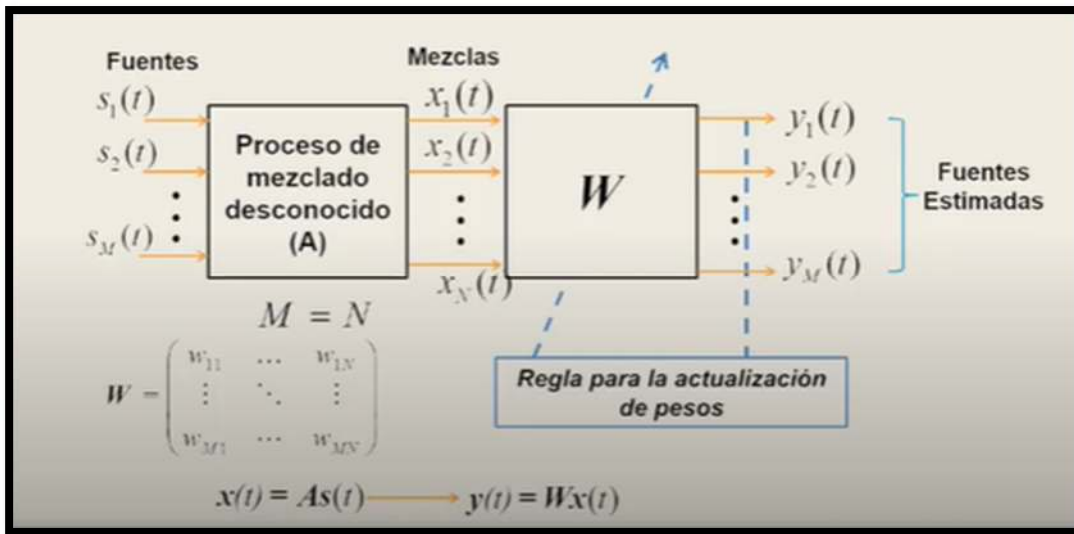
$$\overrightarrow{T_t^A H_t^A}, \overrightarrow{\frac{T_{t+1}^A + H_{t+1}^A}{2} H_{t+1}^B}$$
- 12 angle between the A vector and the vector from A middle to B tail

$$\overrightarrow{T_t^A H_t^A}, \overrightarrow{\frac{T_{t+1}^A + H_{t+1}^A}{2} T_{t+1}^B}$$
- 13 angle between A and B vectors $\overrightarrow{T_t^A H_t^A}, \overrightarrow{T_t^B H_t^B}$

Tabla 1. Cálculo de trayectorias de características, individual y social

Esto llevaba la definición del Análisis de Componentes Independientes, también conocido como ICA (Independent Component Analysis), que se puede hacer de la siguiente manera:

“Dado un conjunto de observaciones de variables aleatorias $\{x_1(t), x_2(t) \dots x_n(t)\}$, siendo t el tiempo o el índice de las muestras, se asume que están generadas por una combinación lineal de componentes independientes”.



6.2.2.1. MODELOS DE MEZCLA GAUSSIANA

Un modelo de mezcla gaussiana involucra la mezcla, es decir, la superposición de múltiples distribuciones gaussianas. Supongamos que se tienen tres distribuciones formadas por muestras de tres clases distintas.

El primer gaussiano representa el nivel de educación de las personas que componen la clase baja. El segundo gaussiano representa el nivel de educación de las personas que componen la clase media, y el tercer gaussiano representa el nivel de educación de las personas que componen la clase alta. Sin saber qué muestras provienen de qué clase, el objetivo será utilizar modelos de mezcla gaussiana para asignar los puntos de datos al grupo apropiado y de esta manera optimizar la asignación de datos a los grupos correspondientes y evitar que algunos datos queden asignados a grupos que no correspondan y generar alguna incoherencia, es decir, que este tipo de modelo ayuda a las técnicas utilizadas en el análisis a agrupar los datos de una manera más eficiente, siempre buscando, desde diferentes algoritmos, optimizar la selección y de igual manera agrupar o desechar datos de información no relacionada y clasificarla como no pertinente para el análisis conductual que se realice.

7. ANÁLISIS COMPARATIVO DE LOS RESULTADOS OBTENIDOS CON LA APLICACIÓN DE LAS TÉCNICAS

A continuación, se puede observar un análisis comparativo de cada técnica estudiada:

