



UNIVERSIDAD NACIONAL DE COLOMBIA

# Modelo de prevención de fraude basado en video. Una aplicación de redes neuronales y modelos estadísticos.

Gerardo Antonio García Arias

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Estadística  
Bogotá, Colombia  
2022-I





UNIVERSIDAD NACIONAL DE COLOMBIA

# Modelo de prevención de fraude basado en video. Una aplicación de redes neuronales y modelos estadísticos.

**Gerardo Antonio García Arias**

Director:

PhD. Carlos Eduardo Alonso Malaver

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Estadística  
Trabajo final de grado  
Bogotá, Colombia  
2022-I



La estadística es el único tribunal de apelación para juzgar el nuevo conocimiento.

P. C. Mahalanobis

A mis hijos, Ammy e Ian, a mi compañera de nave espacial Jennifer Rodríguez.



# Agradecimientos

A mis hijos y esposa, quienes inspiran, acompañan y apoyan en el día a día del camino de la vida.

A mi madre María Arias D.E.P., a mis padres putativos Nancy Parada y Jairo Pelayo y, a toda su familia por su integridad, perseverancia, entrega y, enseñanzas.

A la Universidad Nacional de Colombia que acogió a mi familia desde el pregrado, en especial al Departamento de Estadística y al Área de Gestión y Fomento Socioeconómico de Bienestar Universitario.

A la Facultad de Ciencias que mediante el Examen Final de Ciencias Efi-Ciencias y Exención de Derechos Académicos por excelencia académica, impulsaron la maestría del autor.

Al Instituto Pedagógico Arturo Ramírez Montúfar por acoger a mi hija y guiarla hacia los valores y excelencia académica de la comunidad de la Universidad Nacional de Colombia.

A mi director PhD. Carlos Eduardo Alonso Malaver por su guía, enseñanza y constante aporte al presente. A cada uno de los docentes asociados al Departamento de Estadística; en especial al D.Sc. Luis Hernando Vanegas Penagos por su instrucción en los Modelos Lineales Generalizados y Muestreo y, al Ph.D. Alvaro Mauricio Montenegro Díaz por su docencia en el aprendizaje de máquina y métodos de computación estadística. Finalmente, a cada uno de mis compañeros ,colegas y amigos que nos acompañaron en cada aprendizaje y resultado obtenido.



# Resumen

## **Modelo de prevención de fraude basado en video. Una aplicación de redes neuronales y modelos estadísticos.**

Dada una nueva tipología de fraude, en la que cajeros automáticos son bloqueados con el propósito de crear una distracción de los usuario; lo cual, permite el cambio de una tarjeta asociada a un producto financiero y la captura de su clave. El presente trabajo, propone un modelo lineal generalizado como herramienta de pronóstico de la probabilidad de ocurrencia de fraude, mediante la estructuración de una base de datos, extraída de vídeos por medio de redes neuronales convolucionales. Estos modelos estiman la presencia de personas, encontrando el punto de partida para realizar el rastreo de cada individuo por medio de redes neuronales siamesas. La metodología, permite la construcción de covariables en función de la ubicación espacio-temporal de las personas en el lugar de los hechos, insumo que permite la identificación del modelo lineal.

**Palabras clave:** (Modelo de Prevención de Fraude, Modelos Lineales Generalizados, Redes Neuronales, Detección de Objetos, Seguimiento por Vídeo).

# Abstract

## **Video based prevention fraud model. An application of neuronal networks and statistical models.**

Given a new type of fraud, in which ATMs are blocked with the purpose of creating a distraction for users that allows them to change a card associated with a financial product and capture the password. This paper proposes a Generalized Linear Model as a tool for forecasting the probability of fraud occurrences, by structuring a database extracted from videos by means of Convolutional Neural Networks. This model estimate the presence of people, by finding the starting point to track each individual with a Siamese Neural Networks. The proposal enable the construction of covariates based on the spatio-temporal location of the people at the scene of the events. Input that allows the identification of the lineal model.

**Keywords:** Fraud Prevention Model, Generalized Linear Model, Neural Networks, Object Detection, Tracker Video).

# Contenido

<b>Agradecimientos</b>	<b>vii</b>
<b>Resumen</b>	<b>ix</b>
<b>Lista de símbolos</b>	<b>xii</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Problema de investigación</b>	<b>4</b>
2.1. Justificación . . . . .	4
2.2. Pregunta de investigación . . . . .	4
2.3. Objetivos . . . . .	4
2.3.1. Objetivo General . . . . .	4
2.3.2. Objetivos Específicos . . . . .	4
2.4. Tipología de fraude . . . . .	5
2.5. Metodología . . . . .	6
2.5.1. Covariables . . . . .	7
2.5.2. Conjunto de datos . . . . .	8
<b>3. Marco teórico</b>	<b>10</b>
3.1. Reconocimiento de objetos . . . . .	10
3.1.1. Redes neuronales artificiales . . . . .	10
3.1.2. Perceptrón multicapa . . . . .	12
3.1.3. Estimación de parámetros . . . . .	15
3.1.4. Redes Neuronales Convoluciones . . . . .	17
3.2. Rastreador en vídeo . . . . .	18
3.2.1. Redes siamesas convolucionales . . . . .	20
3.3. Redes neuronales y la estadística . . . . .	23
3.4. Modelos lineales generalizados . . . . .	24
<b>4. Resultados</b>	<b>27</b>
4.1. Reconocimiento de objetos . . . . .	27
4.2. Rastreo de objetos . . . . .	29
4.3. Construcción de variables y análisis descriptivo . . . . .	33

---

4.4. Modelado . . . . .	37
4.4.1. Validación del modelo . . . . .	41
<b>5. Conclusiones y recomendaciones</b>	<b>44</b>
5.1. Conclusiones . . . . .	44
5.2. Recomendaciones . . . . .	45
<b>A. Selección de covariables</b>	<b>46</b>
<b>B. Análisis de sensibilidad.</b>	<b>52</b>
<b>C. Redes neuronales y estructuración de recorridos.</b>	<b>55</b>
<b>Referencias</b>	<b>56</b>

# Lista de símbolos

Subíndice	Término
Adam	<i>Adaptive moments</i>
AN	<i>Artificial neuron</i>
ANN	<i>Artificial neural network</i>
CE	<i>Criterion estimation</i>
CNN	<i>Convolutional neuronal network</i>
DNN	<i>Deep neuronal network</i>
FED	Familia exponencial de dispersión GLM
	<i>Generalized linear models</i>
HOG	<i>Histogram orient gradient</i>
LBP	<i>Local binary pattern</i>
LSTM	<i>Long short-term memory</i>
MLP	<i>Multi layer perceptron</i>
NN	<i>Neuronal network</i>
OR	<i>Object recognition</i>
RoW	<i>Response of a candidate window</i>
RPN	<i>Region proposal network</i>
SGD	<i>Stochastic gradient descendent</i>
SSE	<i>Sum of squared estimate of errors</i>
TFN	Tasa de falsos negativos
TFP	Tasa de falsos positivos
VOT	<i>Video object tracking</i>
VPN	Valor predictivo negativo
VPP	Valor predictivo positivo

# 1. Introducción

En años recientes, varios bancos comerciales han retirado gradualmente la seguridad física, de aquellos lugares con grupos de cajeros automáticos propios de la entidad. Dándose así, el escenario para que se materialicen nuevas tipologías de fraude. De particular atención, en éste trabajo se analizará el tipo de fraude conocido como “cambiazo”, nombre que se da por el hecho que la tarjeta del cliente es cambiada al acercarse a los centros tecnológicos, en pro de realizar sus transacciones con su producto financiero.

En general, un grupo organizado de personas ejecuta al menos una de las actividades subyacentes al fraude, que van desde distraer o preparar la escena para su ejecución, hasta finalmente extraer los fondos de la cuenta. Construyendo elaboradas estrategias, que en la mayoría de casos no permiten individualizar cargos legales o bloquear las cuentas antes de ser afectados sus fondos. Por otro lado, el marco legal asociado, no se posiciona como un factor de protección ante el fraude, ya que, cuando se originan capturas los implicados resultan sin consecuencias judiciales. Adicionalmente, debido a su organización y estrategia, no se tienen herramientas digitales para bloquear el producto cuando el delito está en ejecución.

El proceso generalmente inicia con un bloqueo de las máquinas dispensadoras de dinero - cajeros electrónicos -, de tal forma que los sistemas transaccionales del banco no reciben información que permita diferenciar entre una falla usual y la efectuada como preparación del fraude. Además, las intervenciones realizadas al cajero no son perceptibles por los usuarios. Ante la entrada de una persona que se adecua a los perfiles susceptibles de fraude, los delincuentes simulan una transacción habitual y, dado que el cliente no puede efectuar sus operaciones ofrecen ayuda, distrayendo y cambiando la tarjeta por una similar y capturando la clave de la misma.

Posterior a la realización del fraude, en general, se abre una reclamación y una acción legal del cliente hacia la entidad financiera, la cual, finaliza con una sentencia a favor del banco debido a la permisividad de la persona hacia los delincuentes. Sin embargo, este proceso genera una serie de costos al ente económico, es decir, gastos que pueden tratarse como pérdidas. Unido a lo anterior, son afectados los activos del titular del producto financiero, persona que puede llegar a cancelar el producto afectado o todos los productos a su nombre; hecho que implica otro tipo de pérdida. A lo que se suma la menor oportunidad de establecer nuevos contratos financieros, dado que la confianza en la entidad financiera decrece. Atado a lo

anterior, puede llegar a presentarse la materialización de un riesgo reputacional, dándose la posibilidad de deserción de otros clientes o la disminución de la confianza para establecer nuevas colocaciones de los productos de la entidad.

Por tanto, para disminuir estos riesgos se busca construir un modelo que pronostique o estime la probabilidad de fraude, en función de información no proporcionada por los cajeros. Para ello, una fuente disponible son las cámaras instaladas en cada centro tecnológico, que graban lo ocurrido a la entrada del lugar y el uso de los dispositivos. De esta manera, el propósito es combinar técnicas de aprendizaje de máquina y herramientas estadísticas, para llegar a una solución. El aprendizaje de máquina se usará para realizar el procesamiento de imágenes y, el rastreo en vídeo con el fin de extraer variables asociadas a la probabilidad de fraude. Una vez obtenida la información se requiere una herramienta de pronóstico para determinar si se tiene evidencia de fraude, herramienta que es un modelo estadístico.

Para ello, se han planteado técnicas para el rastreo de objetos como los abordados por Maggio & Cavallaro (2011), quienes exponen el algoritmo de vecinos más cercanos, de correspondencia de gráficos, de hipótesis múltiples, basados en el gradiente, el rastreador bayesiano, el filtro de Kalman, entre otros. Pero, son Wang et al. (2019) quienes establecen el estado del arte, usando redes neuronales siamesas, las cuales requieren solo de una caja delimitadora inicial para los objetos a seguir, procesando 55 fotogramas por segundo. De esta forma, se debe usar un modelo auxiliar que identifique cada persona al ingresar a cada lugar, para lo cual, según Fiaz et al. (2019) las redes neuronales convolucionales han mejorado el desempeño del reconocimiento de objetos, siendo así el “mejor” modelo. Cabe resaltar que en el contexto de la estadística, las redes neuronales artificiales con alta dimensionalidad son considerados como modelos no paramétricos, Warren (1994).

Haciendo uso de estos dos modelos, se obtiene la posición de cada persona a través del tiempo y espacio, generándose la posibilidad de construir variables o características, como la distancia recorrida en el lugar, el tiempo de permanencia, el número de personas, el tiempo de acompañamiento de los usuarios aledaños, etc. Por tanto, se transforman los datos desestructurados como los vídeos, a una base de datos estructurada donde se alojarán las variables independientes extraídas o entradas, como son denominadas en el contexto de redes. Para luego, estimar o entrenar un modelo lineal generalizado para variable respuesta binaria y con función de enlace probit, que explica y pronostica la realización de la variable aleatoria, presencia o ausencia de fraude. De esta manera, se adiciona un ejemplo de la estrecha relación de las redes neuronales y la estadística, que como dicen Wang et al. (2019) no son metodologías competidoras sino complementarias, que en este casos se combinan para dar solución a un problema complejo de acuerdo al estado de arte de la industria.

Particularmente, esta tipología de fraude es explicada por la distancia máxima recorrida por

los usuarios, el número de personas presentes en el lugar, el patrón de regreso que se manifiesta en el mecanismo del ilícito y, el paso por dos o más cajeros de alguna persona. Dicha componente sistemática proporciona una precisión o clasificación correcta del 94.23 % de los eventos analizados con una detección del 86.67 % de los fraudes. Adicionalmente, se espera que solo el 15 % de las situaciones analizadas sean estimadas como fraude, por lo cual, la propuesta generaría una menor carga operativa para el Banco, con la adición de una mejor protección de los activos de sus clientes.

Sin embargo, el resultado obtenido debe ser optimizado en función de reducir los tiempos de procesamiento, que actualmente se encuentran en un fotograma cada dos segundos, teniendo-se la opción de adaptar la red convolucional usada, la mejora del proceso de construcción de variables y el uso de GPU para la totalidad del proceso. Complementariamente, como trabajo futuro se pueden evaluar otras metodologías y propuestas de variables que permitan reducir la razón de falsos descubrimientos que actualmente se encuentra en el 30 %, por ejemplo, incluyendo la detección de personas sospechosas previamente identificadas.

## 2. Problema de investigación

### 2.1. Justificación

Tener un instrumento que permite detectar una situación de fraudes del tipo descrito en la introducción, tiene una alta relevancia para cualquier entidad bancaria porque:

- Protege la reputación de la entidad bancaria.
- En el caso de llegar a implementar los desarrollos para que se ejecuten en tiempo real, los desarrollos aquí presentado son o serán una herramienta de disuasión.
- En el largo plazo mejora la imagen de la entidad bancaria.

### 2.2. Pregunta de investigación

¿ A partir de vídeos, que modelos son adecuados para estimar o pronosticar la probabilidad de realización de fraude en cajeros de un banco colombiano?.

### 2.3. Objetivos

#### 2.3.1. Objetivo General

Analizar, proponer, identificar y estimar un modelo para la predicción de la probabilidad de realización de fraude extrayendo y construyendo las covariables mediante técnicas propias del aprendizaje de máquina a partir de vídeos de un banco colombiano.

#### 2.3.2. Objetivos Específicos

- Realizar reconocimiento o identificación de personas en imágenes.
- Realizar rastreo o seguimiento de seres humanos en vídeo.
- Identificar el número de personas aledañas a los cajeros automáticos.
- Detectar el recorrido realizado por los clientes alrededor de los cajeros automáticos.

- Estructurar una base de datos con las características (covariables) extraídas de la fuente de información.
- Identificar, seleccionar y estimar un modelo lineal o no lineal que permita estimar la probabilidad de realización de fraude.
- Realizar un análisis de sensibilidad para el modelo seleccionado.

## 2.4. Tipología de fraude

A continuación se presenta el paso a paso del fraude:

1. Hace presencia el primer sospechoso quien vandaliza los cajeros; por ejemplo, con pegante, cinta y/o elementos similares; retirándose luego del cajero.
2. Ante la presencia de un usuario ingresan al menos un sospechoso, con la opción de cómplice(s), quien va a ejecutar el fraude, también conocido como cambiazo. Teniendo en cuenta que el usuario no puede realizar la transacción correspondiente debido a la preparación previa, el sospechoso ofrece su “colaboración”.
3. El cliente permisivo da la oportunidad de que visualicen la clave de su cuenta y, además el sospechoso manipula la tarjeta la cual es cambiada sin percatarse el usuario.
4. Tan pronto realizan el cambiazo se retiran de los cajeros.

En los pasos 2 a 3 el sospechoso o su cómplice aparenta que los cajeros están en buen estado y que realiza una transacción exitosa. Adicionalmente, en una buena proporción de los casos, entre los pasos 2 a 3 el cliente intenta realizar sus transacciones, pero al no lograr su objetivo se retira del lugar. En ese instante un delincuente, distrae al usuario y lo hace retornar a los cajeros para materializar el fraude, ver figura 2-1. La tipología expuesta se ejecuta con uno o más delincuentes, con modalidades más frecuentes en pareja y en solitario.

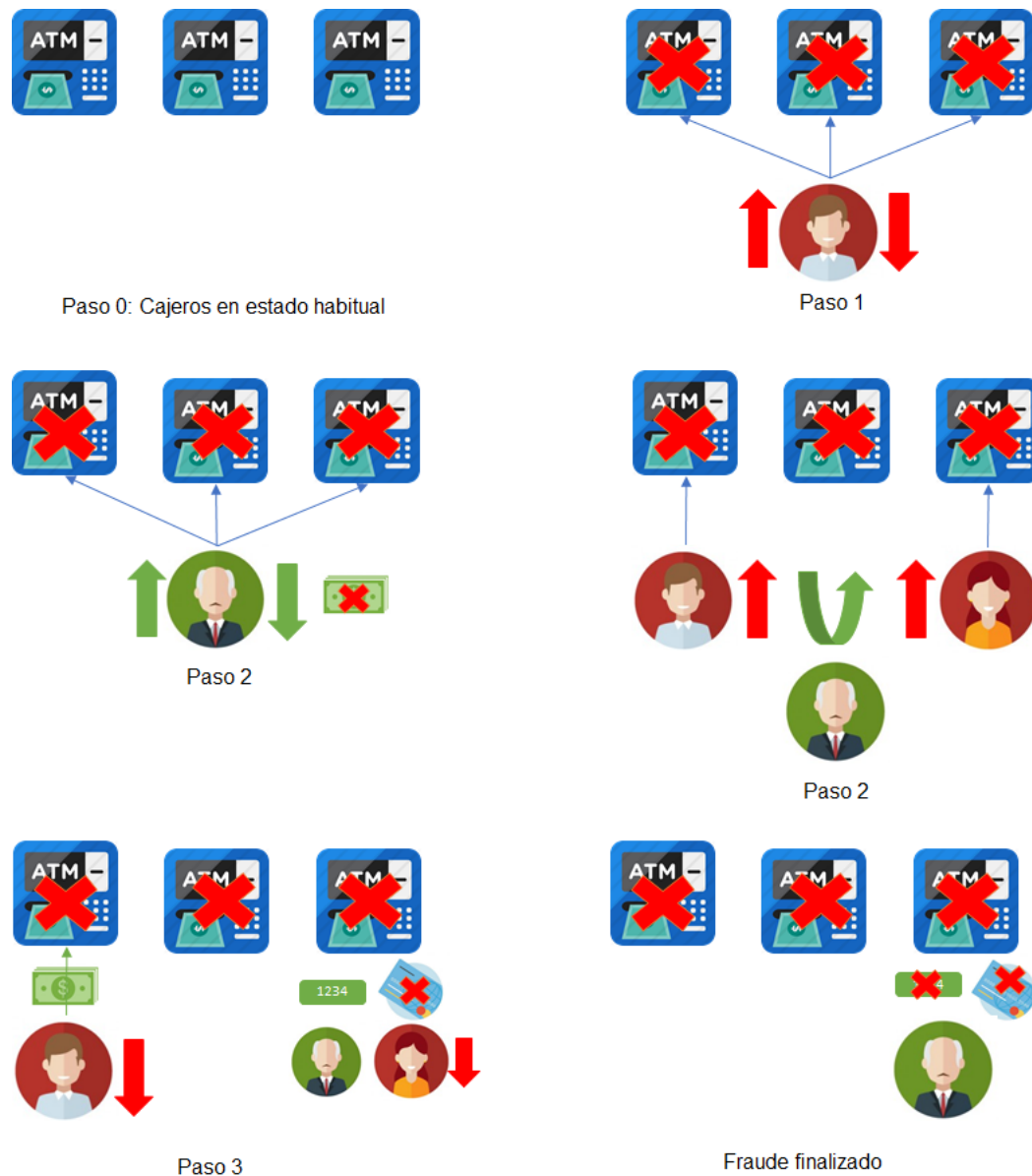


Figura 2-1.: Tipología de fraude

## 2.5. Metodología

Del conocimiento previo y de las características de la tipología empleada (ver 2.4), se espera que el delincuente o grupo de ellos estén aledaños a la víctima de fraude, en un tiempo y espacio menor que un acompañante natural del usuario. Por ello, se hace necesario realizar el rastreo en vídeo de cada persona, labor que es factible dada la evolución reciente del hardware de las computadoras, el nacimiento de optimizadores novedosos y la construcción de algoritmos para dar solución a sistemas de ecuaciones con alta dimensionalidad. Esto,

proporciona los métodos y herramientas necesarias para desplegar las técnicas propias del aprendizaje de máquina como las mostradas en la sección 3.

