



PROYECTO FINAL DE
INGENIERIA INDUSTRIAL

**Utilización de Redes Neuronales en la Optimización de
Campañas Publicitarias Digitales**

Autores:

Matías Shulz
Manuel María Zuretti

Tutor o Docente Guía:

2015

Contenido

1. Resumen Ejecutivo	1
2. Online Advertising Networks	3
3. Redes Neuronales	7
3.1 Topología de Redes	9
3.1.1 Clasificación según Datos de Entrada y Aprendizaje	9
3.1.2 Topología según su Conexión	10
3.2 Procesamiento de la Información en las Neuronas	11
3.3 Aprendizaje y Validación	12
3.3.1 El Aprendizaje Supervisado.....	13
• Regla del Perceptrón.....	14
• Regla Delta	15
• Regla Delta generalizada	17
3.3.2 Aprendizaje No Supervisado	19
3.4 Funciones de Activación y Salida.....	20
3.4.1 Función Escalón (Hardlim)	21
3.4.2 Función de Transferencia lineal (Purelin)	22
3.4.3 Función de Transferencia sigmoideal (logsig)	22
3.4.4 Principales funciones de transferencia	23
3.5 Usos y Aplicaciones	24
4. Datos Recopilados.....	26
4.1 Data Ad Network	26
4.2 Data Campaña	27
4.3 Data Censo	29
4.4 Visualización de las Variables de Entrada	30

4.4.1	Población Hispana	30
4.4.2	Población Hispana Online	31
4.4.3	Origen de la Población Hispana	33
4.4.4	Edad y Sexo	35
4.4.5	Household Income	39
4.4.6	Generaciones Hispanas	40
5.	Complementos de las RNA.....	43
5.1	Complementos de Entrada.....	43
5.2	Complementos de Salida	44
6.	Armado de la Red Neuronal Artificial y sus Complementos.....	47
6.1	Complementos de Entrada.....	47
6.1.1	Realización de la Normalización y el PCA.....	49
6.2	Red Neuronal Artificial.....	51
6.3	Complementos de Salida	51
6.4	Armado de la RNA	52
6.5	Armado de los Complementos de Salida	54
7.	Optimización de Red Neuronal Artificial.....	55
7.1	Ciclos de entrenamiento.....	55
	1000 Ciclos de Entrenamiento	55
	750 Ciclos de Entrenamiento	57
	1500 Ciclos de Entrenamiento	58
	500 Ciclos de Entrenamiento	58
	100 Ciclos de Entrenamiento	59
7.2	Cantidad de Capas Ocultas y de Neuronas Por Capa	59
	12 Neuronas	60
	16 Neuronas	60
	20 Neuronas	60

16 & 12 Neuronas por capa oculta	61
20 & 12 Neuronas por capa oculta	62
24 & 20 Neuronas por capa oculta	63
7.3 Constante de Aprendizaje	64
7.3.1 Constante de aprendizaje 0,3	64
7.3.2 Constante de aprendizaje 0,2	65
7.3.3 Constante de aprendizaje 0,5	66
7.4 Exigencia al CTR	66
7.4.1 Clasificación Bueno-Muy Bueno:	67
7.4.2 Clasificación Malo-Bueno	68
7.5 Configuración Final de la RNA	71
8. Conclusión	74

1. Resumen Ejecutivo

En el siguiente Proyecto Final de Ingeniería Industrial se estudiará el poder de las Redes Neuronales Artificiales (RNA) frente a la predicción de hechos o variables. La investigación realizada intentará anticipar el rendimiento de campañas publicitarias digitales, teniendo en cuenta sus características y los aspectos de las regiones geográficas de los mercados objetivos.

La problemática que se encuentra en la distribución de publicidad digital es la falta de información que se tiene sobre el usuario que la recibe. Generalmente, la entrega de anuncios online se hace en función a las páginas web que un individuo ha visitado anteriormente, asumiendo que el contenido de estas identifica sus intereses. El inconveniente reside en que, en la mayoría de los casos, las distribuidoras de publicidad (Ad Networks) sólo cuentan con el historial de sitios web del 20% de la audiencia diaria. El poder inferir sus necesidades a partir de datos obtenidos del censo y de información recolectada en campañas anteriores, puede mejorar sustancialmente la capacidad de llegar a los segmentos buscados.