TÉCNICA	CLÚSTER	ETIQUETADO BASADO EN REGLAS	TRACKING
INICIO AUTOMÁTICO	Si, debido a que solo necesita los datos de entrada sin importar orden o tipo, usando sus pasos de ejecución y condición de parada, no necesita acoplarse a alguna otra técnica de agrupación.	No, debido a que para poder tener un inicio automático necesita un repertorio de eventos iniciales y esto se logra a través de otras técnicas de agrupación.	No, debido a que necesita de agrupación previa de datos para poder identificar el objeto, realizar seguimientos y generar con cada procesamiento una mejor clasificación.
ADAPTABILIDAD	Si, al ser una técnica de agrupación automático e independiente, puede ser utilizado por otras técnicas para realizar el agrupamiento y la clasificación inicial	Si, al integrarse con técnicas de agrupamiento lineal soluciona problemas cuando en el repertorio de datos hay eventos simultáneos que no han podido definirse, y con cada iteración se entrena y evoluciona a una mejor selección.	Si, al utilizar otras técnicas de selección, modelos de extrapolación de imágenes y modelos de asignación de grupos que evitan una asignación incoherente de algún dato al un grupo.
COSTO COMPUTACIONAL	Alto, debido al cálculo de distancias entre el dato y el centroide hace que cuando haya un alto número de datos el costo computacional se más alto	Bajo, al tener sus reglas definidas y no tener la necesidad de realizar algún cálculo matemático.	Bajo, al segmentar su clasificación con una técnica inicial, procesa los datos clasificados con diferentes modelos y evita ambigüedades, haciendo que su costo computacional sea bajo utilizando un "divide y vencerás".

Realizando un análisis comparativo con las técnicas utilizadas, podemos validar que, aunque su costo computacional sea mayor, la mejor técnica de selección es la de clúster por su automatización independiente y adaptabilidad con otras técnicas, ya que permite un inicio de selección automático independiente de otras técnicas y métodos, pero también, una selección de datos y agrupamiento más exacto solo teniendo la restricción de agrupamiento de datos lineales, al ser el complemento perfecto con otras técnicas de clasificación, permite un aprendizaje de máquina no supervisado más completo y evolutivo.

8. CONCLUSIONES

- Se realizó la consulta de diferentes artículos y libros especializados, en los cuales se encuentran métodos, técnicas y modelos interesantes que pueden ser aplicados ampliamente en el estudio de diferentes cambios en las conductas sociales de individuos o grupos de individuos.
- Se identificó que las técnicas utilizadas no sólo se usan en campos de la salud, biología o mercadeo, sino que también se utilizan para analizar fenómenos sociales al extraer componentes independientes y clasificar información, obteniendo resultados y pudiendo identificar características que por simple inspección visual no se podrían identificar, por ejemplo, tendencias musicales, migraciones masivas, tendencias de consumo, actividades delincuenciales, pandemias, entre otras. Estas técnicas permitirían identificar patrones de comportamiento, características físicas individuales y en grupo, efectos económicos y sociales, y crecimientos poblacionales.
- Las técnicas descritas son técnicas complementarias que utilizadas de manera conjunta facilitan la clasificación y procesamiento de información, favoreciendo la toma de decisiones que puedan ayudar a prevenir o tratar fenómenos sociales.
- Debido a que el análisis no contiene técnicas excluyentes, se concluye que al unir técnicas que se puedan complementar al clasificar, tratar y procesar información para obtener decisiones que puedan ayudar a prevenir o tratar fenómenos sociales.
- La conclusión de usos específicos de estas técnicas es muy difícil por su amplia utilización en diversos campos, sin embargo, por la implementación de visión por computadora el mejor campo de uso es el de tratamiento de información obtenido por movimientos, comportamientos de grupos de individuos como humanos, animales, microbiológicos.
- La técnica de clúster utilizada en altos volúmenes de datos, a pesar de su alto costo computacional, es idónea para generar el repertorio de eventos iniciales necesarios

para que las demás técnicas puedan ser automáticas, es decir, independientes de la intervención de un experto.

BIBLIOGRAFÍA

- [1] J. C. Fentress y F. P. Stilwell, «Grammar of a movement sequence in inbred mice,» *Nature*, vol. 244, nº 5410, pp. 52-53., 1973.
- [2] E. Braun, B. Geurten y M. Egelhaaf, «Identifying prototypical components in behaviour using clustering algorithms,» *PloS one*, vol. 5, nº 2, p. e9361, 2010.
- [3] F. De Chaumont, R. D. S. Coura, P. Serreau, A. Cressant, J. Chabout, S. Granon y J. C. Olivo-Marin, «Computerized video analysis of social interactions in mice,» *Nature methods*, vol. 9, nº 4, pp. 410-417, 2012.
- [4] O. Bauer, A. M. Le Sourd, G. Nardi, T. Bourgeron y F. Chaumont, «Unbiased analysis of mouse social behaviour using unsupervised machine learning,» de *IEEE 14th International Symposium on Biomedical Imaging*, Melbourne, 2017.
- [5] J. Pinho Lucas, «Métodos de clasificación basados en asociación aplicados a sistemas de recomendación,» Universidad de Salamanca, Salamanca, 2010.
- [6] J. A. Hernández Pérez, R. Hernández León y T. C. Autor, «Clasificación multi-etiquetas basada en reglas de asociación de clases,» *Control, Cibernética y Automatización*, vol. 2, nº 1, 2014.
- [7] D. Santos, L. Dallos y P. . A. Gaona-García., «Algoritmos de rastreo de movimiento utilizando técnicas de inteligencia artificial y machine learning,» *Información tecnológica*, vol. 31, nº 3, pp. 23-38., 2020.
- [8] Lorbach, M., Poppe, R., Van Dam, E. A., Noldus, L. P. y Veltkamp, R. C., «Automated recognition of social behavior in rats: The role of feature quality,» de *International Conference on Image Analysis and Processing*, Cham, 2015.