Para la solución del problema, cada  $j$ -ésimo vídeo es descompuesto en una serie de fotogramas  $ft_{ji_{i \geq 1}}$  indexados por el tiempo, por lo cual, para la construcción de una base de datos con las variables insumo para un modelo estadístico, que explique la probabilidad de fraude a partir de vídeo, se realiza un proceso inicial de identificación de cada cliente o usuario de los centros tecnológicos del banco. Lo anterior, por medio de un modelo de detección de objetos, como lo son las redes neuronales convolucionales, ver 3.1. De esta forma, se tiene el punto de partida que requieren las redes siamesas (3.2) que realizan la tarea de rastreo, construyéndose en la secuencia  $ft_{j1}, \dots, ft_{jn_j}$  la posición de cada persona.

De esta manera, en la etapa de detección y rastreo se realiza la reconstrucción de  $\mathfrak{R}^2$ , en el que cada individuo se representa como un punto con coordenadas  $(x_{(\cdot)}, y_{(\cdot)})$ . Solucionado esto, por ejemplo, en cada instante se tiene el número de individuos, dándose además la posibilidad de construir cada covariable que este en función del tiempo y espacio planteadas en 2.5.1.

Luego, se construye y estructura una base de datos con la variable respuesta (fraude, no fraude), obteniéndose todos los insumos para la identificación, estimación, selección y análisis de sensibilidad de un modelo estadístico. Dicho modelo, se propone pertenezca a la familia de modelos lineales generalizados. Un resumen de la metodología planteada se muestra en la figura 2-2.

### 2.5.1. Covariables

A partir de una investigación previa al presente trabajo, se han identificado algunas variables candidatas a ser factores que modifican la probabilidad de fraude y, cuya información se extrae de los vídeos. A continuación, se enumeran dichas características:

1. Distancia recorrida dentro del lugar.
2. Distancia entre seres humanos y el cajero.
3. Número de individuos.
4. Cantidad de movimientos que realiza cada persona alrededor del cajero.
5. Tiempo de permanencia en el cajero o de ejecución de transacciones.
6. Tiempo de acompañamiento bruto.
7. Tiempo de acompañamiento relativo.

8. Identificación del patrón de regreso.
9. Identificación del paso por dos o más cajeros.

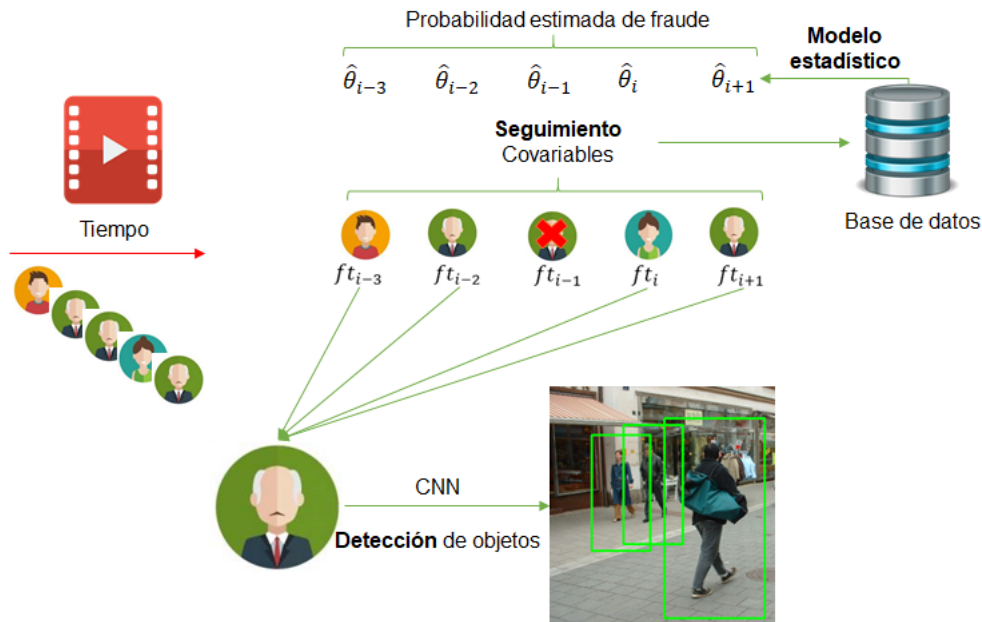
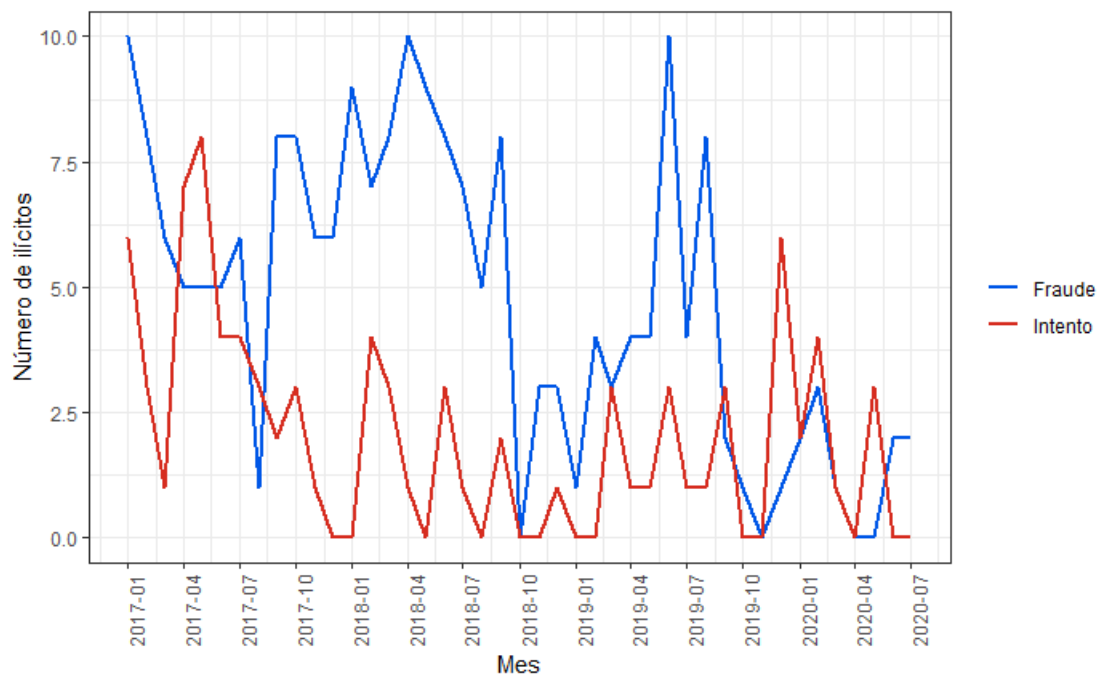


Figura 2-2.: Metodología.

### 2.5.2. Conjunto de datos

De la tipología de fraude a analizar, se han efectuado 300 casos desde 2017 hasta agosto de 2020. Adicionalmente, se identificaron 167 situaciones en las que se evidenció la metodología del cambiazo pero por alguna razón los delincuentes no logran capturar la tarjeta del cliente. Este último subconjunto de vídeos, se plantea sea insumo adicional para estimar y validar el modelo, con el objetivo de pronosticar la posible realización de fraude en una etapa temprana del cambiazo.

Se puede apreciar, en la figura 2-3, que el número de fraudes por año en los centros tecnológicos han venido reduciéndose constantemente. Sin embargo, esto se debe a la implementación de personal que vigila e identifica visualmente el fraude (costos fijos elevados) y por otra parte, la habilidad de los delincuentes los lleva a plantear variantes, en las que conducen a los usuarios de los lugares vigilados por el banco a otros escenarios en los que no se cuenta con esos protocolos. Complementariamente, la figura 2-3 puede mostrar que al parecer no existe un comportamiento mensual estacional.



**Figura 2-3.:** Cantidad de realizaciones de fraude, mensual.

Del origen de los datos, se puede decir que las videograbadoras son administradas por un software especializado que permite descargar fragmentos de vídeo, según las necesidades de la entidad financiera. Debido a que la biometría de los clientes está de por medio, dicho software cifra los vídeos para salvaguardar los datos del cliente y cumplir con todos los protocolos de seguridad de información, que debe cumplir la entidad. Alternativamente, la aplicación permite descargar en casos especiales, los vídeos requeridos sin encriptado, pero, en una ventana de tiempo limitada. Por consiguiente, se lograron obtener datos desde agosto de 2019 hasta la actualidad, disponiéndose de 23 realizaciones de cambiaso y 20 intentos del mismo. Por lo anterior, se plantea según el cronograma del presente trabajo de grado, contemplar todos los fraudes o intentos de este ocurridos desde agosto de 2019 hasta febrero de 2021. Complementariamente, se dispone de más de 200 vídeos de diferentes lugares y situaciones en las cuales los usuarios del banco realizan sus transacciones con normalidad, completando así el conjunto de datos a analizar.

	2017	2018	2018	2020
Fraude	74	77	42	10
Intento	42	15	19	10

**Tabla 2-1.:** Cantidad de realizaciones de fraude, anual.

## 3. Marco teórico

Los elementos que se presentan a continuación, son los conceptos y desarrollos que se perciben como necesarios para un mejor entendimiento del trabajo presentado y, cuya finalidad es estimar la probabilidad de fraude dadas las características que se pueden observar o extraer a partir de un vídeo. Para ellos, se registran las acciones que circundan los lugares donde confluyen los clientes bancarios, como se describe en la sección 2. La revisión literaria se puede pensar en dos grandes ramas, una relacionada con el estado del arte de los algoritmos de análisis de vídeos y, la segunda es aquella que cubre los desarrollos y conceptos asociados a los modelos candidatos a emplear en la sección 2.5.

Unido a lo anterior, una de las tareas centrales en el análisis de vídeos es el seguimiento de objetos, afirmación que es planteada por Wang et al. (2019). Sin embargo, antes de iniciar dicha tarea, como menciona el autor, los algoritmos desarrollados requieren una detección inicial o reconocimiento.

Previo al uso de información, en función de la detección y rastreo de objetos, se construirá una base de datos insumo de un modelo estadístico, modelo que en primera instancia se plantea como un modelo lineal generalizado con respuesta tipo Bernoulli.

### 3.1. Reconocimiento de objetos

El reconocimiento consiste en detectar instancias de objetos de una clase determinada en imágenes o vídeos (humanos, vehículos, entre otros). Sin embargo, esta tarea es un problema de alta dimensionalidad, por lo cual, los modelos estadísticos “clásicos” tales como los modelos de regresión y clasificación tienen una aplicabilidad limitada, que se plantea desde la maldición de la dimensionalidad, problema abordado por Lee & Verleysen (2007).

#### 3.1.1. Redes neuronales artificiales

Recientemente los métodos basados en redes neuronales convolucionales (CNN - Convolutional Neural Network ) han sido empleados en tareas de visión por computadora, ganando popularidad al mejorar el rendimiento en reconocimiento de objetos (OR), entre otros (Fiaz

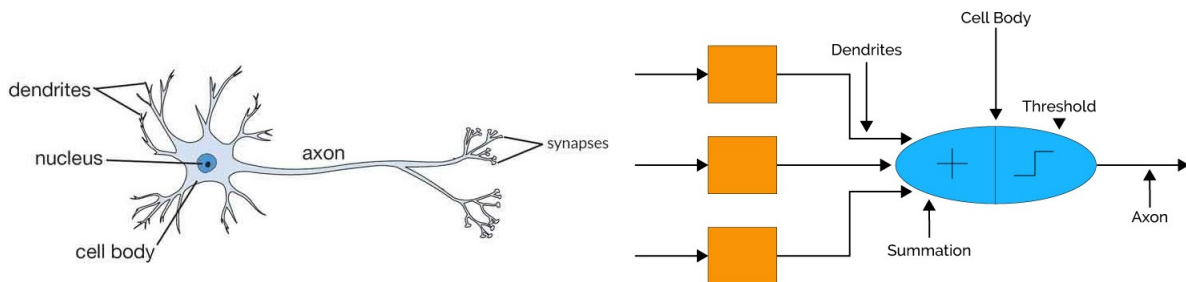
et al. 2019). Para formular este tipo de modelos se especifica o menciona a continuación, su motivación, componentes, la estructura o arquitectura y, criterios de estimación (EC).

Las redes neuronales artificiales (ANN - artificial neural network) son modelos inspirados en el funcionamiento del cerebro de los mamíferos, el cual contiene una red compuesta de billones de neuronas (células nerviosas), interconectadas por medio de trillones de conexiones, Arbib (2003). Así que, se pueden entender como un conjunto de unidades básicas de procesamiento enlazadas y, que dadas unas variables independientes (entradas) procesan la información y producen estimaciones (salidas), Khan et al. (2018).

El perceptrón o neurona artificial (AN) es la unidad básica de estas redes. Están inspiradas en la célula nerviosa la cual consta de cuatro partes con las siguientes funciones:

- *Dendrita*: receptor de señales, desde otras neuronas.
- *Cuerpo celular*: acopia todas las señales y genera una señal combinada de ellas.
- *Axón*: medio de transporte de la señal de salida. Cuando dicha combinación supera un umbral, la neurona produce una señal que viaja a través del axón a otras neuronas.
- *Sinapsis*: punto de interconexión de una neurona con otras, transmite la señal con una intensidad que depende de la fuerza de las conexiones (pesos sinápticos).

Cuando el perceptrón recibe información a través de un vector  $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})^t$ , las dendritas son análogas a las entradas ponderadas  $x_{k1}w_1, \dots, x_{kp}w_p$ , el cuerpo celular ejerce la labor de sumar las señales a través de la combinación lineal  $z_k = \sum_i^p x_i w_i$ ,  $z_k \in \mathcal{R}$ . Finalmente,  $z_k$  es transformado a través de una función  $g(\cdot)$  denominada de activación  $x_k^* = g(z_k)$ , en que  $x_k^*$  ejerce la labor del axón. Así, la ANN es modelada imitando el trabajo básico de una neurona biológica.



**Figura 3-1.:** Estructura de una neurona biológica y artificial (Salman et al. 2018, Figura 3.4.)

Algunas de las arquitecturas o estructuras de las ANN son:

- *Perceptrón simple*
- *Red neuronal de funciones base radiales*
- *Perceptrón multicapa*
- *Redes neuronales recurrentes*
- *Red neuronal de memoria a corto y largo plazo (LSTM)*
- *Red de Hopfield*
- *Máquinas de Boltzmann*
- *Red neuronal convolucional (CNN)*
- *Red neuronal modular*

Una presentación más profunda de las redes anteriores puede ser consultada en Bengio (2009).

En particular, las CNN son versiones regularizadas de perceptrones multicapa (MLP), en los que cada neurona en una capa está conectada a todas las neuronas de la siguiente capa, organizadas de forma jerárquica e inspiradas en el funcionamiento de la corteza visual de los animales, por lo que presentan un alto desempeño en reconocimiento de objetos. Las ANN se pueden clasificar en dos categorías. La primera, *feed-forward*, en las cuales la información de neurona a neurona es transmitida en un solo sentido, en arquitecturas organizadas jerárquicamente como las CNN. La segunda, *feed-backward*, posee conexiones que transmiten señales a neuronas previamente interconectadas que exhiben habilidad de memoria, de almacenar información y, de relación de secuencias en su memoria interna (Khan et al. 2018).

### 3.1.2. Perceptrón multicapa

El MLP es una de las estructuras más usadas. Un ejemplo son las redes de aprendizaje profundo (DNN), que son MLP con una o más capas ocultas en que la profundidad esta dada por la cantidad de capas ocultas. En general los MLP se caracterizan por tres aspectos:

- **Arquitectura en capas:** ANN compuestas por niveles o capas de unidades de procesamiento (AN) en orden jerárquico. Cada capa contiene un número determinado de AN. Usualmente, la primera capa (capa de entrada) alimenta el MLP o recibe la información a través de covariables y, la última capa (capa de salida) realiza predicciones o estimaciones. Las capas que se encuentran en medio de las capas de entrada y salida son denominadas capas intermedias u ocultas, las cuales se encargan del procesamiento, que no es más que realizar combinaciones lineales y transformaciones afín de las variables independientes del modelo.

- Neuronas artificiales: implementan la función de activación, que dada una información de entrada realiza una transformación de ella para transmitirla a la siguiente capa de la ANN.
- Interconexiones densas: las AN están interconectadas y pueden comunicarse con cualquier otra a través de un parámetro que indica la fuerza de la conexión entre ellas. Son ANN tipo *feed-forward*, luego, la información se desplaza secuencialmente de la capa de entrada a la de salida, es decir, cada AN esta directamente conectada a todas la AN de la siguiente capa.