La elección de la temática de investigación y el campo de aplicación se hicieron de manera conjunta. Las RNA son de perfecta aplicación en la publicidad online por tres atributos que las caracterizan: la capacidad de inferir correlaciones poco claras entre variables, el poder de trabajar con información incompleta y la posibilidad de ser utilizadas en tiempo real. En la mayoría de los casos, la información que se registra de un usuario no es más que su

ubicación, la hora de contacto e información acerca de la configuración de su ordenador. Es por esta razón que se adhieren datos del censo, logrando incrementar el número de variables manipuladas en gran medida. Sin embargo, esto hace que se torne dificultoso encontrar la relevancia de cada dato a la hora de clasificar a un cliente. Las redes neuronales pueden ser entrenadas para buscar los patrones de comportamiento y así, eliminar los datos que no aporten información. También pueden generalizar ciertas conductas, logrando independizarse de la necesidad de contar con todas las variables para tomar una decisión. Esto indica que pueden trabajar perfectamente ante situaciones nuevas, adaptándose a cualquier imprevisto. Asimismo, como las neuronas que componen al sistema funcionan en paralelo, el tiempo de respuesta de las redes no supera la fracción de segundo.

El objetivo principal del proyecto es buscar una aplicación nueva para una tecnología que data de los años 80. Las oportunidades que se presentan en el mundo digital son incontables, pero no se encuentran al alcance de todos. El progreso del campo elegido no sólo se basa en innovaciones tecnológicas, sino también en la capacidad de adaptar herramientas existentes a las necesidades actuales.

2. Online Advertising Networks

Una Online Advertising Network, o Ad Network, tiene como objetivo primario conectar a los anunciantes, con los espacios de publicidad digital. La diferencia fundamental con empresas que publicitan en otros medios, ya sea en televisión o diarios, es que la publicidad digital permite la segmentación y el seguimiento de usuarios y también posibilita un informe del rendimiento de la campaña publicitaria.

Existen varios tipos de Ad Networks y se clasifican según como distribuyan la oferta publicitaria a través de los diferentes espacios de difusión:

- **Blind Networks:** Las redes ciegas no ofrecen una lista de sitios a sus anunciantes. Por el contrario, les ofrecen bajos precios a cambio de que ellos cedan el control sobre el destino de sus publicidades. Generalmente, el tipo de anunciantes que opta por estas redes, no hace énfasis en la importancia de la asociación de marca y simplemente van detrás de un resultado.
- **Targeted Networks:** A veces llamadas redes de segunda generación, tienen un foco específico en “Targeting Technologies” y pueden ser contextuales o por comportamiento. Las redes contextuales buscan contextualizar los anuncios con el contenido que ofrece el sitio web. Se basan en que el contenido que el usuario está buscando identifica sus necesidades. Por ejemplo, si una persona se encuentra en un sitio de deportes, la publicidad estará en la misma categoría. Las redes que segmentan dependiendo del comportamiento de sus usuarios son llamadas “Behavioral Networks” y usan tecnologías que posibilitan analizar para cada persona, sus decisiones y acciones en tiempo real. Luego, infieren patrones de comportamiento y sirven publicidades relacionadas a dichas acciones.
- **Vertical Networks:** Proveen completa transparencia sobre el destino de los anuncios. El objetivo es conectar a los anunciantes con

usuarios interesados en un tema específico. Generalmente cuentan con un inventario Premium que luego comercializan a precios altos.

- Mobile Networks: Sumado a las Online Advertising Networks, existen redes que focalizan sus esfuerzos en la distribución de publicidad en dispositivos móviles.

En la figura 2.1 se muestra de manera simplificada a los actores principales en la cadena de valor de la publicidad online.



Figura 2.1 Cadena de Valor de la Publicidad Online

Para lograr entender la función de cada uno de los actores que forman la cadena de valor, primero se debe entender que es lo que simbolizan las flechas. Estas hacen referencia al producto intercambiado entre cada actor y son llamadas inventario publicitario. Este inventario se define como los espacios publicitarios disponibles para publicar anuncios. Cada vez que un individuo ve un anuncio digital, se dice que se realizó una impresión.