Considerando una sola capa oculta, como se muestra en la figura **3-2**, el enfoque matemático del MLP con  $p$  covariables, una capa oculta con tres neuronas y una variable respuesta es el siguiente:

Sean  $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})^t$  la  $k$ -ésima observación en la capa de entrada,  $w_{ij}^{(1)}$  el parámetro de la conexión de la neurona  $i$  de la capa de entrada con la  $j$ -ésima de la capa oculta;  $g^{(1)}(\cdot)$  y  $g^{(2)}(\cdot)$  la función de activación de la capa oculta y de salida respectivamente. De esta manera, el procesamiento de información en la capa oculta esta dada por  $\mathbf{x}^{(1)} = g^{(1)}(\mathbf{z}^{(1)})$  en que:

$$\begin{aligned} z_1^{(1)} &= x_{k1}w_{11}^{(1)} + \dots + x_{kp}w_{p1}^{(1)} = \mathbf{x}_k^t \mathbf{w}_1^{(1)} \\ z_2^{(1)} &= x_{k1}w_{12}^{(1)} + \dots + x_{kp}w_{p2}^{(1)} = \mathbf{x}_k^t \mathbf{w}_2^{(1)} \\ z_3^{(1)} &= x_{k1}w_{13}^{(1)} + \dots + x_{kp}w_{p3}^{(1)} = \mathbf{x}_k^t \mathbf{w}_3^{(1)} \\ \mathbf{z}^{(1)} &= (z_1^{(1)}, z_2^{(1)}, z_3^{(1)})^t \\ \mathbf{w}_j^{(1)} &= (w_{1j}^{(1)}, \dots, w_{pj}^{(1)})^t \end{aligned}$$

Finalmente, la predicción o estimación del MLP está dado por  $\hat{\mathbf{y}} = \mathbf{x}^{(2)} = g^{(2)}(\mathbf{z}^{(2)})$  donde:

$$\begin{aligned} z_1^{(2)} &= (z_1^{(1)}w_{11}^{(2)} + z_2^{(1)}w_{21}^{(2)} + z_3^{(1)}w_{31}^{(2)}) = \mathbf{z}^{(1)t} \mathbf{w}_1^{(2)} \\ \mathbf{z}^{(2)} &= (z_1^{(2)})^t \\ \mathbf{w}_l^{(2)} &= (w_{1l}^{(2)}, w_{2l}^{(2)})^t \end{aligned}$$

Además, si se tienen  $n$  unidades experimentales, entonces  $X_{n \times p} = (\mathbf{x}^1, \dots, \mathbf{x}^p)^t$  es la matriz diseño.  $\mathbf{W}_{p \times j}^{(1)} = (\mathbf{w}_1^{(1)}, \dots, \mathbf{w}_j^{(1)})^t$  y  $\mathbf{W}_{j \times l}^{(2)} = (\mathbf{w}_1^{(2)}, \dots, \mathbf{w}_l^{(2)})^t$  son matrices de parámetros, en que  $p, j, l$  son el número de variables independientes, el número de neuronas en la primera capa oculta y de salida respectivamente, entonces:

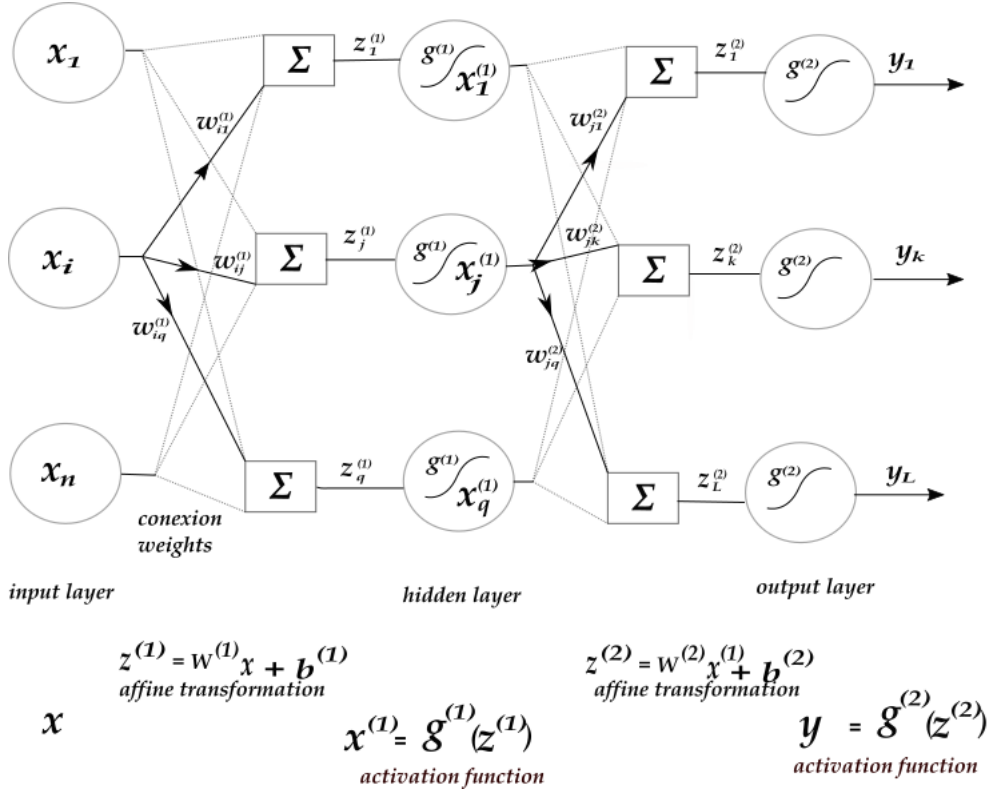


Figura 3-2.: Perceptrón multicapa. Sharma, Avinash (2020)

$$\mathbf{z}^{(1)} = \mathbf{X}\mathbf{W}^{(1)} \quad (3-1)$$

$$\mathbf{X}^{(1)} = \mathbf{g}^{(1)}(\mathbf{z}^{(1)}) \quad (3-2)$$

$$\mathbf{z}^{(2)} = \mathbf{X}^{(1)}\mathbf{W}^{(2)} \quad (3-3)$$

$$\mathbf{X}^{(2)} = \mathbf{g}^{(2)}(\mathbf{z}^{(2)}) \quad (3-4)$$

$$\hat{\mathbf{y}} = \mathbf{X}^{(2)} \quad (3-5)$$

Fácilmente, usando la definición matricial de la arquitectura en la expresión 3-1 a 3-5, se puede generalizar hasta  $L$  capas con tantas neuronas como se desee en cada capa de la ANN. De este manera las ANN son funciones de  $\mathcal{R}^p \rightarrow \mathcal{R}^l$  si los parámetros de la red son conocidos.

$$\begin{aligned} \mathbf{x}_k^{(L)} &= \mathbf{g}^{(L)}(\mathbf{W}^{(L)}\mathbf{g}^{(L-1)}(\mathbf{W}^{(L-1)} \dots \mathbf{g}^{(1)}(\mathbf{W}^{(1)}\mathbf{x}_k) \dots)) \\ &= \mathbf{g}^{(L)}(\mathbf{z}^{(L)}) \\ &= \mathbf{g}^{(L)}(\mathbf{W}^{(L)}\mathbf{x}_k^{(L-1)}) \end{aligned}$$

### 3.1.3. Estimación de parámetros

Dada una función de pérdida o criterio de estimación  $Q(\cdot)$ , se desea encontrar  $\mathbf{W}$  tal que optimicen tal criterio. Por ello, es necesario hacer uso de las herramientas propias del cálculo vectorial, involucrándose en el proceso el cálculo del gradiente de  $Q(\mathbf{W})$ ,  $\nabla Q(\mathbf{w}^*)$ .

#### Regla Delta generalizada

Este algoritmo calcula derivadas realizando una aplicación simple de la regla de la cadena, llamada usualmente *backpropagation*, haciendo uso de funciones de activación lineales o no lineales en cada AN para modelar la relación entre los dominios de la capa de entrada (covariables) y salida (variable(s) respuesta), donde los errores (residuos) se propagan recursivamente de la capa de salida hacia atrás a través de las múltiples capas de la red (Khan et al. 2018).

Sea  $Q(\mathbf{y}, \mathbf{x}_k^{(L)} | \mathbf{x}_k)$  un criterio de estimación o función de pérdida, entonces usando la regla de la cadena:

$$\begin{aligned} \frac{\partial Q(\mathbf{y}, \mathbf{x}_k^{(L)} | \mathbf{x}_k)}{\partial \mathbf{x}_k} &= \frac{\partial Q(\mathbf{y}, \mathbf{x}_k^{(L)} | \mathbf{x}_k)}{\partial \mathbf{x}_k^{(L)}} \frac{\partial \mathbf{x}_k^{(L)}}{\partial \mathbf{z}^{(L)}} \frac{\partial \mathbf{z}^{(L)}}{\partial \mathbf{x}_k^{(L-1)}} \frac{\partial \mathbf{x}_k^{(L-1)}}{\partial \mathbf{z}^{(L-1)}} \frac{\partial \mathbf{z}^{(L-1)}}{\partial \mathbf{x}_k^{(L-2)}} \cdots \frac{\partial \mathbf{x}_k^{(1)}}{\partial \mathbf{z}^{(1)}} \frac{\partial \mathbf{z}^{(1)}}{\partial \mathbf{x}_k} \\ &= \frac{\partial Q(\mathbf{y}, \mathbf{x}_k^{(L)} | \mathbf{x}_k)}{\partial \mathbf{x}_k^{(L)}} g^{(L)'}(\mathbf{z}^{(L)}) \mathbf{W}^{(L)} g^{(L-1)'}(\mathbf{z}^{(L-1)}) \mathbf{W}^{(L-1)} \cdots g^{(1)'}(\mathbf{z}^{(1)}) \mathbf{W}^{(1)} \end{aligned}$$

Bajo el problema de regresión, con el criterio de estimación suma de cuadrados de los residuos (SSE), para una variable respuesta  $\mathbf{y}$  en el modelo planteado en la figura 3-2:

$$Q(\mathbf{W} | \mathbf{x}_k) = \frac{1}{2} (y_k - x_k^{(2)})^2 = \frac{1}{2} (y_k - g^{(2)}(z_k^{(2)}))^2$$

El gradiente para los parámetros en la capa de salida está dado por:

$$\frac{\partial Q(\mathbf{W} | \mathbf{x}_k)}{\partial w_{jl}^{(2)}} = \left( \frac{\partial Q(\mathbf{W} | \mathbf{x}_k)}{\partial x_l^{(2)}} \right) \left( \frac{\partial x_l^{(2)}}{\partial z_l^{(2)}} \right) \left( \frac{\partial z_l^{(2)}}{\partial w_{jl}^{(2)}} \right) = (y_k - x_k^{(2)}) g^{(2)'}(z_l^{(2)}) x_j^{(1)}$$

Definiendo  $\delta_l = (y_k - x_k^{(2)}) (g^{(2)'}(z_l^{(2)}))$ , la expresión que contempla todos los términos con índice  $j$  es:

$$\nabla_{w_{jl}^{(2)}} = \frac{\partial Q}{\partial w_{jl}^{(2)}} = x_j^{(1)} \delta_l$$

Ahora, el gradiente para los parámetros de la oculta es tal que:

$$\begin{aligned}
\frac{\partial Q(\mathbf{W}|\mathbf{x}_k)}{\partial w_{ij}^{(1)}} &= \left( \frac{\partial Q(\mathbf{W}|\mathbf{x}_k)}{\partial x_l^{(2)}} \right) \left( \frac{\partial x_l^{(2)}}{\partial z_l^{(2)}} \right) \left( \frac{\partial z_l^{(2)}}{\partial x_j^{(1)}} \right) \left( \frac{\partial x_j^{(1)}}{\partial z_j^{(1)}} \right) \left( \frac{\partial z_j^{(1)}}{\partial w_{ij}^{(1)}} \right) \\
&= \delta_l w_{jl}^{(2)} g^{(1)'}(z_j^{(1)}) x_i \\
&= x_i g^{(1)'}(z_j^{(1)}) \delta_l w_{jl}^2 \\
\nabla_{w_{ij}^{(1)}} &= x_i \delta_j
\end{aligned}$$

Donde  $\delta_j = g^{(1)'}(z_j^{(1)}) \delta_l w_{jl}^{(2)}$ . En general:

$$\delta^{(l^*-1)} = g^{(l^*-1)'}(z_j^{(l^*-1)}) w_{jl}^{(l^*)}$$

Mishachev (2017) haciendo uso del producto de Hadamard ( $\circ$ ) y el producto ( $\bullet$ ), tal que  $\mathbf{A}(\bullet)\mathbf{B} = \mathbf{B}\mathbf{A}$ , el gradiente de los parámetros en la capa  $l^*$  está dado por:

$$\nabla_{\mathbf{W}}^{l^*} Q(\mathbf{W}|\mathbf{X}) = \Delta_{(l^*+1)} \bullet \mathbf{W}^{(l^*+1)t} g^{(l^*)'}(\mathbf{z}^{(l^*)}) g^{(l^*-1)t}(\mathbf{z}^{(l^*-1)})$$

$$\Delta_{(l^*+1)} = g^{(L)'}(\mathbf{z}^{(L)}) \bullet \mathbf{W}^{(L)t} \circ g^{(L-1)'}(\mathbf{z}^{(L-1)}) \bullet \mathbf{W}^{(L-1)t} \dots g^{(l^*+1)'}(\mathbf{z}^{(l^*+1)})$$

## Optimización

Sea  $\mathbf{W}$  el conjunto de parámetros a estimar en la ANN,  $\mathbf{W}^r$  la  $r$ -ésima iteración del método de optimización;  $Q(\mathbf{W}|\mathbf{X})$  la función objetivo o criterio de estimación a ser minimizado,  $\nabla Q(\mathbf{W}|\mathbf{X})$  el gradiente y,  $\nabla^2 Q(\mathbf{W}|\mathbf{X})$  la matriz hessiana respectiva.

A partir del trabajo de Goodfellow, Bengio & Courville (2016), capítulo 8, en general los métodos empleados para optimizar  $Q(\mathbf{W}|\mathbf{X})$  son de la forma:

$$\mathbf{W}^{r+1} = \alpha \mathbf{W}^r - \eta_r B_r \nabla Q_{r+1}(\mathbf{W}^r|\mathbf{X}) \quad (3-6)$$

De esta manera, cada componente de la ecuación 3-6 puede cambiar en cada iteración del algoritmo, en el cual  $\eta_r$  representa la rata de aprendizaje. Cuando  $\alpha \neq 1$  el algoritmo pertenecerá a los denominados de impulso, considerados que conllevan un aprendizaje, estimación o convergencia más rápida.  $B_r$  para el gradiente descendiente estocástico (SGD) es la identidad  $I$  y  $\alpha = 1$ ; adicionando  $\alpha \neq 1$  se encuentra ante el SGD con momento. Fijando  $\alpha = 1$  y tomando a  $B_k = \nabla^2 Q(\mathbf{W}|\mathbf{X})$  se está ante el método Newton.

De esta familia de optimizadores, el estado del arte es el algoritmo Adam (*adaptive moments*) dado por:

$$\begin{aligned}
g_{r+1} &= Q_{r+1}(\mathbf{W}^r | \mathbf{X}) \\
m_{r+1} &= \beta_1 m_r + (1 - \beta_1) g_{r+1} \text{ (primer momento sesgado)} \\
v_{r+1} &= \beta_2 v_r + (1 - \beta_2) g_{r+1}^2 \text{ (segundo momento sesgado)} \\
\hat{m}_{r+1} &= \frac{m_r}{1 - \beta_1} \text{ (primer momento corregido)} \\
\hat{v}_{r+1} &= \frac{v_r}{1 - \beta_2} \text{ (segundo momento corregido)} \\
\mathbf{W}_{r+1} &= \mathbf{W}_r - \epsilon \frac{\hat{m}_{r+1}}{\sqrt{\hat{v}_{r+1} + \delta}} \text{ (actualización de parámetros)}
\end{aligned}$$

$\epsilon$  llamado *step size* se sugiere sea 0.001,  $\beta_1, \beta_2 \in [0, 1)$  las tasas de decaída exponencial para la estimación de los momentos, recomendados 0,9 y 0,999 respectivamente.  $\delta$  la constante de estabilización ( $10^{-8}$ ),  $m_0 = 0, v_0 = 0$  el valor inicial para el primer y segundo momento.

Además, Goodfellow et al. (2016) mencionan las estrategias para dar el valor inicial  $\mathbf{W}^0$  y, el criterio de parada para los optimizadores. Para cada uno de los métodos mencionados, se fija un valor de convergencia  $\epsilon > 0$  y “cercano” a cero, tal que dada una función de distancia de  $\zeta(\mathbf{W}_{r+1}, \mathbf{W}_r) : \mathfrak{R}^s \rightarrow \mathfrak{R}$ ,  $\hat{\mathbf{W}} = \mathbf{W}_{r+1}$  si y solo si  $\zeta(\mathbf{W}_{r+1}, \mathbf{W}_r) < \epsilon$ , en que  $s$  es la dimensión de  $\mathbf{W}$ .

### 3.1.4. Redes Neuronales Convoluciones

En Goodfellow et al. (2016), se dice que son un tipo especializado de ANN para el procesamiento o modelado de datos con una topología en cuadrícula (*grid topology*). Por ejemplo, series de tiempo (*1D grid*) que pueden ser modeladas tomando muestras en intervalos de tiempo regulares, imágenes (*2D grid*) que son cuadrículas de píxeles, entre otros. Estas redes utilizan una operación matemática llamada convolución en al menos una de sus capas (percentrón multicapa), la cual es una operación lineal especializada.

**Definición 3.1.1.** Convolución:

Operación  $s(\cdot)$  sobre dos funciones de valor real tal que:

$$s(t) = \int f(a)k(t - a)da = (f * k)(t)$$

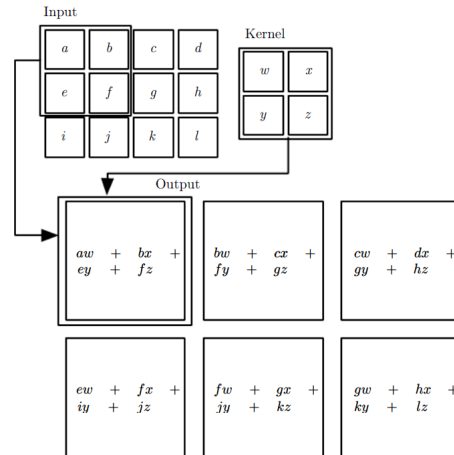
Esta operación está definida siempre y cuando la integral anterior este definida. Además, si se asume  $f(\cdot)$  y  $k(\cdot)$  aplican solamente para los enteros, entonces la convolución discreta es tal que:

$$s(t) = (f * k)(t) = \sum_{a=-\infty}^{\infty} f(a)k(t - a)$$

En el contexto de CNN,  $f(\cdot)$  es un tensor de entrada (datos-covariantes) y  $k(\cdot)$  es un kernel que comúnmente es un arreglo multidimensional de parámetros. En el caso de una imagen bidimensional  $I$  como entrada, se puede usar un kernel bidimensional  $K$ , en cuyo caso  $s(\cdot)$  es:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_l I(m, l)K(i - m, j - l)$$

En la figura 3.1.4, se presenta el bosquejo del funcionamiento de la convolución presentado por Goodfellow et al. (2016).



**Figura 3-3.:** Convolución 2-D

Cabe resaltar que el concepto anterior, guarda una estrecha relación con la estadística. Por ejemplo, si  $X$  y  $Y$  son dos variables aleatorias independientes con funciones de densidad  $f(\cdot)$  y  $k(\cdot)$  respectivamente, entonces la función de densidad de  $Z = X + Y$  es la convolución:

$$(f * k)(t) = \int f(a)k(t - a)da = \int f(t - a)k(a)da$$

## 3.2. Rastreador en vídeo

El seguimiento de objetos en vídeo (VOT), según Wang et al. (2019) es la correspondencia de un objeto entre fotogramas ( $ft$ ) que, dada la localización e identificación de dicho objeto

en el primer fotograma  $ft_1$  en un fragmento de vídeo; el objetivo es estimar su posición en la sub-secuencia de fotogramas  $ft_1, \dots, ft_n$  con la mayor precisión posible. Por otro lado, Maggio & Cavallaro (2011) dicen que el seguimiento en vídeo, es el proceso de estimación a través del tiempo de la localización de uno o más objetos. Además, en Fiaz et al. (2019) se dice que el objetivo de VOT, es identificar una región de interés en fotogramas de un vídeo que consiste de cuatro componentes secuenciales:

1. Identificación del objetivo o objeto(s): la(s) region(es) de interés son etiquetadas (detectadas a priori) usando una representación como un elipse, contorno, centroide o cuadro delimitador.
2. Modelo de la apariencia del objetivo: construir una mejor representación de las características asociadas al objetivo (transformación de covariables) y un modelo estadístico para identificar la región de interés usando metodologías de aprendizaje (estimación).
3. Estimación del movimiento: estima la posición del objetivo en la secuencia  $ft_1, \dots, ft_n$  por medio de una predicción máxima a posteriori.
4. Localización del objetivo

El problema de VOT es simplificado actualizando los modelos de apariencia y movimiento (actualizar las estimaciones de los parámetros asociados), capturando las nuevas características y comportamientos del objetivo a través del tiempo.

Sin embargo, el desempeño de estos rastreadores es afectado por la transformación aplicada a las covariables, clasificadas en *hand-crafted (HC features)* y *deep features*. Las primeras como el histograma de gradientes orientados (HOG), patrones binarios locales (LBP), entre otros. Estos, realizan una representación con desventajas comparadas con aquellas transformaciones basadas en redes profundas, las cuales poseen una mayor capacidad de capturar o representar información multi-nivel, con las variaciones de apariencia del objetivo.

En general, el seguimiento de objetos se centra en estimar una matriz de parámetros  $\mathbf{W}$ , tal que minimice un criterio de estimación con respecto al objetivo respuesta  $\mathbf{z}$ , solucionando un problema de regresión tal que:

$$\| \mathbf{X}\mathbf{W} - \mathbf{z} \|_2^2 + \lambda \| \mathbf{W} \|_2^2$$

Donde  $\mathbf{X}$  es un tensor que representa una imagen o zona de búsqueda,  $\lambda$  es un parámetro de regularización y  $\| \cdot \|_2$  es la norma  $\ell_2$ . De esta manera, se está ante una regresión Ridge cuya solución está dada por:

$$\mathbf{W} = (\mathbf{X}^t\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^t\mathbf{z}$$

Alternativamente, están aquellos métodos basados en filtros de correlación, es decir que, usan una función de similaridades, para comparar la similaridad entre  $\mathbf{X}$  y  $\mathbf{z}$ .

Partiendo de lo anterior, Wang et al. (2019) establece el estado del arte en aplicaciones de VOT en tiempo real, construyendo un rastreador denominado *SiamMask* basado en filtros de correlación. Allí se calcula un *response map* entre un objetivo de rastreo y una sub-región candidata en el dominio de Fourier (Fiaz et al. 2019). De esta manera, se plantea como modelo una red siamés que es función de dos imágenes y calcula en el proceso una medida de similaridad.

*SiamMask* ataca el problema de VOT con velocidad de procesamiento con una ANN, basada en una CNN multi tarea que realiza simultáneamente:

- Estimación de una medida de similaridad entre el objetivo  $\mathbf{z}$  y múltiples regiones candidatas  $\mathbf{X}_{(\cdot)}$ , en una ventana deslizante generando un *dense response map* indicando la ubicación del objetivo.
- Regresión para una caja delimitadora usando una *Region Proposal Network* (RPN).
- Estimación de una máscara de clasificación binaria.

Para lo cual, la función de pérdida usa aditivamente los criterios de estimación empleados en las tres tareas.

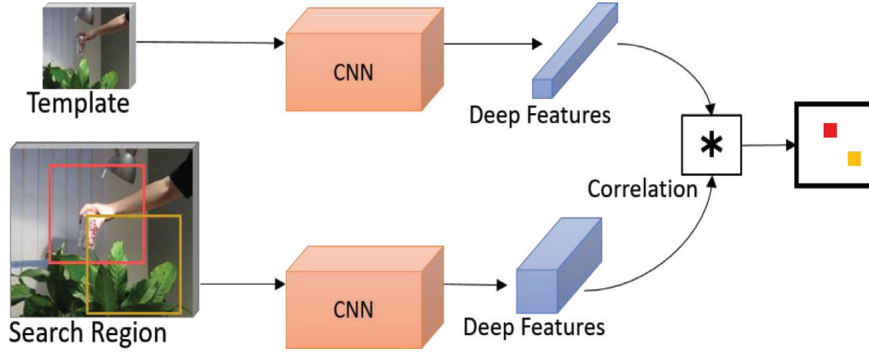
### 3.2.1. Redes siamesas convolucionales

#### SiamFC

Compara una imagen u objetivo  $\mathbf{z}$  contra una imagen de búsqueda  $\mathbf{X}$ , que son procesadas por la misma CNN  $f(\mathbf{W}_\theta)$ , generando dos *feature maps* que son correlacionadas:

$$g(\mathbf{W}_\theta | \mathbf{X}, \mathbf{z}) = f(\mathbf{W}_\theta | \mathbf{X}) \star f(\mathbf{W}_\theta | \mathbf{z})$$

Wang et al. (2019) dicen que  $g(\cdot)$  genera un *response of a candidate window* (RoW), en que  $g^k(\mathbf{W}_\theta | \mathbf{X}, \mathbf{z})$  es la similaridad entre  $\mathbf{z}$  y la  $k$ -ésima sub imagen en  $\mathbf{X}$ ; incorporando una función de pérdida logística  $L_{sim}$ , inversa de la función de enlace *logit* en los modelos lineales generalizados. De esta manera, estima la posición del objetivo en la sub-región con el valor máximo en el RoW. Donde  $1 \leq k \leq \dot{w} \times \dot{h}$ , con  $\dot{w}$  la cantidad de anchos y  $\dot{h}$  de alturas usados en cada RoW.



**Figura 3-4.:** Arquitectura SiamFC, (Fiaz et al. 2019, Figura 3)

### SiamRPN

Adicional a la ANN SiamFC, incorpora una RPN que busca la posición del objetivo, estimando cajas delimitadoras con el criterio de estimación  $smoth L_1$  (3-7) y, en paralelo estima sus correspondiente probabilidades de encontrar el objetivo en el RoW usando la función de pérdida entropía cruzada, 3-8. Dichos criterios están dados por:

$$L_{box} = S_{L_1}(x) = \begin{cases} 0.5x^2 & \text{si } |x| < 1 \\ |x| - 0.5 & \text{en otro caso (e.o.c.)} \end{cases} \quad (3-7)$$

$$L_{score} = C_E(p, q) = \begin{cases} -\sum_{x \in \mathcal{X}} p(x) \log(q(x)) \\ -\int_{\mathcal{X}} P(x) \log(Q(x)) d\tau(x) \end{cases} \quad (3-8)$$

Con  $p(\cdot)$  y  $p(\cdot)$  distribuciones de probabilidad discretas,  $P(\cdot)$  y  $Q(\cdot)$  funciones de densidad de probabilidad con respecto a una medida  $\tau$ .

### SiamMask

Debido a que en la mayoría de métodos de VOT generan una representación de baja confianza sobre el objetivo, en este modelo se da importancia a producir una máscara de segmentación binaria por fotograma. Por ello, además de estimar la probabilidad de presencia del objetivo y una caja delimitadora, se busca estimar la probabilidad de que un píxel pertenezca al objeto buscado. Por lo cual, se extiende la red incorporando dos capas adicionales de neuronas  $h(\mathbf{W}_\phi)$  con  $\mathbf{W}_\phi$  los parámetros a estimar. De esta manera, si  $m_k$  es la máscara predicha entonces:

$$m_k = h(g^k(\mathbf{W}_\theta | \mathbf{X}, \mathbf{z}), \mathbf{W}_\phi)$$

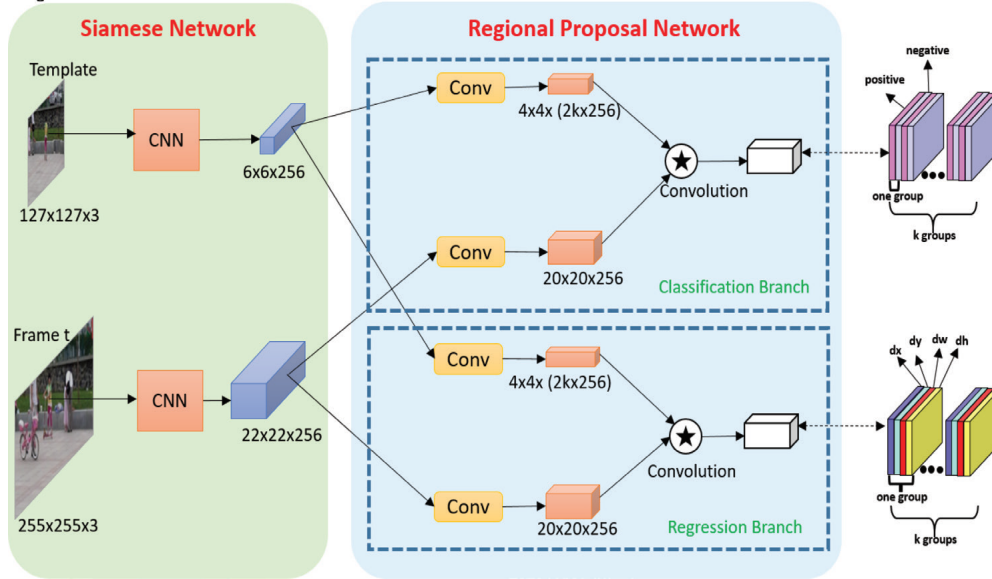


Figura 3-5.: Arquitectura SiamRPN, (Fiaz et al. 2019, Figura 4)

Por tanto, se puede observar que dada una diferente imagen de referencia  $\mathbf{z}$ , la ANN producirá una diferente máscara de segmentación  $m_n$ .

Para la estimación de los parámetros a cada RoW, le es asociada una etiqueta o variables respuestas  $y_n \in \{-1, 1\}$  y, una máscara  $\mathbf{c}_n$  de  $\dot{w} \times \dot{h}$ . Sea  $c_n^{(ij)} \in \{-1, 1\}$ , el valor correspondiente al píxel  $(i, j)$  de la máscara de la RoW  $n$ , entonces el criterio de estimación  $L_{mask}$  está dado por:

$$L_{mask}(\mathbf{W}_\theta, \mathbf{W}_\phi) = \sum_{k=1}^n \left( \frac{1 + y_n}{2\dot{w}\dot{h}} \sum_{i=1}^{\dot{w}} \sum_{j=1}^{\dot{h}} \log(1 + e^{-c_k^{(ij)} m_k^{(ij)}}) \right)$$

De esta forma,  $L_{mask}$  es similar a la función de pérdida en una regresión logística y,  $h(\mathbf{W}_\phi)$  son  $\dot{w} \times \dot{h}$  clasificadores.

Finalmente, Wang et al. (2019) proponen dos arquitecturas que se presentan en la figura 3-6, variando en la presencia o ausencia de  $b_\sigma$ , con criterios de estimación:

$$L_{2B} = \lambda_1 L_{mask} + \lambda_2 L_{sim}$$

$$L_{3B} = \lambda_1 L_{mask} + \lambda_2 L_{score} + \lambda_3 L_{box}$$

Fijando  $\lambda_1 = 32, \lambda_2 = \lambda_3 = 1$

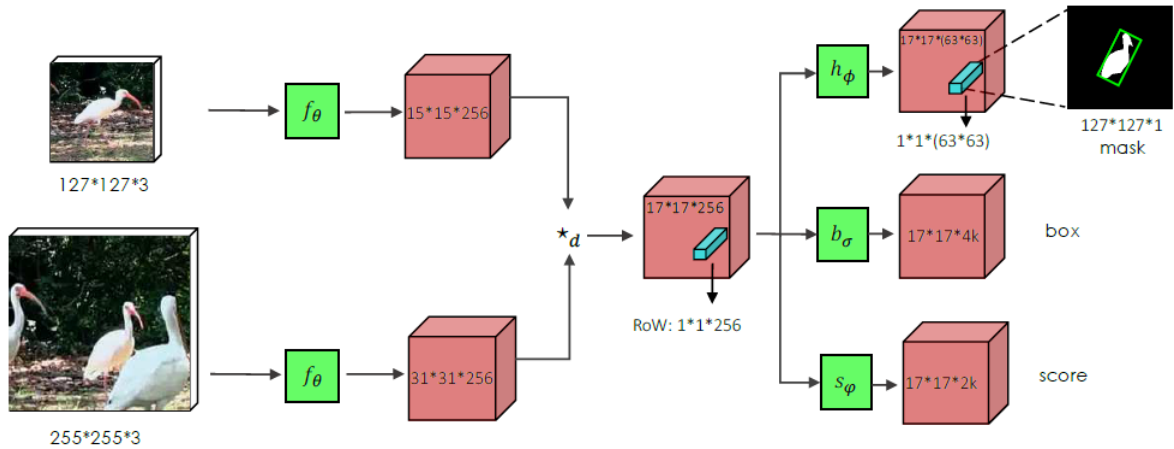


Figura 3-6.: Arquitectura SiamMask, (Wang et al. 2019, Figura 2).

### 3.3. Redes neuronales y la estadística

En Warren (1994) se puede ver la relación entre ANN y los modelos estadísticos, así como, parte de la brecha que los separa. De esta última, se resalta que la terminología de la literatura de NN es diferente a la usada en estadística. Por ejemplo, en ese contexto llaman a las variables características, a las variables independientes entradas, a las variables dependientes objetivos, a los residuos errores, al proceso de estimación entrenamiento, a un criterio de estimación función de pérdida, entre otros. Por otro lado, se afirma que los ingenieros en el campo de NN exponen sus construcciones como cajas negras que no requieren intervención del ser humano, siguiendo la idea de que los datos entran y predicciones son obtenidas.

Por el contrario, la estadística depende de la inteligencia humana para entender el fenómeno bajo estudio, proponer hipótesis y modelos, probar sus supuestos, diagnosticar sus problemas para explicar los datos y, reportar los resultados de forma comprensiva, explicando lo que se investiga. Por lo cual, las ANN no pueden reemplazar la metodología de la estadística, siendo esta última incorporada en ANN a través de criterios de estimación, algoritmos de optimización, intervalos de confianza, diagnóstico, métodos gráficos, intervalos de predicción, etc.

Partiendo de distinguir entre modelos de NN y algoritmos de NN, dichos modelos son una clase de regresiones no lineales, modelos discriminantes, modelos de reducción de datos o sistemas dinámicos no lineales. Dependiendo de como se formule su arquitectura, pueden ser similares a los modelos lineales generalizados, a la regresión polinómica, a la regresión no paramétrica, al análisis discriminante, al análisis de *clustering*, etc. Particularmente, si un perceptrón tiene función de activación lineal se está ante un modelo de regresión lineal, con la función logística la ANN es un modelo lineal generalizado (regresión logística) y con la

función umbral se tiene una función discriminante lineal.

Además, un MLP con una cantidad moderada de neuronas ocultas, puede considerarse un modelo semi-paramétrico y, si aumenta el número de neuronas considerablemente, se está ante un modelo no paramétrico, como es el caso de las CNN. Por ejemplo, en Cyganek (2013) se encuentra la formulación estadística del reconocimiento y seguimiento de objetos planteados en 3.1 y 3.2.

### 3.4. Modelos lineales generalizados

Debido a que la variable respuesta del presente trabajo es discreta, el modelo normal lineal puede ser inadecuado para analizar los datos. Por tanto, se plantea los modelos lineales generalizados (GLM) como alternativa que extienden el modelo lineal estándar, abarcando distribuciones de la variable respuesta no normales y, una posible función no lineal que enlaza la media de la variable respuesta con el predictor lineal (Agresti 2015). Los GLM tienen tres componentes:

- **Componente aleatoria:** especifica la distribución de la variable independiente que pertenece a la familia exponencial de dispersión y, supone que las  $n$  observaciones  $\mathbf{y} = (y_1, \dots, y_n)^t$  son realizaciones de las variables aleatorias independientes  $Y_1, \dots, Y_n$ .
- **Componente sistemática o predictor lineal ( $\eta_k$ ):** combinación lineal de  $p$  covariables medidas a cada unidad experimental  $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})^t$  y un vector de parámetros desconocidos y fijos  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ , con  $k = 1, \dots, n$ ;  $\eta_k = (\mathbf{x}_k^t \boldsymbol{\beta})$ .
- **Función de enlace:** función  $g(\cdot)$  estrictamente monótona y doblemente diferenciable que, enlaza la componente aleatoria con la sistemática, vinculando la media de la variable respuesta  $\mu_k = E(Y_k)$  con el predictor lineal  $\eta_k$ ; es decir,  $\eta_k = g(\mu_k)$ . Por tanto, la media de la variable respuesta depende funcionalmente de las covariables.

**Definición 3.4.1.** Familia exponencial de dispersión:

La distribución de la variable aleatoria  $Y_k$  pertenece a la familia exponencial de dispersión (FED) si su función de densidad o función de probabilidad se puede escribir como:

$$f(y_k; \mu_k, \phi/m_k) = e^{\left\{ \frac{m_k}{\phi} [y_k \theta_k - b(\theta_k) + c(y_k, \phi/m_k)] \right\}}$$

Para los modelos de respuesta binaria  $Y_k \sim \text{Bernoulli}(\mu_k)$  con función de probabilidad:

$$f(y_k; \mu_k, \phi/m_k) = \mu_k^{y_k} (1 - \mu_k)^{1-y_k} I_{\{0,1\}}(y_k)$$

Por tanto, la distribución Bernoulli pertenece a la familia exponencial de dispersión con  $\theta_k = \ln\left(\frac{\mu_k}{1-\mu_k}\right)$ ,  $b(\theta_k) = \ln(1 + e^{\theta_k})$ ,  $m_k = \phi = 1$  y  $c(y_k, \phi/m_k) = 0$ . Así, el modelo puede ser escrito como:

$$\begin{cases} Y_k \sim \text{Bernoulli}(\mu_k) \\ g(\mu_k) = \eta_k = \mathbf{x}_k^t \boldsymbol{\beta} \\ Y_1, \dots, Y_n \text{ independientes.} \end{cases}$$

Eligiendo  $Y_k \sim FED$  se tienen propiedades como la existencia de la estadística suficiente  $\sum_{k=1}^n y_k$  para sus parámetros, bajo condiciones de regularidad  $E(Y_k) = b'(\theta_k)$  y  $\text{var}(Y_k) = \frac{\phi}{m_k} b''(\theta_k) = \frac{\phi}{m_k} V(\mu_k)$ ; con  $V(\mu_k)$  la función de varianza.

Los parámetros del modelo son estimados vía máxima verosimilitud, con propiedades deseables como consistencia, eficiencia, normalidad asintótica e invarianza funcional. Bajo el supuesto de independencia, en los GLM la función de verosimilitud está dada por:

$$L(\boldsymbol{\beta}) = \prod_{k=1}^n \exp\left\{ \frac{m_k}{\phi} [y_k \theta_k - b(\theta_k) + c(y_k, \phi/m_k)] \right\}$$

Sin embargo, maximizar  $L(\boldsymbol{\beta})$  es equivalente a maximizar su logaritmo  $\ell(\boldsymbol{\beta}) = \ln(L(\boldsymbol{\beta}))$  donde:

$$\ell(\boldsymbol{\beta}) = \sum_{k=1}^n \frac{m_k}{\phi} [y_k \theta_k - b(\theta_k) + c(y_k, \phi/m_k)]$$

En general, debido a la no linealidad de la función de enlace,  $\hat{\boldsymbol{\beta}}$  no puede estimarse de forma cerrada, por lo cual se debe utilizar métodos de optimización, expuestos en la subsección 3.1.3. Para esta familia de modelos se utilizan frecuentemente los algoritmos Newton-Raphson y Scoring de Fisher. En el contexto de la ecuación 3-6, el algoritmo Newton-Raphson se obtiene con  $\alpha = \eta_r = 1$  y  $B_r = (\nabla^2 \ell(\boldsymbol{\beta}))^{-1}$  ( $B_r$  es la inversa de la matriz hessiana de segundas derivadas de  $\ell(\boldsymbol{\beta})$  con respecto a  $\boldsymbol{\beta}$ ). De forma similar, Scoring de Fisher es obtenido tomando  $\alpha = \eta_r = 1$  y  $B_r = [E(\nabla^2 \ell(\boldsymbol{\beta}))]^{-1}$  ( $B_r$  es la inversa de la matriz de información de Fisher).