En la cadena de valor mostrada en la figura 2.1, los Publishers son los creadores del inventario. Ellos insertan espacios publicitarios en los sitios web, aplicaciones, e-mails o videojuegos que crean, estableciendo un flujo constante de inventario, que tiene como destino a los Anunciantes.

Las Ad Networks y las Agencias tienen como función principal conectar a los Publishers con los Anunciantes. Las primeras consiguen espacios publicitarios de varios Publishers y luego de agregarles valor, las venden a las Agencias o a los

Anunciantes. Las Agencias cumplen dos funciones esenciales: diseñan los anuncios (ya sea un banner animado o un video publicitario de 30 segundos) y compran el espacio publicitario. Generalmente, los Anunciantes destinan un presupuesto determinado para campañas online; el mismo es recibido por las Agencias, quienes se encargan crear el marketing mix necesario para llegar al segmento deseado.

Una vez que un anuncio alcanza a un individuo, el mismo se cataloga como un **Viewer**. Esto indica que la impresión llegó a destino. Si el usuario hace click sobre el ad, se convierte en un **Clicker**. Y finalmente, si el usuario alcanza una meta deseada, como por ejemplo completar un formulario o realizar una compra online, este pasa a ser Converter. En la figura 2.2 se puede ver una proporción estimada de la cantidad de Viewers que se convierten en Clickers y de la cantidad de Clickers que llegan Converters.

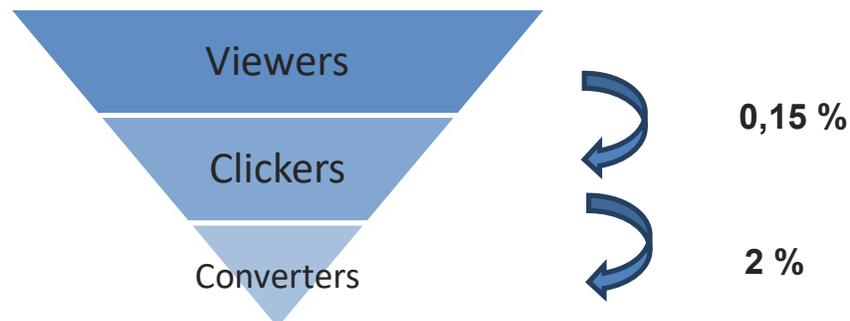


Figura 2.2 Clasificación de usuarios según su nivel de respuesta frente a los anuncios

Los valores mostrados en la tabla 2.2 son meramente un estimativo de lo que suelen ser los ratios de conversión de Viewers a Clickers y de Clickers a Converters. Estos valores pueden incrementar o decrementar si por ejemplo, la Ad Network que distribuye los anuncios lo hace de manera “Blind” o “Targeted”. Estos ratios, particularmente el de Viewers a Clickers, determinan el rendimiento de la campaña. En el negocio de la publicidad online, el índice más representativo

del rendimiento de las campañas relaciona la cantidad de Impresiones o Views, con la cantidad de Clicks. Este es llamado Click-Through Rate (CTR), y se calcula mediante la siguiente fórmula:

$$CTR = \frac{Clicks}{Impressions} \times 100$$

Figura 2.3 Fórmula del índice Click-Through Rate

Uno de los problemas principales del CTR es que no tiene en cuenta que se realiza luego de clickear en el ad. Es indiferente si el usuario, luego de hacer el click, accede a ver un video de 30 segundos entero o si lo cierra luego de los primeros 5. Otro conflicto que el índice no reconoce es que un mismo usuario puede hacer más de un click en una misma campaña publicitaria. Cuando esto ocurre, el anunciante cree que la audiencia interesada en su producto es mayor que la real.

Dejando de lado estos inconvenientes, el índice en cuestión puede revelar si la campaña se ha enviado a los segmentos deseados. Un valor menor a 0,05% indicaría que la calidad del segmento elegido es muy baja o directamente que la distribución fue realizada ciegamente. Los valores normales de CTR rondan valores del 0,13% y con las redes de segunda generación, se pueden incrementar notablemente.

3. Redes Neuronales

Las Redes Neuronales Artificiales (RNA), inspiradas en el funcionamiento del sistema nervioso biológico, son un paradigma de aprendizaje de procesamiento automático. Se trata de la interconexión de neuronas o perceptrones que logran inferir un resultado o una variable de salida, a partir de diferentes variables de entrada.

Los primeros modelos de RNA se retrotraen a los años 40, aunque su aplicación no fue hasta mediados de los 80, donde el científico Americano John Joseph Hopfield desarrolló la Red de Hopfield y Rumelhart ideó el algoritmo de “Backpropagation” (propagación hacia atrás).

Las ventajas que tienen las RNA son numerosas y varias de ellas tienen que ver con la semejanza entre redes y cerebros. Entre las más relevantes para el proyecto se encuentran:

- Aprendizaje: Las redes son capaces de realizar tareas basadas en un entrenamiento o experiencia previa. Estas pueden identificar patrones, eliminando la necesidad de especificar distribuciones de probabilidad en las variables de entrada. Además, las neuronas son elementos dinámicos y pueden adaptarse a nuevas condiciones.
- Organización Automática: Además de compilar información en cada neurona en particular, la red puede modificarse enteramente para un fin específico. Esto quiere decir que las RNA se auto-organizan para poder inferir una mayor cantidad de resultados. Lo que produce en la organización es la generalización. Esto es la capacidad de responder apropiadamente a situaciones nunca antes vistas, una facultad muy importante si no se cuentan con datos claros, precisos y completos.
- Operación en Tiempo Real: Debido a que las redes trabajan en paralelo, tienen la capacidad de trabajar a altas velocidades. Esta facultad la hace esencialmente apta para ser usada en procesos

relacionados con e-Marketing, como por ejemplo, tomar decisiones en milésimas de segundo.

Para entender el paralelismo entre las neuronas biológicas y las artificiales, llamadas perceptrones, se debe entender el funcionamiento de las primeras. La información proveniente de otras neuronas llega a través de las Dendritas. Cada una de las neuronas que pasa información de entrada, tiene un peso o una relevancia diferente en el resultado final. La información pesada llega al Núcleo y luego de ser analizada, sale por el Axón hacia otra neurona. El cerebro humano se compone de billones de neuronas conectadas entre sí.