Se puede ver que:

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{k=1}^n \frac{m_k}{\phi} \frac{(y_k - \mu_k)}{V(\mu_k) g'(\mu_k)} x_{kj}$$

$$\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_{j'}} = \sum_{k=1}^n \frac{m_k x_{kj'}}{\phi} \left[ (Y_k - \mu_k) \frac{\partial f_k}{\partial \beta_j} - \frac{x_{kj}}{V(\mu_k) [g'(\mu_k)]^2} \right]$$

$$E \left( \frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_{j'}} \right) = \sum_{k=1}^n - \frac{m_k x_{kj'}}{\phi} \frac{x_{kj}}{V(\mu_k) [g'(\mu_k)]^2}$$

Donde  $f_k = (V(\mu_k) g'(\mu_k))^{-1}$ . Con ello, se tienen los insumos para hallar  $\hat{\boldsymbol{\beta}} = \arg \max_{\boldsymbol{\beta}} (\ell(\boldsymbol{\beta}))$ .

Complementando, en el capítulo 4 de Agresti (2015) se puede ver el detalle de la estimación de  $\boldsymbol{\beta}$ , sus propiedades (consistencia, distribución asintótica, etc), criterios de bondad de ajuste, selección y diagnóstico del modelo. Además, se dedica el capítulo 5 a modelos de respuesta binaria asunto del presente trabajo.

## 4. Resultados

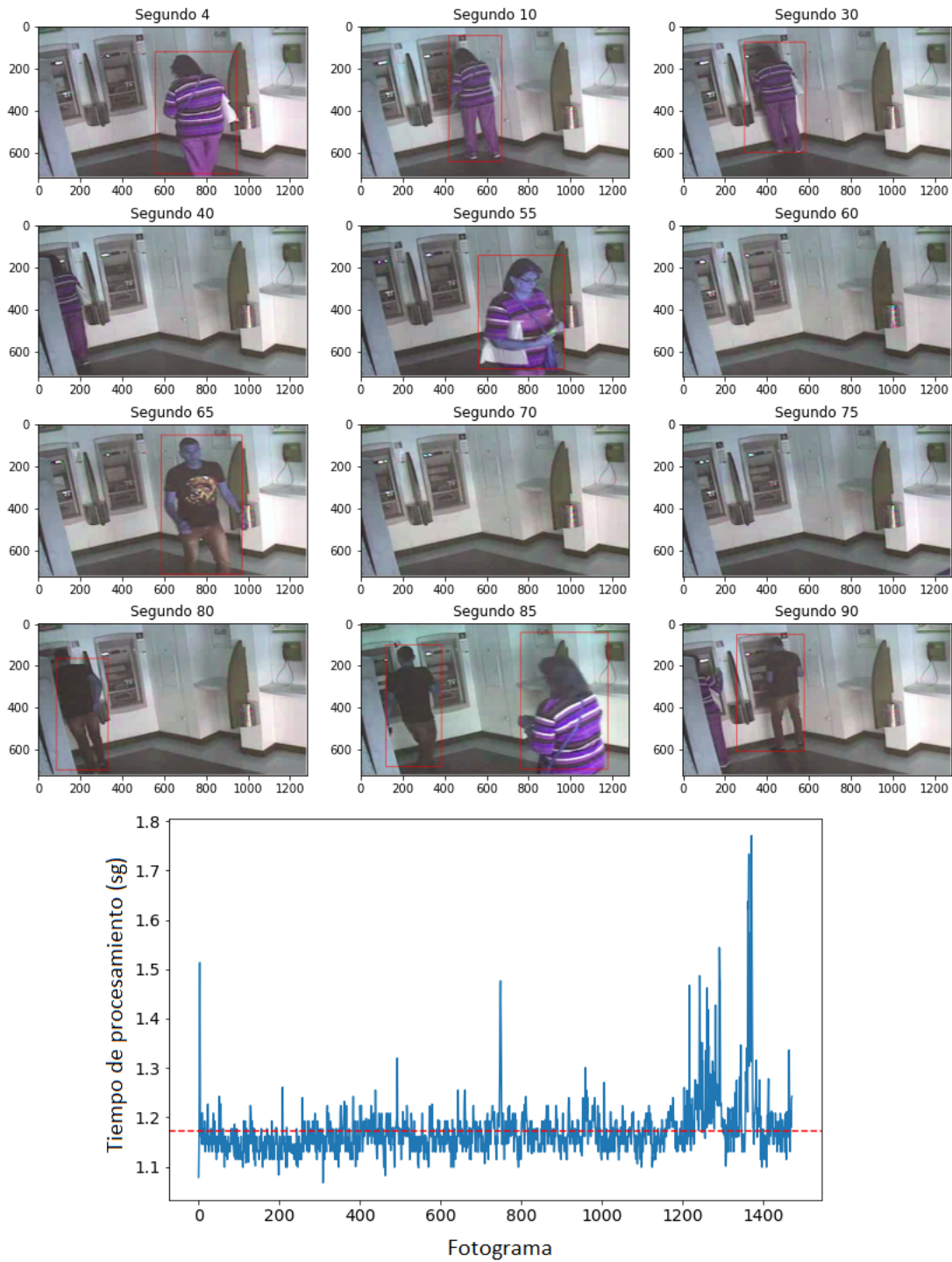
En esta sección se presentan los resultados obtenidos del tratamiento de los datos, realizando un recorrido a través de la metodología planteada en la sección 2.5. Los datos son procesados con una máquina equipada con procesador *Intel Core i5* de séptima generación, memoria RAM (*Random Access Memory*) de 20 GB (*Gigabytes*) DDR4 (*Double Data Rate Fourth Generation*), disco duro de estado sólido, tarjeta gráfica *Intel HD Graphics 620* con una velocidad máxima de 1050 MHz (*Megahertz*) y capacidad de 32 GB. Wang et al. (2019) realizaron su trabajo en GPU (*Graphics Processing Unit*), con una tarjeta gráfica *Nvidia GeForce RTX 2080* a 1800 MHz y capacidad de 616 GB. De esta manera, la GPU disponible no es adecuada, por lo cual, se decide realizar el entrenamiento del modelo en CPU *Central Processing Unit*.

El software utilizado para la detección, rastreo de objetos y transformación de datos es Python 3.7, referencia en el desarrollo e implementación del aprendizaje de máquina, con el uso principal de las librerías Tensorflow 1.14, Keras 2.3 y OpenCV 4.4. Cerrando el proceso identificando, estimando y analizando el modelo lineal generalizado por medio del software estadístico R Project.

### 4.1. Reconocimiento de objetos

Haciendo uso de las redes neuronales convolucionales se realiza la detección de los usuarios con el modelo *faster\_rcnn\_inception\_v2\_coco* desarrollado en Tensorflow y disponible en OpenVIVO (2020). Este modelo, como mencionan Ren, He, Girshick & Sun (2017) hace parte del estado del arte de las ANN para detección de objetos (OR), con algoritmos que encuentran regiones prometedoras de búsqueda. En este caso una red de propuestas regionales (RPN) y, que se combina con la red Fast R-CNN que reduce el tiempo de detección; logrando procesar 5 fotogramas por segundo con una arquitectura VGG-16.

En la figura 4-1, se aprecian los resultados de un caso particular de realización de fraude, del cual se pueden identificar las personas que transitan en el lugar con la resolución y calidad de vídeo disponibles. En general, se producen detecciones satisfactorias, sin embargo, se evidencia la dificultad de identificación de los usuarios debido a obstáculos. Del despliegue del modelo, se logra procesar en CPU en promedio una imagen cada 1.17 segundos. A continuación, en la figura 4-1 se presenta el tiempo de procesamiento de un fragmento de



**Figura 4-1.:** Aplicación: detección de objetos - Faster R-CNN

vídeo de aproximadamente un minuto procesado a  $24 \text{ ft/sg}$  .

## 4.2. Rastreo de objetos

Una vez detectado el objeto, se tiene el punto inicial para realizar su rastreo por medio de la red SiamMask. Este modelo, a diferencia de la CNN, propone un cuadrilátero como región delimitadora del objetivo, con lados que pueden ser no paralelos a los ejes del plano cartesiano. En el caso particular expuesto, hacia el segundo 4 del vídeo se realiza la detección plena del usuario del cajero con Faster R-CNN y, posteriormente el respectivo seguimiento hasta que sale del plano de grabación de la cámara hacia el segundo 60, ver figura **4-2**. Con respecto al desempeño, en la figura **4-3** se aprecia el tiempo medio de procesamiento de 1.2 segundos por fotograma -línea roja- y alcanzando la identificación de la persona camino al cuarto segundo sobre el fotograma 83. Posteriormente, SiamMask rastrea la persona logrando un desempeño con una media 31.6 % menor comparada con la CNN (0.82 *ft/sg*), línea verde. Luego, se aprecia el decrecimiento de la probabilidad de encontrar el objeto rastreado hacia el fotograma 1250 ante la salida de la persona del plano de la vídeo grabadora, proponiéndose un criterio de parada que ante la ausencia de la persona rastreada finalice su búsqueda.

Complementariamente, en la figura **4-2** se ilustra parte del recorrido de la persona, visto a través del centroide de su región delimitadora. Partiendo de ello, surge la necesidad de suavizar la curva formada por la trayectoria, con el fin de construir adecuadamente las covariables para el modelo de predicción. Posteriormente, combinando el despliegue del reconocimiento de objeto y, su rastreo con su criterio de parada y, la programación de algoritmos auxiliares; se construye una base de datos estructurada. En ella se almacena la ubicación de cada individuo, con asignación de una identificación única a cada persona detectada en los vídeos, hasta un máximo de 4 personas, debido a que de la naturaleza del patrón de fraude son atípicas las realizaciones del mismo con la presencia de 5 personas o más en el lugar. Se presenta una vista previa de los datos extraídos de las grabaciones en la tabla **4-1**.

Una vez combinados los procedimientos, para un vídeo de 5 minutos el tiempo de procesamiento esta alrededor de:

- 1 hora y 38 minutos a 24 *ft/sg*, reconocimiento cada 24 *ft* y rastreo continuo (100 vídeos en una semana).
- 50 minutos a 12 *ft/sg*, reconocimiento cada 12 *ft* y rastreo continuo.
- 25 minutos a 6 *ft/sg*, reconocimiento cada 6 *ft* y rastreo continuo.

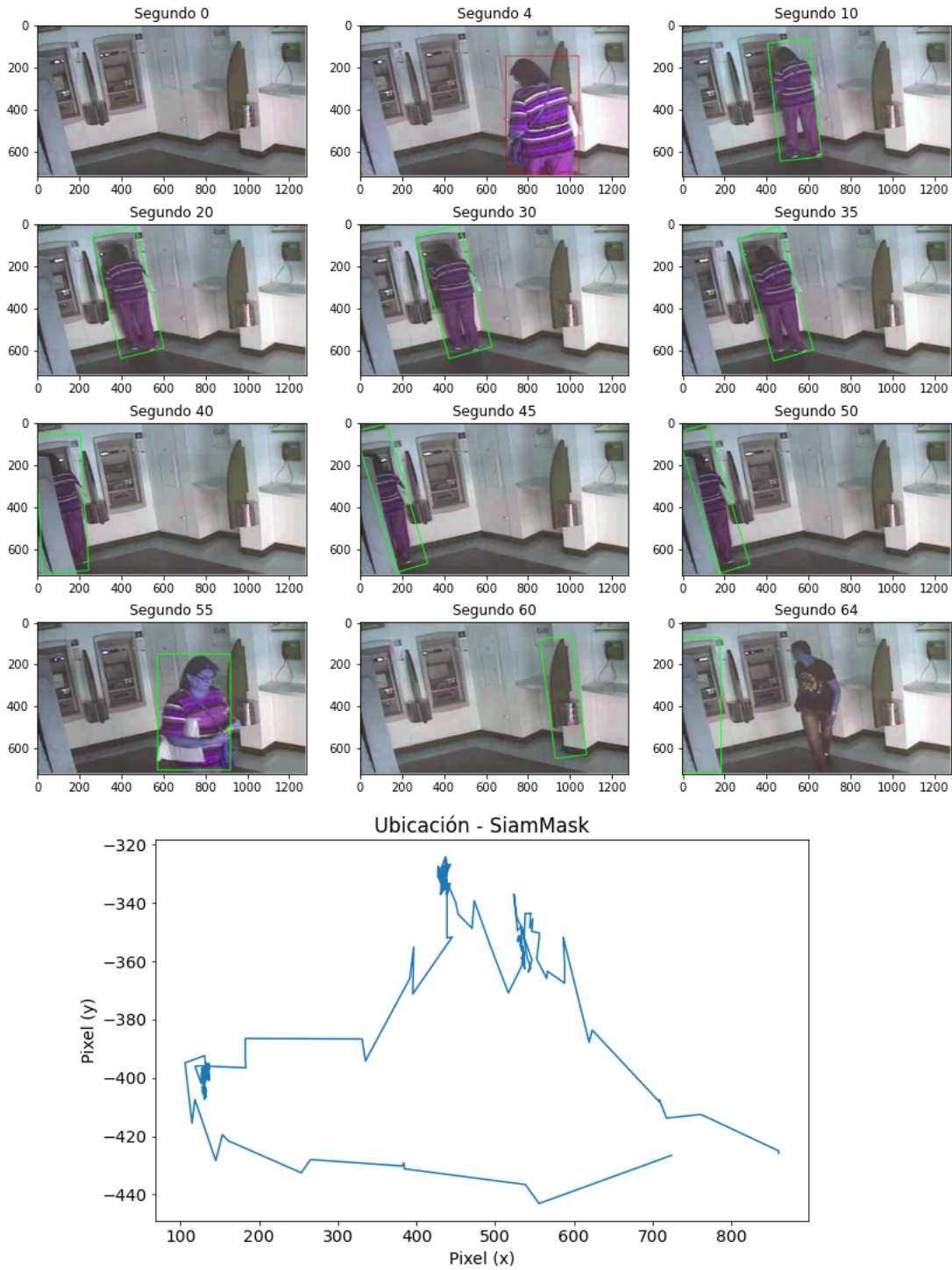
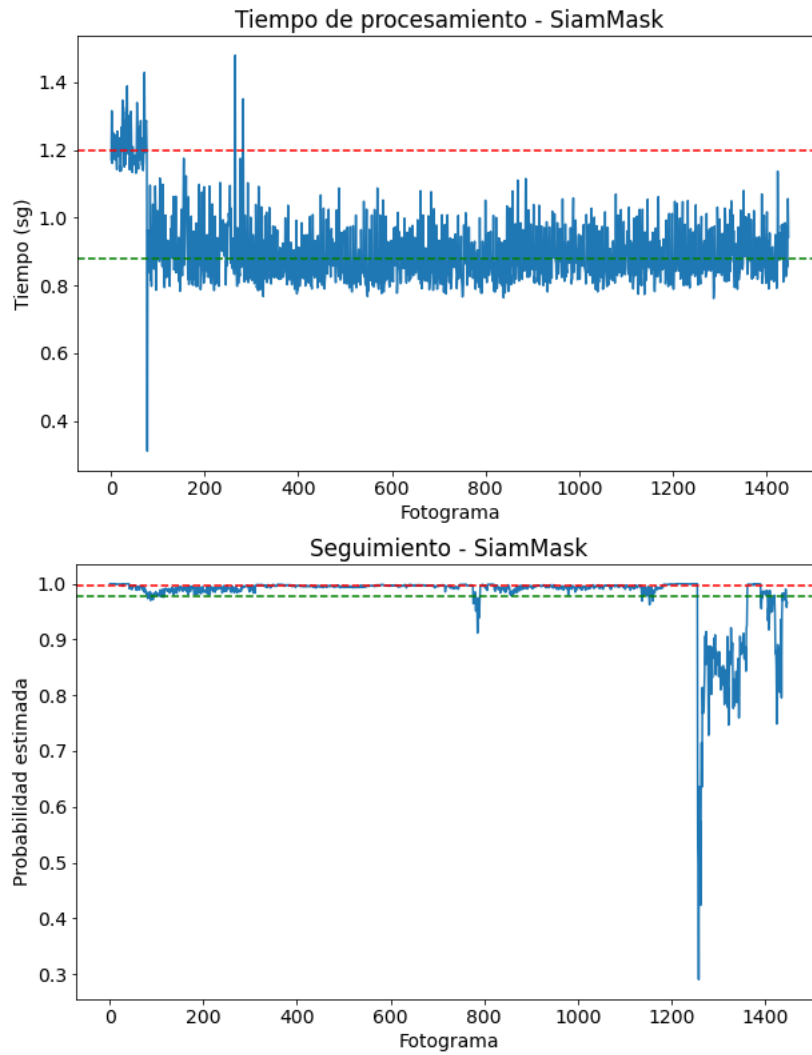


Figura 4-2.: Aplicación: rastreo de objetos - SiamMask



**Figura 4-3.:** Aplicación: desempeño - SiamMask(CPU)

Frame	$x_1$	$y_1$	$x_2$	$y_2$	$x_3$	$y_3$	$x_4$	$y_4$
0								
1								
2								
3								
4	136	250						
5	177	257						
6	236	261	225	309				
7	295	267	261	305				
8	337	275	314	302				
9	361	283	356	299				
10	400	288	410	296				
11	449	285	451	293				
12	500	283	487	291				
13	559	279	539	288				
14	614	276	560	284				
15	674	274	610	281				
16	697	271	654	278				
17	737	277	675	275				
18	796	274	728	273				
19	846	272	763	270				
20	880	269	816	267				
21	901	266	837	272				
22	940	263	875	270				
23	987	260	930	267				
24	1013	256	964	263				
25	1066	253	1011	259				
26	1090	259	1061	256				
27	1109	244	1094	253				
28	1152	248	1152	250				
29	1149	249	1150	257				
30	1153	256	1154	247	198	949	294	931
31	1150	249	1148	249	249	940	333	923
32	1160	256	1150	256	321	923	378	906
33	1145	250	1158	250	404	910	439	891
34	1149	252	1151	257	438	900	510	875
35	1155	249	1162	250	520	886	584	862

**Tabla 4-1.:** Estructuración de base de datos

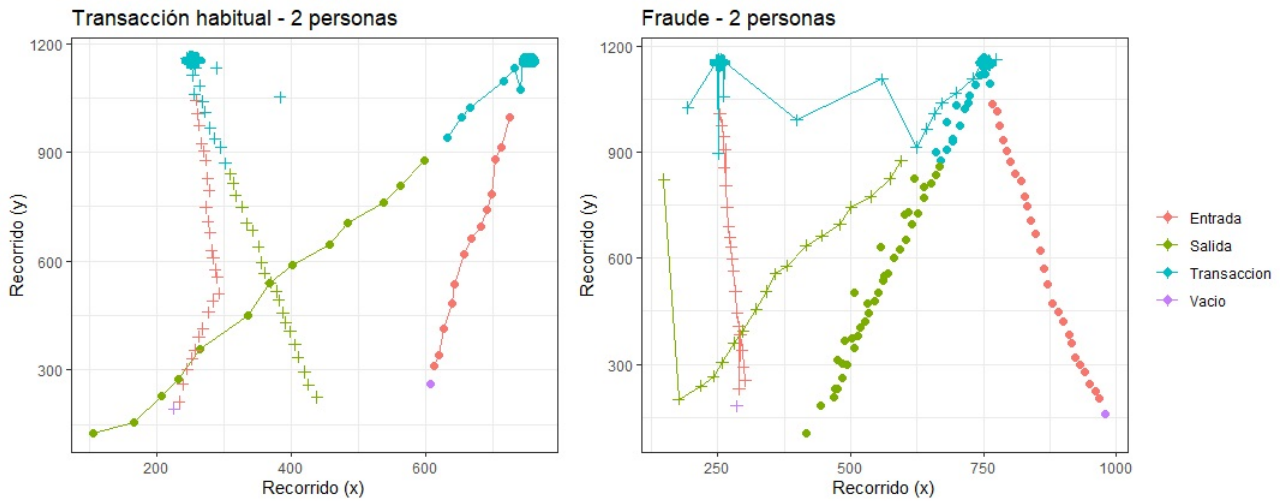
### 4.3. Construcción de variables y análisis descriptivo