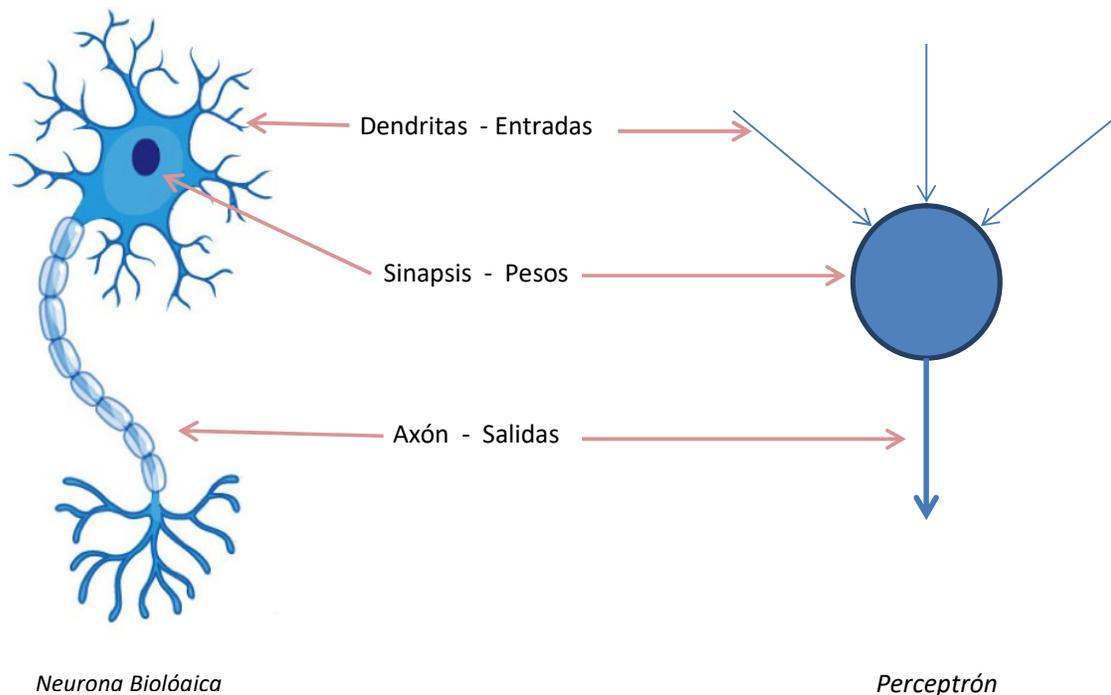


Figura 3.1 Comparación entre la Neurona Biológica y la Neurona Artificial o Perceptrón

En la figura 3.2 se muestra un esquema básico de una RNA. Las variables de entrada, o Inputs, ubicados a la izquierda, son los valores de los datos que ingresan a la red por medio de la Capa de Entrada. La información pasa por las

Capas Ocultas, donde se computa y se destina a la Capa de Salida. Dentro de una RNA puede haber tantas Capas Ocultas como se crean necesarias.

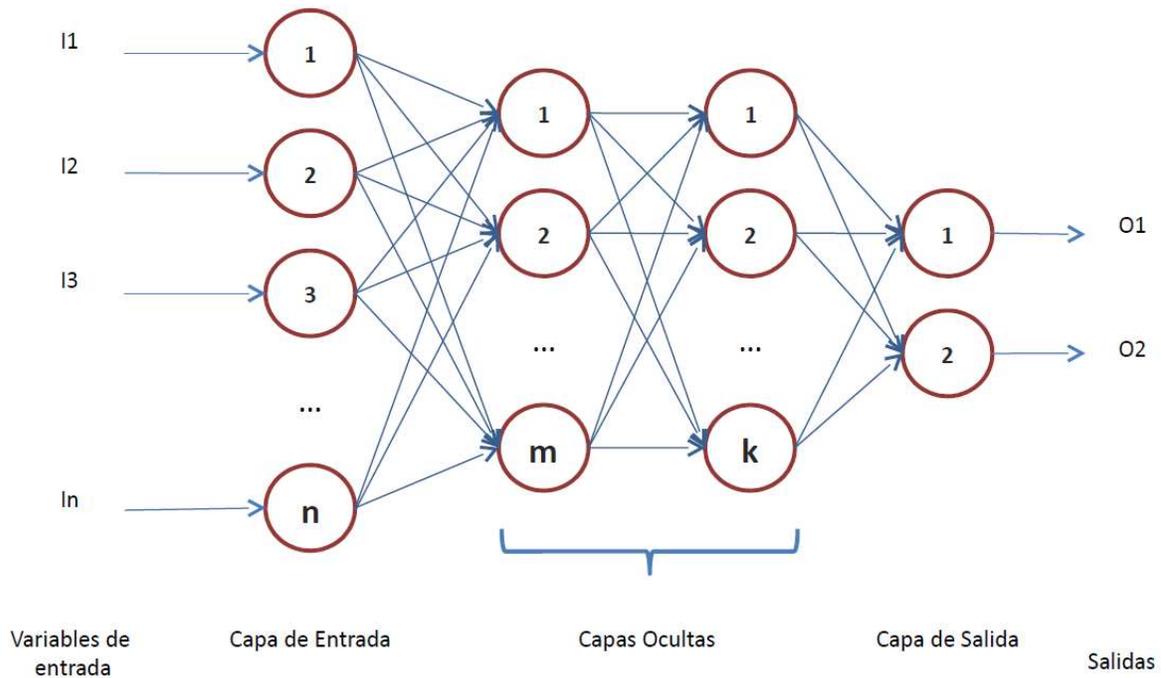


Figura 3.2 Esquema básico de una Red Neuronal Artificial

3.1 Topología de Redes

Las RNA se pueden clasificar de varias maneras, dependiendo de las características y capacidades que posea. Las tres clasificaciones más utilizadas en el área de estudio son según: las conexiones entre neuronas, el tipo de aprendizaje de la red y la configuración de los datos de entrada.

3.1.1 Clasificación según Datos de Entrada y Aprendizaje

La primera clasificación de las RNA se suele hacer en función a la configuración de los datos de entrada y al tipo de aprendizaje que estas pueden

realizar. En la Figura 3.1.1 se esquematizan de manera clara los nombres de las distintas redes, dependiendo de estos dos factores.

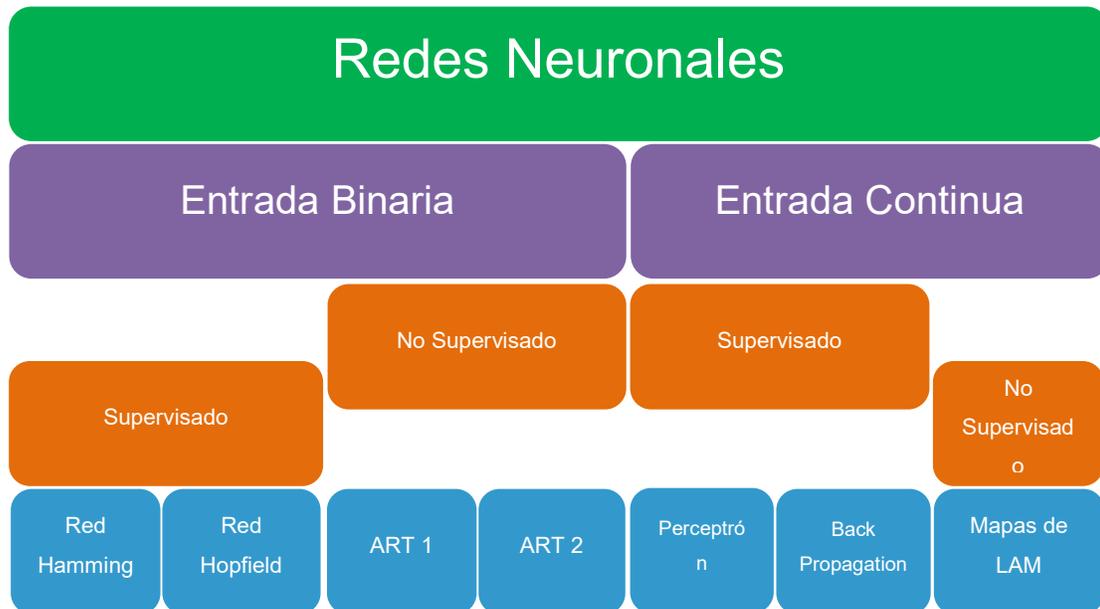
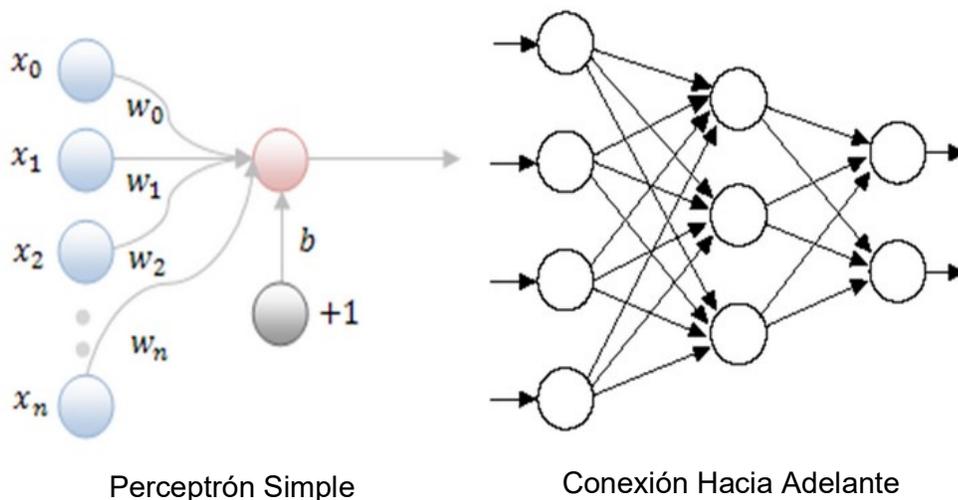


Figura 3.1.1 Clasificación de redes según datos de entrada y aprendizaje

3.1.2 Topología según su Conexión

El segundo tipo de clasificación de redes se suele hacer dependiendo del tipo de conexiones entre neuronas. Se puede diferenciar dos tipos básicos de redes neuronales: con conexión hacia adelante y con conexiones recurrentes. En los siguientes esquemas se pueden identificar y entender las diferentes topologías de conexiones entre neuronas.