Siguiendo la metodología planteada, en función de los recorridos de cada vídeo se programan y construyen las siguientes variables:

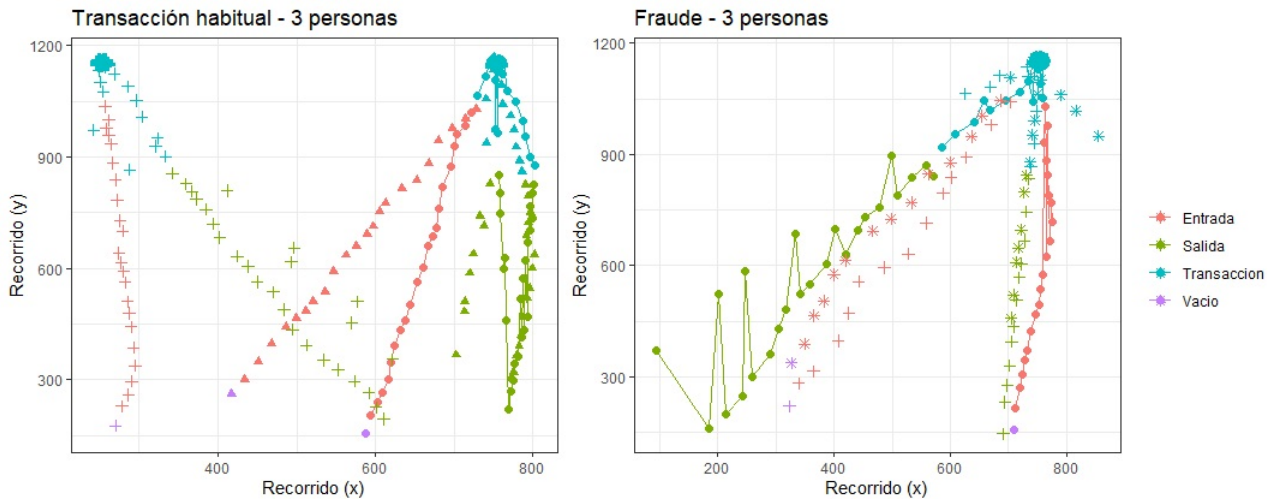
- Distancia recorrida: usuario 1, usuario 2, usuario 3, usuario 4, máxima .
- Tiempo desde el inicio de operaciones en el cajero: usuario 1, usuario 2, usuario 3, usuario 4, máximo.
- Distancia del usuario  $i$  al usuario  $j$ , para  $i, j = 1, 2, 3, 4$ .
- Distancia del usuario  $i$  al cajero  $k$ , para  $k = 1, 2, 3$ .
- Variables indicadoras de si el usuario  $i$  pasa por el cajero  $k$ .
- Variable indicadoras de si el usuario  $i$  pasa por al menos 2 cajeros .
- Variables indicadoras de si el usuario  $i$  una vez paso por al menos un cajero, inicia la salida del lugar.
- Variables indicadoras de regreso, es decir, si el usuario  $i$  una vez inicia la salida del lugar regresa a por lo menos un cajero.
- Distancia del usuario  $i$  al usuario  $j$ , para  $i, j = 1, 2, 3, 4$ .
- Variables indicadora de si el usuario  $i$  acompaña al usuario  $j$  en función de la distancia entre ellos.
- Identificación de la etapa de la transacción en función de la distancia a los cajeros (entrada, transacción, salida).
- Variables indicadora de si el usuario  $i$  realiza una irrupción al usuario  $j$  en función del porcentaje de acompañamiento entre ellos y la etapa de la transacción.

Dado lo anterior, en la figura 4-4 se presentan un caso de fraude y un proceso habitual de uso del cajero. En el caso de fraude, el usuario identificado con el simbolo +, una vez inicia las operaciones en la máquina presenta un regreso (intento de salida del lugar), aproximadamente en la cordenada (80, 800), con el posterior paso a un segundo cajero y, la intersección de su recorrido con un segundo usuario; concretando el patrón habitual del fenómeno. Por el contrario, los recorridos habituales tienen trayectorias sin irrupciones y con una baja intersección de los mismos.

Ahora, en la figura 4-5 se observa un fraude de comportamiento atípico en el que el sujeto activo del fraude aborda desde la entrada del lugar a la víctima y su acompañante, haciendo



**Figura 4-4.:** Recorridos según la etapa de la transacción - 2 usuarios

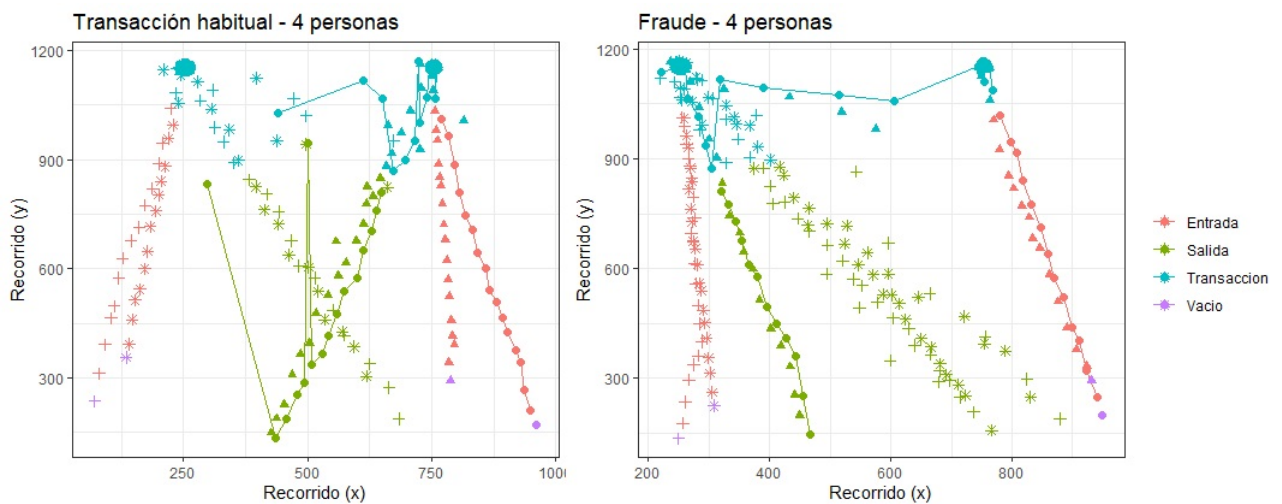


**Figura 4-5.:** Recorridos según la etapa de la transacción - 3 usuarios

uso de un solo cajero para cometer el ilícito. Por otro lado, en los recorridos de las figuras 4-4 a 4-6, se observan reconocimientos de los usuarios alejados de su trayecto respectivo, debido a un mayor cambio del centroide de la región delimitadora, hecho que aumenta la distancia recorrida.

De la misma manera, que en la figura 4-4, los recorridos en la ilustración 4-6 con la presencia de 4 usuarios, muestran como en el fraude el patrón se da mediante el paso por más de un cajero y, la irrupción de los sujetos activos del fraude a su víctima, generando una intersección

en los trayectos. En el escenario sin ilícito, ante de la presencia de una mayor cantidad de personas y, la obstrucción de las mismas en el plano de grabación, ocurren puntos en el tiempo donde el rastreador se confunde entre las personas presentes, materializándose en observaciones que se alejan abruptamente de su trayecto y, que además se acercan al recorrido de otro usuario; lo cual sugiere, la técnica de rastreo disminuye su precisión a medida que aumenta el número de objetos a procesar; sin embargo, al desaparecer el solapamiento de los usuarios en el plano de grabación, el rastreador realiza adecuadamente la detección, continuando con la estimación adecuada de las trayectorias.



**Figura 4-6.:** Recorridos según la etapa de la transacción - 4 usuarios

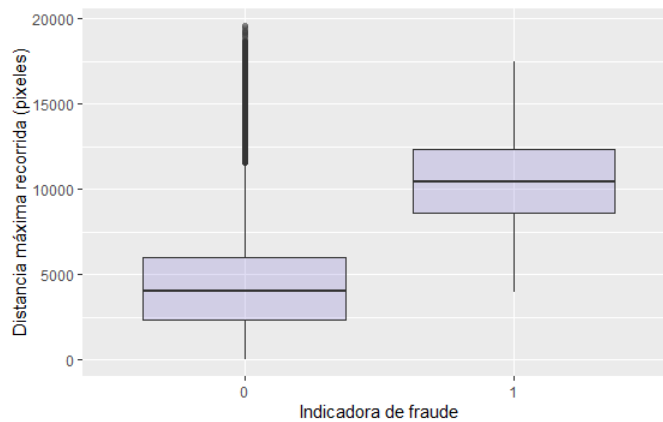
Ahora, la prevalencia de fraude para fotografías con a los sumo 3 personas es al menos el doble comparado con la presencia de 4 personas o más, hecho que sugiere que la probabilidad de fraude decrece a medida que el número de usuarios aumenta. Aquellos fraudes con un usuario, corresponden a fotografías de tramos avanzados del patrón de fraude, es decir, ya se encuentran materializados; por lo cual, se espera que este patrón de fraude se ejecute ante la presencia de 2 o 3 usuarios en el plano de grabación. Ver tabla 4-2.

Usuarios	0	1	2	3	>3
Video	0.00 %	0.00 %	11.20 %	34.00 %	21.20 %
Fotogramas	0.00 %	9.94 %	6.71 %	10.40 %	3.47 %

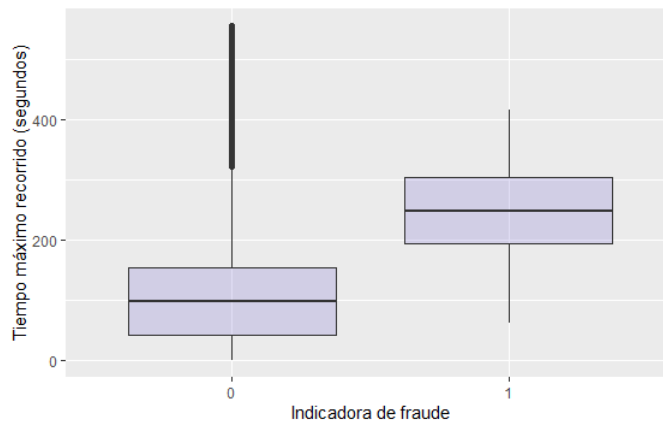
**Tabla 4-2.:** Distribución de fraude dado el número de usuarios

Para los usuarios detectados se calcula la distancia recorrida sobre el plano de grabación, en particular en la figura 4-7, se observa la distribución de la distancia máxima recorrida dado si es o no fraude, de la cual se observa que la distancia en píxeles para una situación de

fraude es en general mayor. Lo anterior debido a que el percentil 75 de fotogramas asociados a transacciones normales es inferior al percentil 25 de escenarios de fraude, los cuales representan el 7.84 % de los fotogramas. De la misma manera, la afirmación se mantiene para el tiempo máximo en transacciones dado el fraude (figura 4-8).



**Figura 4-7.:** Distancia máxima recorrida dado el fraude.



**Figura 4-8.:** Tiempo máximo de transacciones dado el fraude.

Además, se identifica si alguno de los usuarios ha pasado por lo menos a 2 cajeros, que en el evento de fraude sucede en 92.50 % de los casos, contra un 6.04 % en los escenarios de no ocurrencia del ilícito. También, se busca detectar el patrón de regreso (ejecución de operaciones seguido de desplazamiento hacia la salida y regreso a realizar transacciones de nuevo), que se efectúa en un 79.20 % de los eventos de fraude contra un 7.25 % en transacciones habituales. Complementando, se define y reconoce el evento de irrupción (en función del porcentaje de acompañamiento, etapa de la transacción y distancia entre usuarios) como el abordaje del

usuario por una persona que no lo acompaña; en los ilícitos el 35.40 % presenta irrupción, fenómeno que no ocurre en situaciones habituales (no fraude).

Fraude	Paso 2 o más cajeros	Regreso	Irrupción
0 (No)	6.04 %	7.25 %	0.00 %
1 (Si)	92.50 %	79.20 %	35.4 %

**Tabla 4-3.:** Indicadoras de patrón de regreso, irrupción y paso por 2 o más cajeros dado el fraude

El consolidado de las anteriores cifras se presentan en la figura 4-10, donde se puede adicionar que los eventos con patrón de regreso tienden a generar mayores distancias recorridas, así como tiempos en transacción; con distribuciones marginales unimodales, asimétricas y con posibles valores esperados diferentes dada la ocurrencia o no de fraude o, si presenta o no patrón de regreso. Descriptivamente, muestra la posibilidad de que el modelo lineal generalizado pueda tener dificultades para diferenciar las operaciones no ilícitas con patrón de regreso de los eventos de fraude.

## 4.4. Modelado

Con el objetivo de estimar la probabilidad esperada de fraude dadas las características observadas a través del vídeo, se estima un modelo lineal generalizado con funciones de enlace logit, probit y complemento log-log en el software estadístico R Project. Previamente, se particiona el conjunto de datos en 70 % para entrenamiento o estimación del modelo y el 30 % para su validación, bajo muestreo aleatorio simple estratificado por la variable respuesta.

Con ello, se realiza la identificación de la componente sistemática del modelo por medio de las metodologías *stepwise (backward, forward)*, siguiendo los procedimientos en Vanegas & Rondón (2018). Son seleccionadas las variables provistas en la tabla 4-4 mediante la función *step\_glm(.)* de los mismos autores, con el previo retiro de la covariable tiempo máximo en transacciones altamente correlacionada con la distancia; hecho que induce en la estimación de un parámetro negativo sobre el tiempo que se contradice con el patrón esperado de fraude (posible efecto de multicolinealidad). En adición, se retira la variable irrupción que resulta no significativa para el modelo con el valor  $p$  asociado menor al 5 % para la hipótesis nula  $H_0 : \beta_k = 0$ , ver Anexo A.

La matriz de correlación de las variables seleccionadas está dado por la tabla 4-4 en la que:

- $V_1$  : número de personas.

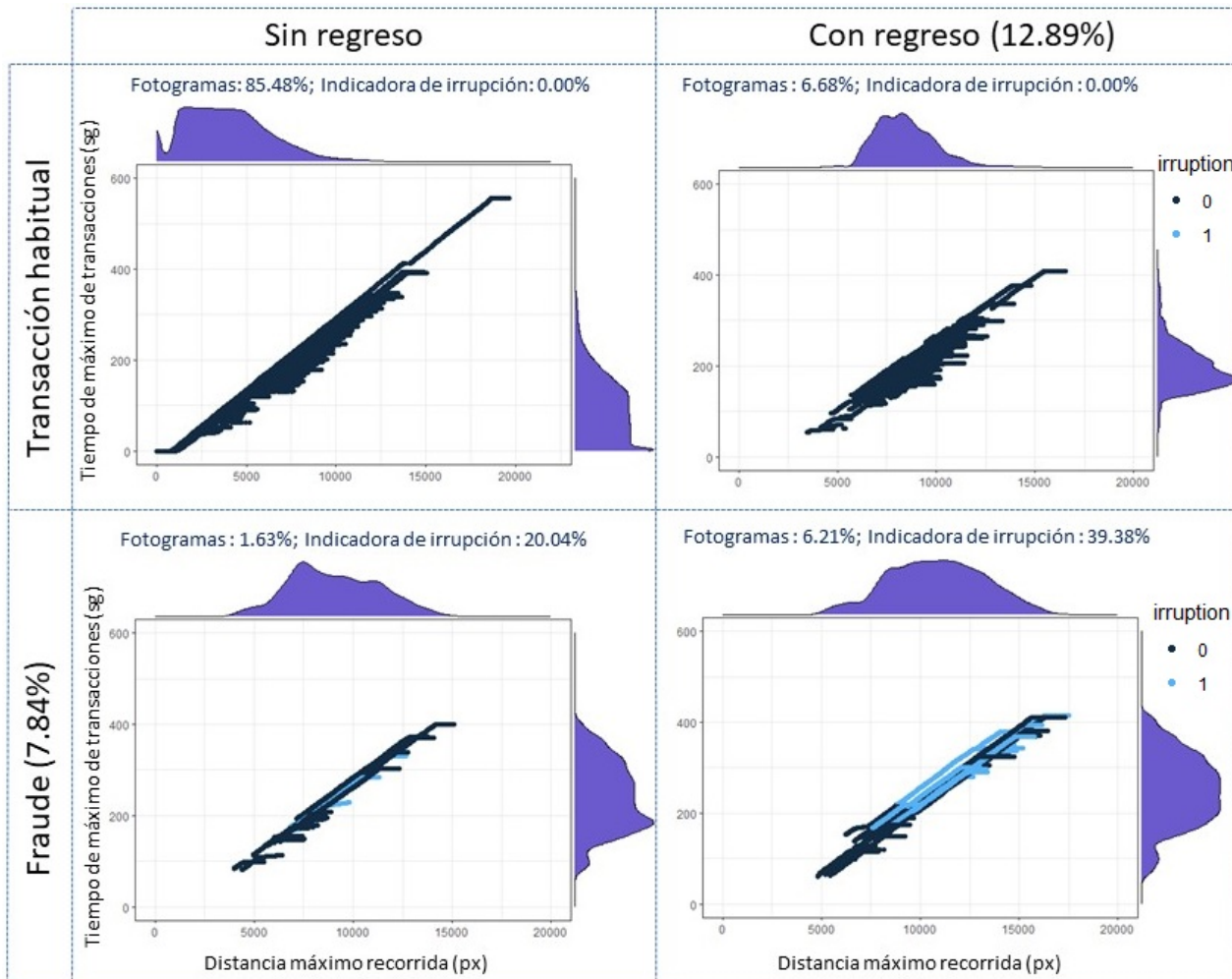


Figura 4-9.: Distribución de covariables dado el evento de fraude.

- $V_2$  : distancia máxima recorrida.
- $V_3$  : tiempo máximo recorrido.
- $V_4$  : paso por 2 o más cajeros.
- $V_5$  : patrón de regreso.
- $V_6$  : evento de irrupción.

La componente sistemática seleccionada para las funciones de enlace son expuestas en la tabla 4-5. Complementariamente, es seleccionada la función de enlace probit con menor AIC. Adicional a ello, del análisis descriptivo y entendimiento del patrón de fraude, se plantea que el efecto de la distancia máxima recorrida depende de si hubo o no al menos un regreso de un usuario. El modelo planteado esta dado por la ecuación  $refeq_{glm}$ .