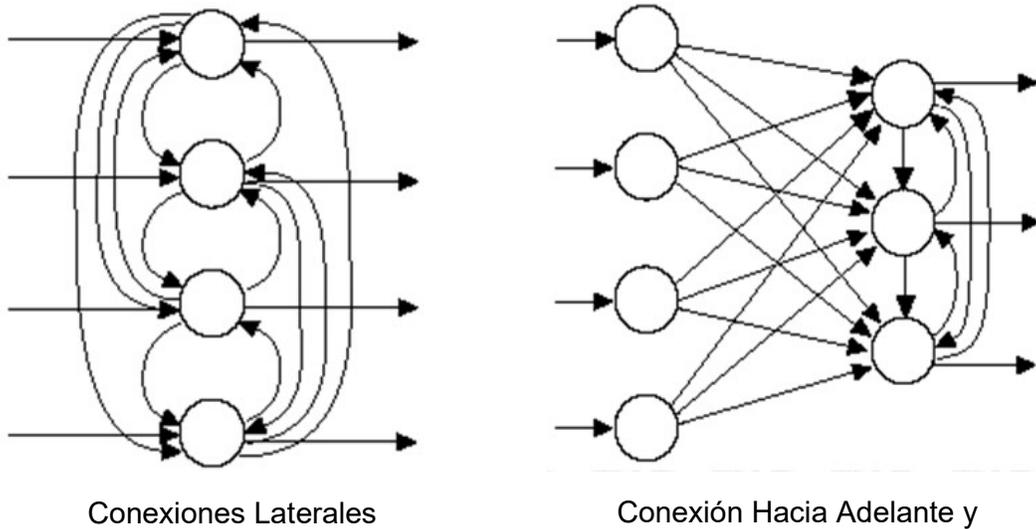


Figura 3.1.2 Topología de redes según el tipo de conexiones entre neuronas

3.2 Procesamiento de la Información en las Neuronas

Toda la información proveniente de las neuronas de entrada, se computa y se convierte en una entrada global (gini). La combinación de las diferentes entradas se puede realizar de varias maneras, aunque las más utilizadas son las siguientes:

Sumatoria de las entradas por sus pesos,
$$gin = \sum_j n * w$$

Productoria de las entradas por sus pesos,
$$gin = \prod_j n * w$$

Máximo de las entradas por sus pesos,
$$gin = \max_j n * w$$

Siendo n_j una neurona entrante determinada y w_j el peso que se le asigna. Cada neurona o dato entrante tendrá un valor propio.

Luego de ser transformada en una entrada global, la información pasa por una *Función de Activación*. La función calcula el estado de la neurona, llevando la entrada global menos un umbral Θ , a un valor determinado. Las funciones más conocidas son la lineal, la sigmoidea y la tangente hiperbólica. Estas tres funciones serán tratadas en capítulos posteriores.

La Función de Activación devuelve un valor a la Función de salida, quien hace una última modificación y envía un dato a las demás neuronas. Generalmente suelen adoptarse una de las siguientes funciones:

- Binaria: Dependiendo del valor devuelto por la Función de Activación, la Función de salida devuelve un 0 o un 1.
- Ninguna: La función devuelve el mismo valor que entró.

3.3 Aprendizaje y Validación

Como la RNA debe ser entrenada con un conjunto de ejemplos ya existentes, es necesario proveer a la red de un conjunto de datos pasados en los cuales se pueda basar. El proceso de aprendizaje es denominado *proceso de entrenamiento o acondicionamiento*, mientras que el conjunto de datos que se procesan son llamados *conjunto de datos de entrenamiento*.