	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$
$V_1$	1.0000	- 0.0189	- 0.0431	0.1413	- 0.0084	0.0464
$V_2$	- 0.0189	1.0000	0.9868	0.1692	0.5957	0.3246
$V_3$	- 0.0431	0.9868	1.0000	0.1231	0.4892	0.2858
$V_4$	0.1413	0.1692	0.1231	1.0000	0.2868	0.3800
$V_5$	- 0.0084	0.5957	0.4892	0.2868	1.0000	0.3459
$V_6$	0.0464	0.3246	0.2858	0.3800	0.3459	1.0000

**Tabla 4-4.:** Correlación de covariables

Función de enlace	Componente sistemática	AIC
Logit	Fraude $\sim V_2 + V_4 + V_5 + V_1$	52249
Probit	Fraude $\sim V_2 + V_4 + V_5 + V_1$	51916
Complemento log-log	Fraude $\sim V_2 + V_4 + V_5 + V_1$	53324

**Tabla 4-5.:** GLM: componente sistemática

$$\begin{cases} Y_k \sim \text{Bernoulli}(\mu_k) \\ \Phi(\mu_k) = \beta_0 + \beta_1 * V_2 + \beta_2 * V_5 + \beta_3 * V_4 + \beta_4 * V_1 + \beta_5 * V_2 * V_5 \\ Y_1, \dots, Y_n \text{ independientes.} \end{cases} \quad (4-1)$$

Bajo la hipótesis nula  $H_0 : \beta_5 = 0$ , es decir que el efecto sobre la probabilidad de fraude de la distancia máxima recorrida no depende de si existe o no el patrón de regreso; bajo el test de Wald el valor p asociado  $2 * 10^{-16}$  es menor al nivel de significancia (0.05), luego existe evidencia estadística para rechazar la hipótesis nula. Por tanto, el efecto de la distancia máxima recorrida depende de si existe o no el patrón de regreso. Ver figura **4-10**.

Dado que la significancia de los parámetros del modelo se pueden ver afectados por el número de observaciones (fotogramas) en el conjunto de datos de entrenamiento  $n > 200000$ , como validación final se realizan 100 muestras estratificadas por la variable respuesta para cada uno de los tamaños muestrales 500, 1000, 1500, ..., 10000 y, se evalúa la media y mediana del p valor asociado al test de Wald para cada iteración. En la figura **4-10**, se concluye que a partir de muestras de 2500 fotogramas (aproximadamente la indexación de 3 vídeos) se encuentra que se rechaza la hipótesis nula evaluada  $H_0 : \beta_5 = 0$ .

En el anexo A se presenta este procedimiento bajo el orden de selección de variables del método *backward*, confirmándose la estructura del modelo seleccionado, con la adición de un

ejemplo de un parámetro adicionado al modelo que resulta significativo con todos los datos, pero al realizar la simulación los p valores tienden a ser mayores al 5 %.

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.0319	-0.1470	-0.0741	-0.0385	3.4024

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercepto)	-3.231e+00	2.791e-02	-115.771	< 2e-16 ***
Distancia máxima recorrida	2.110e-04	2.828e-06	74.616	< 2e-16 ***
Patrón de regreso	-2.183e-01	4.621e-02	-4.724	2.31e-06 ***
Paso por 2 o más cajeros	2.896e+00	3.516e-02	82.378	< 2e-16 ***
Número de usuarios	-1.743e-01	9.021e-03	-19.321	< 2e-16 ***
Interacción: patrón de regreso y distancia máxima recorrida	1.488e-04	4.936e-06	30.143	< 2e-16 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 132523 on 243136 degrees of freedom  
Residual deviance: 50950 on 243131 degrees of freedom  
AIC: 50962

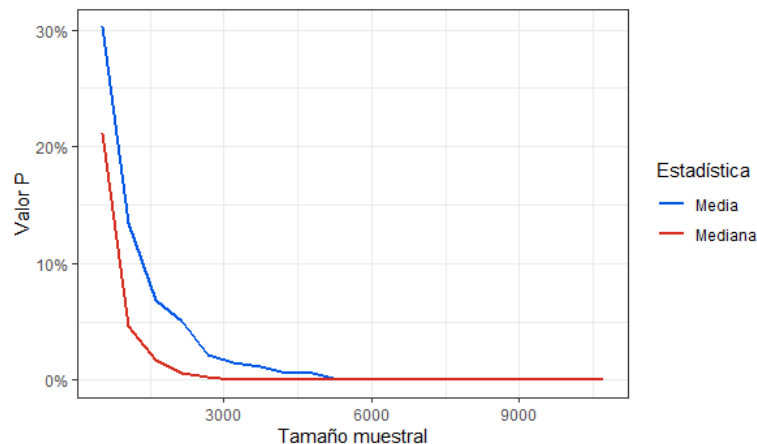


Figura 4-10.: Simulación hipótesis nula:  $\beta_5 = 0$

Se puede ver que si se supone el modelo bajo la hipótesis nula  $\beta_6 = 0$ , el estadístico de Wald con todos los datos de entrenamiento sugiere el efecto de la distancia máxima recorrida depende de si al menos un usuario para por más de un cajero, sin embargo la simulación planteada sugiere lo contrario, ver figura 4-11.

$$\begin{cases} Y_k \sim \text{Bernoulli}(\mu_k) \\ \beta_0 + \beta_1 * V_2 + \beta_2 * V_5 + \beta_3 * V_4 + \beta_4 * V_1 + \beta_5 * V_2 * V_4 \\ Y_1, \dots, Y_n \text{ independientes.} \end{cases}$$

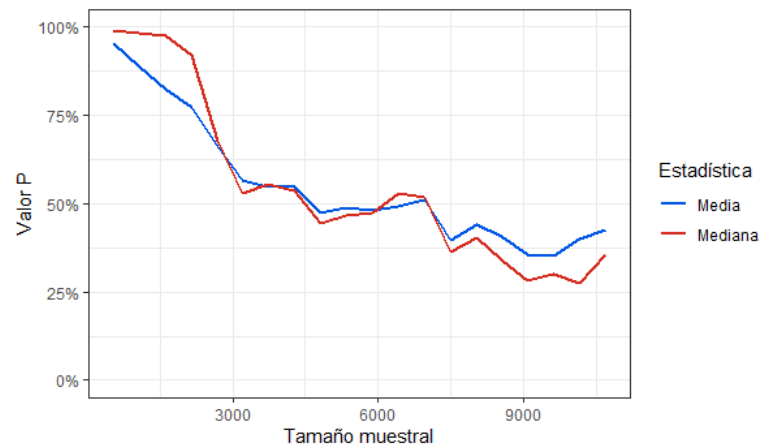


Figura 4-11.: Simulación hipótesis nula:  $\beta_6 = 0$

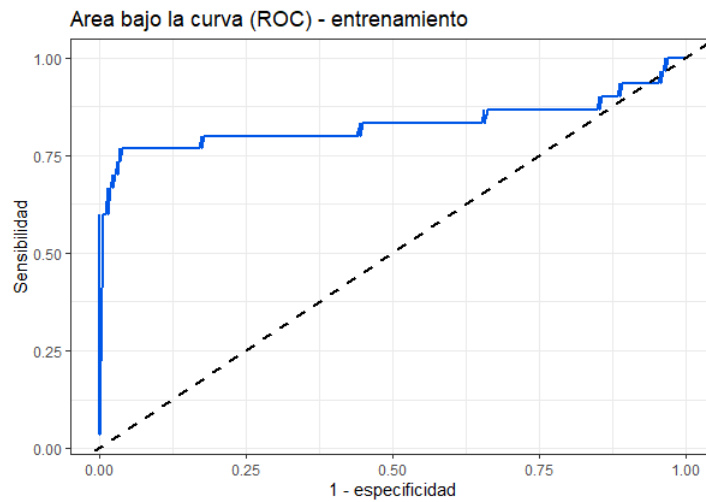
#### 4.4.1. Validación del modelo

Los estadísticos KS, Gini y AUC reflejan un buen poder discriminativo del modelo con resultados estables en el conjunto de datos de validación, sin presentar algún decrecimiento de los indicadores (table 4-6). Así mismo, la curva ROC construida mediante la función `roc.curve` escrita por Vanegas & Rondón (2018) relacionada en la figura 4-12, muestra un resultado adecuado alejándose del rendimiento del clasificador no informativo o al azar.

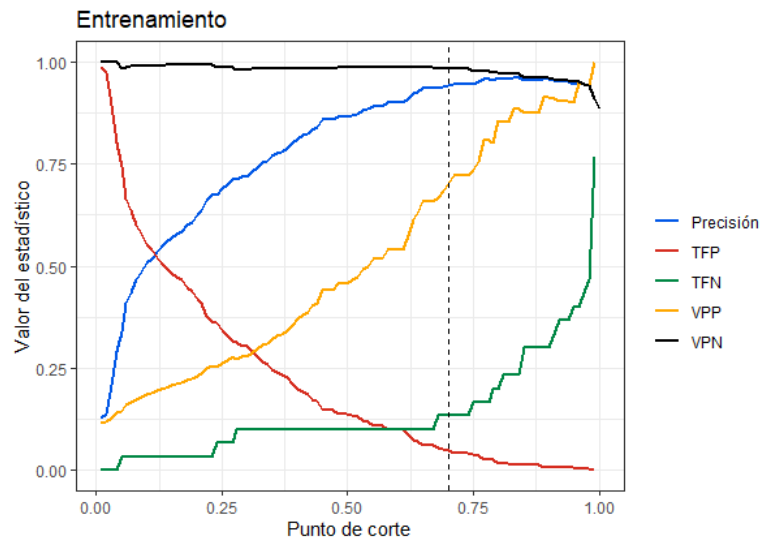
Ahora, para elegir el punto de corte óptimo se analiza la precisión, la tasa de falsos positivos (TFP), la tasa de falsos negativos (TFN), el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN). Como se puede observar en la figura 4-13, a medida que el punto de corte se acerca a uno, la TFP decrece y el VPP aumenta. Sin embargo, al superar 0.7 la TFN presenta una tendencia creciente. Seleccionándose como punto de corte 0.7 como aquel que balancea TFP, TFN y el VPP, adicionando alta precisión.

Estadística	Entrenamiento	Validación
KS	73.2	85.0
Gini	66.3	95.2
AUC	83.2	97.6

**Tabla 4-6.:** Estadísticos de poder discriminativo



**Figura 4-12.:** Curva ROC



**Figura 4-13.:** Punto de corte

Dado el punto de corte, en el conjunto de datos de entrenamiento y validación se tiene una precisión del 94.23 % y un 97.35 % respectivamente. Se espera además que, 13 de cada 100 fraudes no sean detectados (validación 8/100) y, que de cada 100 transcurso de transacciones estimadas como fraude 30 sean falsos positivos (14/100 validación). Complementando, con la sugerencia de alertar el 15 % de trayectorias. Ver tablas 4-7 y 4-8.

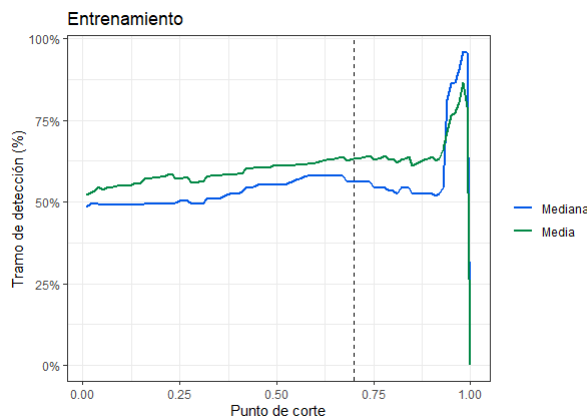
	Estimación							
	Matriz de confusión				Marginal fila		Marginal Columna	
	No	Si	No	Si	No	Si	No	Si
Fraude								
No	219	11	84.23 %	4.23 %	95.22 %	4.78 %	98.21 %	29.73 %
Si	4	26	1.54 %	10.00 %	13.33 %	86.67 %	1.79 %	70.27 %

**Tabla 4-7.:** Matriz de confusión entrenamiento

	Estimación							
	Matriz de confusión				Marginal fila		Marginal Columna	
	No	Si	No	Si	No	Si	No	Si
Fraude								
No	98	2	86.73 %	1.77 %	98.00 %	2.00 %	98.99 %	14.29 %
Si	1	12	0.88 %	10.62 %	7.69 %	92.31 %	1.01 %	85.71 %

**Tabla 4-8.:** Matriz de confusión validación

Finalmente, en la figura 4-14 se muestra que se espera detectar el fraude una vez transcurrido un 65 % del patrón de fraude, lo cual implica generación de alertas tempranas que ayuden a impedir la finalización del ilícito o en su defecto proteger los fondos de la víctima.



**Figura 4-14.:** Tramo de detección

# 5. Conclusiones y recomendaciones

## 5.1. Conclusiones

La metodología planteada permite transformar una fuente de información des-estructurado, como lo es el vídeo, en una base de datos estructurada. Detalladamente, las redes neuronales convolucionales y siamesas pre-entrenadas, realizan una detección y rastreo con la precisión suficiente para plasmar los trayectos de los usuarios. Identificando el número de personas en el plano de grabación y, detectando los recorridos alrededor de los cajeros automáticos.

Lo anterior, permite la formulación, construcción y extracción de características (covariables). Insumo que se orienta a la estimación y validación de un modelo lineal generalizado con respuesta binaria y función de enlace probit. El modelo estadístico facilita la adecuada estimación o pronóstico del patrón de fraude, con una precisión del 94.23 %, identificando más del 86 % de fraudes. Representando así, un caso de éxito de aplicación del modelo lineal propuesto. La ventaja del resultado obtenido al depender de la distancia recorrida, número de usuarios, y las variables indicadoras de paso por dos cajeros o más y, del patrón de retorno; es que facilita su aplicación a cualquier espacio donde se alojen cajeros automáticos, dado que se estandarice adecuadamente la información relacionada con la distancia recorrida.

Por consiguiente, la combinación de las técnicas de aprendizaje de máquina y modelos estadísticos, proporciona una solución asertiva, que desde el contexto de negocio induce una optimización del proceso de prevención de fraude. Puesto que, bajo la intervención actual, la entidad financiera monitorea visualmente los centros tecnológicos; hecho que implica gastos fijos y un proceso no optimizado de prevención de fraude. La solución planteada determina, de ser necesario, monitorear solo el 15 % de los escenarios de operaciones en cajeros, con la opción de priorización del control de fraude de acuerdo a la probabilidad de fraude estimada. Lo anterior, podría paulatinamente extenderse en un proceso automático, que impactará en una reducción en la afectación de los fondos del cliente. Lo anterior ayudará al Banco con una disminución a la exposición de materialización de riesgo reputacional, con la adición de reducción de gastos en capital humano que puede destinar a otras actividades que involucren mayor rentabilidad.

## 5.2. Recomendaciones

Partiendo del procesamiento medio por fotograma determinado en las figuras 4-1 y 4-2, se puede decir que por cada fotograma analizado, la presente metodología le toma alrededor de 2 segundos por imagen. Por lo cual, el proceso en general debe ser optimizado partiendo de reentrenar una red para la detección exclusiva de personas, con arquitectura símil al modelo Faster R-CNN que detecta más de 50 objetos diferentes. Repercutiendo en una reducción de tiempo máquina en la fase de detección.

Complementando, existe la opción de evaluar el rendimiento de la metodología planteada con una rata de  $ft/sg$  inferior ( $ft/sg \rightarrow 1$ ), con la finalidad de procesar la menor cantidad de imágenes posibles manteniendo las propiedades del modelo estimado.

En la misma línea, la programación orientada a la construcción de las covariables debe ser optimizada y, la solución expuesta en su totalidad debe ser transformada al procesamiento en GPU, que es natural para el procesamiento de imágenes.

Por otra parte, pueden explorarse otras técnicas diferente al modelo lineal generalizado que puedan mejorar las propiedades del resultado, en especial la razón de falsos descubrimientos,  $FP/(FP + VP)$ . La cual, actualmente se ubica en el 30 %. Además, se debe evaluar el efecto del desbalanceo, que en el conjunto de datos se acerca al 10 % la tasa de malos de vídeos; sin embargo, en la vida real es menor al 1 %.

Finalmente, la metodología actual puede ser complementada con la inclusión de identificación facial, que permita identificar la presencia de personas sospechosas de las listas negras de la entidad financiera relacionadas con casos de fraude.

# A. Selección de covariables

En el presente anexo se expone el resumen de los métodos empleados para identificar la componente sistemática del GLM, expuesto en la sección 4.4. Del conjunto de covariables plasmada en 4.3, bajo la metodología *stepwise* (forward) se muestra la abreviación para la identificación de la componente sistemática  $Fraude \sim V_2 + V_4 + V_5 + V_1$  de la siguiente forma:

Family: binomial  
 Link: probit

Initial model:  
 Fraude ~ 1

Step 0 :

	Df	AIC	BIC	Deviance+	Pearson <sup>^</sup>	p-value*
+ V_2	1	7.4330e+04	7.4351e+04	4.3910e-01	4.3730e-01	0.0000e+00
+ V_5	1	7.7349e+04	7.7370e+04	4.1640e-01	0.0000e+00	0.0000e+00
+ V_3	1	9.1216e+04	9.1237e+04	3.1170e-01	2.8110e-01	0.0000e+00
+ V_6	1	1.0187e+05	1.0189e+05	2.3130e-01	2.3200e-02	0.0000e+00
+ V_4	1	1.1254e+05	1.1256e+05	1.5080e-01	0.0000e+00	0.0000e+00
+ V_1	1	1.3247e+05	1.3249e+05	4.0000e-04	2.0000e-04	1.24e-13
<none>		1.3253e+05	1.3254e+05	0.0000e+00	0.0000e+00	

Step 1 : + V\_2

	Df	AIC	BIC	Deviance+	Pearson <sup>^</sup>	p-value*
+ V_3	1	51065.8364	51097.0406	0.6147	0.5546	0.0000
+ V_4	1	59961.1390	59992.3431	0.5476	0.5428	0.0000
+ V_5	1	62365.8005	62397.0046	0.5294	0.4008	0.0000
+ V_6	1	63860.5584	63891.7626	0.5182	0.4912	0.0000
+ V_1	1	73647.7713	73678.9754	0.4443	0.4262	7.087e-151
<none>		74330.2523	74351.0551	0.4391	0.4373	

Step 2 : + V\_3

	Df	AIC	BIC	Deviance+	Pearson <sup>^</sup>	p-value*
+ V_6	1	43827.6578	43869.2633	0.6693	0.5756	0.0000
+ V_4	1	44084.6011	44126.2066	0.6674	0.5834	0.0000
+ V_5	1	50467.8568	50509.4623	0.6192	0.5252	1.691e-132
+ V_1	1	51058.8940	51100.4996	0.6148	0.5499	0.0028
<none>		51065.8364	51097.0406	0.6147	0.5546	

Step 3 : + V\_6

	Df	AIC	BIC	Deviance+	Pearson <sup>^</sup>	p-value*
+ V_4	1	38571.9938	38624.0007	0.7090	0.5917	0.0000
+ V_5	1	43299.8989	43351.9058	0.6733	0.5947	3.187e-117
+ V_1	1	43820.7929	43872.7998	0.6694	0.5799	0.0029
<none>		43827.6578	43869.2633	0.6693	0.5756	

Step 4 : + V\_4

	Df	AIC	BIC	Deviance+	Pearson <sup>^</sup>	p-value*
+ V_1	1	37972.8746	38035.2828	0.7135	0.6171	9.558e-133
+ V_5	1	38133.1434	38195.5516	0.7123	0.6234	7.069e-98
<none>		38571.9938	38624.0007	0.7090	0.5917	

Step 5 : + V\_1

	Df	AIC	BIC	Deviance+	Pearson <sup>^</sup>	p-value*
+ V_5	1	37463.3449	37536.1546	0.7174	0.6591	2.947e-113
<none>		37972.8746	38035.2828	0.7135	0.6171	

Step 6 : + V\_5

+ Adjusted R-squared based on the residual deviance  
<sup>^</sup> Adjusted R-squared based on the Pearson statistic  
\* p-value of the likelihood-ratio test

Final model:

Fraude ~ V\_1 + V\_2 + V\_3 + V\_4 + V\_5 + V\_6

El resumen del modelo seleccionado está dado por:

Call:

```
glm(formula = Fraude ~ Fraude ~ V_1 + V_2 + V_3 + V_4 + V_5 + V_6,
     family = binomial(link = "probit"), data = data_train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.4053	-0.0802	-0.0364	-0.0178	3.4099

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-4.640e+00	3.580e-02	-129.59	<2e-16	***
V_2	1.506e-03	1.525e-05	98.70	<2e-16	***
V_3	-4.205e-02	4.943e-04	-85.07	<2e-16	***
V_4	2.482e+00	3.817e-02	65.04	<2e-16	***
V_5	-5.451e-01	2.415e-02	-22.57	<2e-16	***
V_6	6.807e+00	1.514e+01	0.45	0.653	
V_1	-2.609e-01	1.071e-02	-24.37	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 132523 on 243136 degrees of freedom  
 Residual deviance: 37449 on 243130 degrees of freedom  
 AIC: 37463

Number of Fisher Scoring iterations: 17

Ahora, bajo la hipótesis nula  $H_0 : \beta_6 = 0$  (*irrupción*) y el tests de Wald, dado que el valor p asociado a la prueba es menor al nivel de significancia (5%), existe evidencia estadística para afirmar que el evento de fraude no depende del patrón de irrupción de un usuario a otro que no lo acompaña. Por tanto, el modelo seleccionado es:

Call:

```
glm(formula = Fraude ~ V_2 + V_3 + V_4 + V_5 + V_1,
     family = binomial(link = "probit"), data = data_train)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-3.6088	-0.0851	-0.0364	-0.0169	3.3640

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-4.733e+00	3.353e-02	-141.15	<2e-16 ***
V_2	1.551e-03	1.473e-05	105.30	<2e-16 ***
V_3	-4.289e-02	4.790e-04	-89.55	<2e-16 ***
V_4	2.538e+00	3.579e-02	70.90	<2e-16 ***
V_5	-5.728e-01	2.312e-02	-24.77	<2e-16 ***
V_1	-2.615e-01	1.020e-02	-25.64	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 132523 on 243136 degrees of freedom  
 Residual deviance: 42767 on 243131 degrees of freedom  
 AIC: 42779

Number of Fisher Scoring iterations: 9

Sin embargo, los parámetros asociados al tiempo máximo en transacción y la variable indicadora del patrón de retorno son negativos, hechos que sugieren que a mayor tiempo de transacción o que dado el patrón de regreso sobre los cajeros se espera una menor probabilidad de fraude; lo cual, de acuerdo al contexto del problema es contradictorio. Apoyados en que la correlación del tiempo máximo de transacción y la distancia máxima recorrida es mayor al 95 %, ver tabla 4-4, se procede a retirar del modelo la covariable relacionada con el tiempo:

Call:

```
glm(formula = Fraude ~ V_2 + V_4 + V_5 + V_1,
     family = binomial(link = "probit"), data = data_train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9803	-0.1269	-0.0520	-0.0214	3.5908

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.664e+00	2.744e-02	-133.51	<2e-16 ***
V_2	2.664e-04	2.421e-06	110.03	<2e-16 ***
V_4	2.806e+00	3.570e-02	78.59	<2e-16 ***
V_5	1.124e+00	1.321e-02	85.07	<2e-16 ***
V_2	-1.728e-01	9.118e-03	-18.95	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 132523 on 243136 degrees of freedom  
 Residual deviance: 51906 on 243132 degrees of freedom  
 AIC: 51916

Number of Fisher Scoring iterations: 8

Complementando, se evalúa si el efecto de la distancia máxima recorrida depende del paso por uno o más cajeros, el patrón de regreso o el número de personas en el lugar; obteniéndose el modelo expuesto en 4.4.

Acompañándose de la evaluación de la inclusión de cada covariable realizando 100 muestras aleatorias simples y estratificadas para diferentes tamaños muestrales. Como se observa en la figura **A-1** tomando el camino de selección de la metodología *forward*, se observa que la media y la mediana de los valor p, obtenidos al evaluar con el test de Wald la hipótesis nula  $\beta_{(\cdot)} = 0$ , tienden a cero a medida que el tamaño de la muestra aumenta. Conclusión que se extiende a la hipótesis de que el efecto de la distancia máxima recorrida sobre la realización o no de fraude, depende de si hay o no patrón de retorno ( $\Phi(\mu_k) = \beta_0 + \beta_1 * V_2 + \beta_2 * V_5 + \beta_3 * V_4 + \beta_4 * V_1 + \beta_5 * V_2 * V_5$ ).

Finalmente, se puede ver que otras estructuras como  $\Phi(\mu_k) = \beta_0 + \beta_1 * V_2 + \beta_2 * V_5 + \beta_3 * V_4 + \beta_4 * V_1 + \beta_6 * V_2 * V_4$ , reflejan valores p que tienden a mostrar que, el evento de fraude no

depende de alguna otra covariable propuesta o, que el efecto de las variables seleccionadas depende de otra covariable.

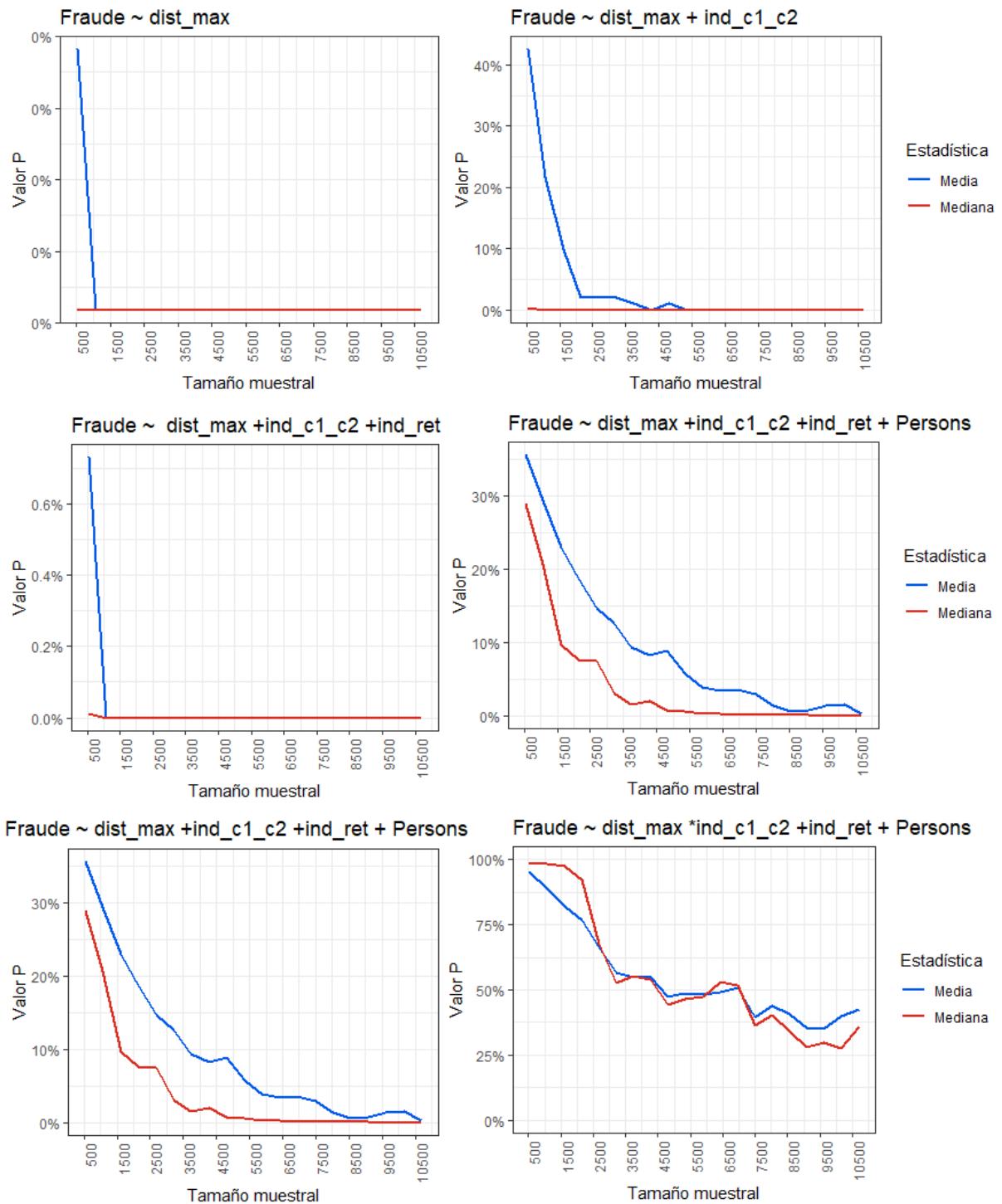


Figura A-1.: Valor p asociado al estadístico de Wald dado el tamaño muestral.

## B. Análisis de sensibilidad.

Los residuos estandarizados del modelo, muestran una distribución aproximadamente simétrica, que de acuerdo al test de Shapiro-Wilk el estadístico  $W = 0,99963$  tiene asociado un valor  $p$  igual a 0.4879, luego no existe evidencia estadística para afirmar que los residuos estandarizados no sean aproximadamente normales. Además, los mismos residuos comparados con las estimaciones de la variable respuesta, se encuentran distribuidos aleatoriamente lo que sugiere un modelo adecuado para los datos, ver figura B-1.

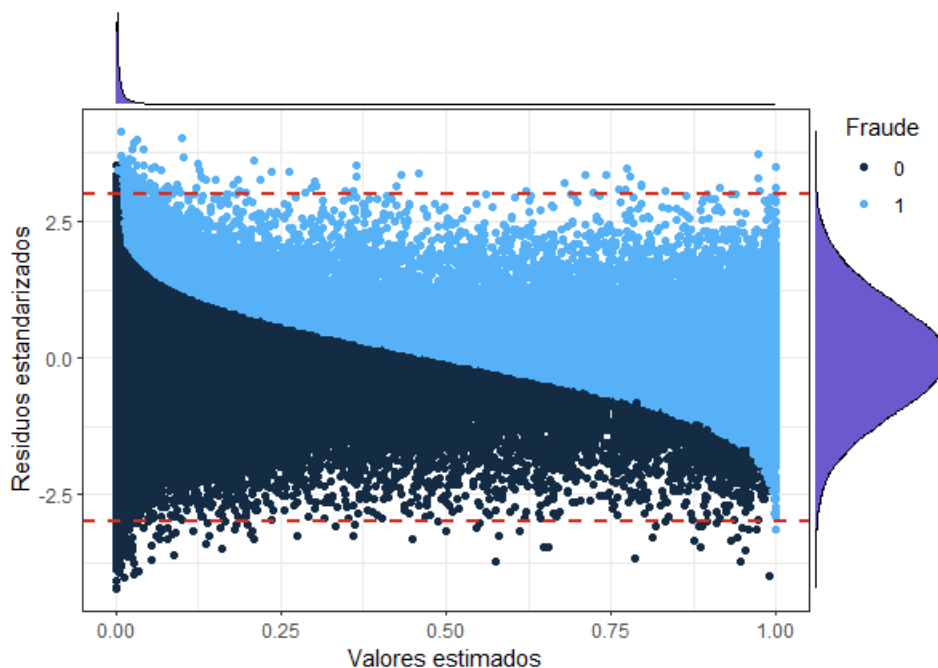


Figura B-1.: Residuos estandarizados

Ahora, es de interés validar si existen observaciones atípicas que influyan la estimación de los parámetros del modelo, para lo que mediante el la distancia de Cook se identifican 360 fotogramas con el estadístico mayor o igual a 0.02 (percentil 99). Al retirarlas de la estimación del modelo, se encuentra que los signos de sus parámetros permanecen iguales con bajas variaciones en su estimación; por tanto, las conclusiones obtenidas se mantienen y no existe evidencia de observaciones influyentes:

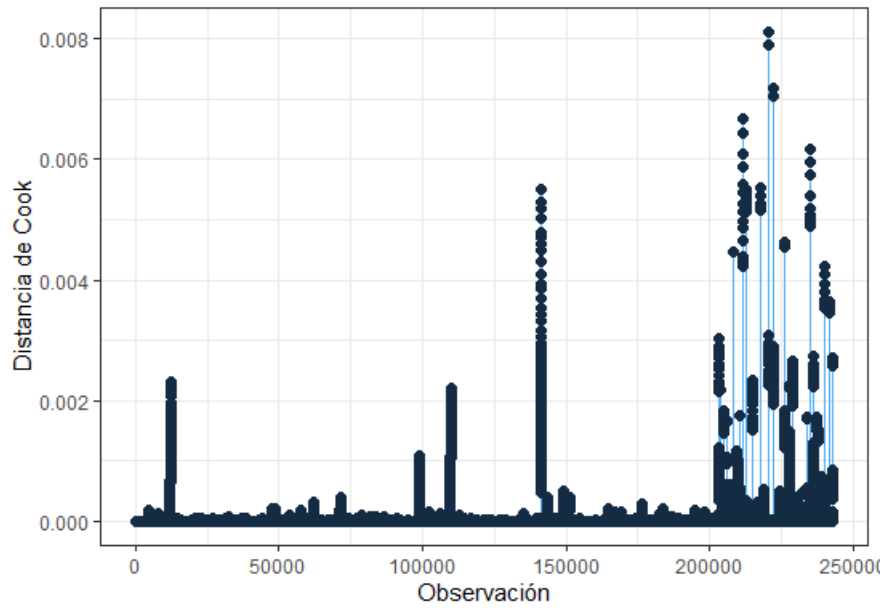


Figura B-2.: Distancia de Cook

	Estimate	Std. Error	Pr(> z )		Estimate*	Std. Error*	Pr(> z )*		Change(%)
(Intercept)	-3.231e+00	2.791e-02	0		-3.288e+00	2.879e-02	0		-1.763
V_2	2.110e-04	2.828e-06	0		2.154e-04	2.885e-06	0		2.113
V_5	-2.183e-01	4.621e-02	0		-2.679e-01	4.811e-02	0		-22.716
V_4	2.896e+00	3.516e-02	0		3.705e+00	5.520e-02	0		27.908
V_1	-1.743e-01	9.021e-03	0		-1.661e-01	9.199e-03	0		4.714
V_2:V_5	1.488e-04	4.936e-06	0		1.565e-04	5.118e-06	0		5.150

Finalmente, mediante el qqplot y sus bandas simuladas al 95 % de confianza, de acuerdo a los procedimientos expuestos por Vanegas & Rondón (2018), se observa que la distribución de los errores del modelo están de acorde a los supuestos del mismo, por tanto, se concluye que el modelo propuesto es adecuado para explicar el evento de fraude, ver figura B-3.

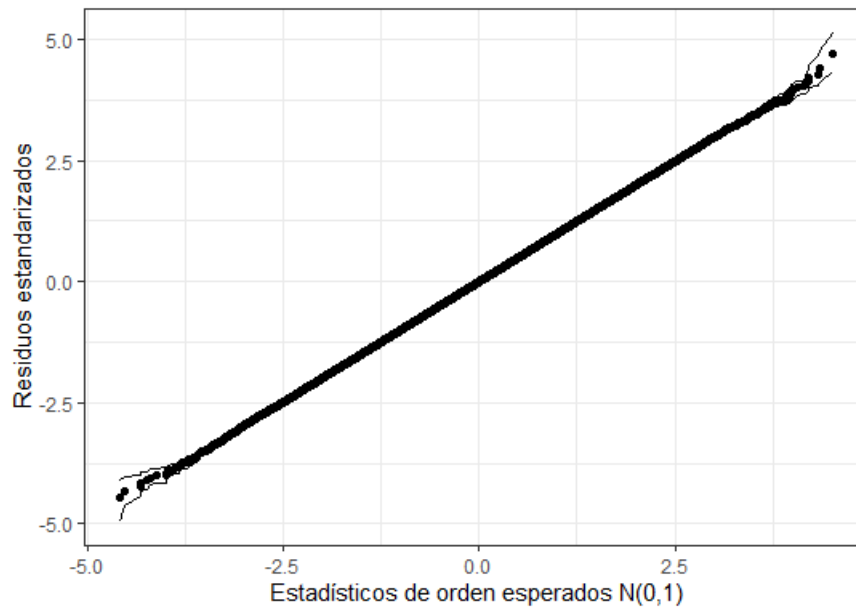


Figura B-3.: QQ plot: residuos estandarizados

## C. Redes neuronales y estructuración de recorridos.

Para el despliegue de las redes neuronales y la extracción de los recorridos de personas en vídeo se debe obtener la red pre-entrenada Faster R-CNN de OpenVIVO (2020) y clonar la red neuronal SiamMask del repositorio de Wang et al. (2019), realizando la combinación de los 2 modelos de la forma expuesta en <https://github.com/geagarciaar/TrabajoFinalGrado>.

# Referencias

- Agresti, A. (2015), *Foundations of Linear and Generalized Linear Models*, i edn, Wiley, United States of America.
- Arbib, M. (2003), *The Handbook of Brain Theory and Neural Networks*, i edn, Advisory Board, United States of America.
- Bengio, Y. (2009), 'Learning deep architectures for ai', *Foundations and Trends in Machine Learning* **2**, 1–127.
- Cyganek, B. (2013), *Video tracking: theory and practice*, i edn, John Wiley and Sons, United Kingdom.
- Fiaz, M., Mahmood, A. & Ki Jung, S. (2019), Deep siamese networks toward robust visual tracking, in L. Mazzeo, ed., 'Visual Object Tracking with Deep Neural Networks', IntechOpen, chapter 1, pp. 1–21.
- Goodfellow, I., Bengio, Y. & Courville, A. (2016), *Deep learning*, i edn, MIT Press.
- Khan, S., Rahmani, H., Ali Shah, S. A. & Bennamoun, M. (2018), 'A guide to convolutional neural networks for computer vision', *Synthesis Lectures on Computer Vision* **8**, 1–207.
- Lee, J. & Verleysen, M. (2007), *Nonlinear Dimension Reduction*, i edn, Springer, United States of America.
- Maggio, E. & Cavallaro, A. (2011), *Video tracking: theory and practice*, i edn, John Wiley and Sons, India.
- Mishachev, N. (2017), 'Backpropagation in matrix notation', *arXiv* **8**, 1–7.
- OpenVIVO (2020), 'Faster RCNN inception v2 COCO', [https://github.com/openvinotoolkit/open\\_model\\_zoo/tree/master/models/public/faster\\_rcnn\\_inception\\_v2\\_coco](https://github.com/openvinotoolkit/open_model_zoo/tree/master/models/public/faster_rcnn_inception_v2_coco). Online; accedido 03 de septiembre de 2020.
- Ren, S., He, K., Girshick, R. & Sun, J. (2017), 'Faster r-CNN: Towards real-time object detection with region proposal networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(6), 1137–1149.

- 
- Salman, K., Hossein, R., Syed, A., Ali, S. & Mohammed, B. (2018), *A Guide to Convolutional Neural Networks for Computer Vision*, i edn, Morgan Claypool.
- Sharma, Avinash (2020), ‘Understanding Activation Functions in Neural Networks’, <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>. Online; accedido 01 de agosto de 2020.
- Vanegas, H. & Rondón, L. (2018), ‘Notas de clase: Modelos lineales generalizados’.
- Wang, Q., Zhang, L. & Bertinetto, L. (2019), ‘Fast online object tracking and segmentation: A unifying approach’, *IEEE Conference On Computer Vision And Pattern Recognition* pp. 1328–1338.
- Warren, S. (1994), ‘Neural networks and statistical models’, *Proceedings of the Nineteenth Annual SAS Users Group International Conference* pp. 1–13.