Si las diferentes funciones en la red no pueden cambiar, mientras que los pesos en la misma son dinámicos, el aprendizaje de la red se basa en la *adaptación de pesos*. En este caso, se encuentra una similitud con las neuronas biológicas, las cuales crean, destruyen y modifican conexiones con otras neuronas cercanas. La variación de los pesos entre un paso del entrenamiento y otro se describe como $\frac{dw}{dt}$ y se dice que la red ha finalizado su entrenamiento cuando los valores de los pesos permanecen estables ($\frac{dw}{dt} = 0$) o cuando

alcanzan un valor aceptable. La fórmula utilizada para calcular el cambio de los pesos es:

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

$$\text{Peso Nuevo} = \text{Peso Viejo} + \text{Cambio de Peso}$$

En las redes neuronales la esencia del aprendizaje se encuentra en la modificación colectiva de los pesos de los elementos de procesado. Una definición de aprendizaje en redes neuronales podría ser la siguiente: Proceso por el cual los parámetros libres de una red neuronal son ajustados a través de un proceso continuo de estimulación por parte del entorno en donde se sitúa el sistema. El tipo de aprendizaje viene determinado por la forma en la que tienen lugar dichos cambios. Esta definición, implica tres hechos fundamentales:

La red neuronal se encuentra estimulada por el entorno, cambia como consecuencia de esos estímulos y responde de manera diferente al entorno a causa de los cambios que se han producido en sus estructura interna.

Hay esencialmente dos métodos de aprendizaje:

- Aprendizaje Supervisado
- Aprendizaje No Supervisado

Otro método para diferenciar redes, aunque menos utilizado, es el de catalogarlas según si el aprendizaje se realiza durante su funcionamiento (online) o si es necesario desconectar la red para entrenarla (offline).

3.3.1 El Aprendizaje Supervisado

Se caracteriza por tener un agente externo o supervisor entrenando a la red. Este determina la respuesta que el sistema debe otorgar a partir de una entrada determinada. En caso que la salida no corresponda con la esperada, se modifican los pesos de las neuronas, con el fin de disminuir el error entre el valor respuesta y el esperado.

La adaptación sucede cuando el sistema compara directamente la salida que proporciona la red con la salida que se desearía obtener de dicha red. Existen tres tipos básicos: por corrección de error, por refuerzo y estocástico.

En el aprendizaje por corrección de error el entrenamiento consiste en presentar al sistema un conjunto de pares de datos, representando la entrada y la salida deseada para dicha entrada (este conjunto recibe el nombre de conjunto de entrenamiento). El objetivo, es minimizar el error entre la salida deseada y la salida que se obtiene.

El aprendizaje por refuerzo, es más lento que el anterior. No se dispone de un ejemplo completo del comportamiento deseado pues no se conoce la salida deseada exacta para cada entrada sino que se conoce el comportamiento de manera general para diferentes entradas. La relación entrada-salida, se realiza a través de un proceso de éxito o fracaso, produciendo este una señal de refuerzo que mide el buen funcionamiento del sistema. La función del supervisor es más la de un crítico que la de un maestro.

El aprendizaje estocástico consiste básicamente en realizar cambios aleatorios de los valores de los pesos y evaluar su efecto a partir del objetivo deseado.

- **Regla del Perceptrón**

Supongamos que tenemos un conjunto de muestras de aprendizaje consistentes en vectores de entrada x y salidas deseadas $d(x)$ (en tareas de clasificación $d(x)$ es, usualmente, $+1$ o -1). La regla de aprendizaje del perceptrón es muy simple y se desarrolla a través de los siguientes pasos.

1. Iniciar la red con un conjunto aleatorio de pesos para las conexiones.
2. Seleccionar un vector de entrada x del conjunto de muestras de entrenamiento.

3. Si el perceptrón proporciona una salida incorrecta, modificar todas las conexiones mediante la expresión:

$$\Delta w_i = d(x)x_i \quad (1.1)$$

4. Volver al paso 2.

Cuando la red responde de forma correcta los pesos de las conexiones no se modifican.

- **Regla Delta**

Para una red con una única capa con un único elemento de procesado en la salida con una función de activación lineal, la salida viene dada por:

$$y = \sum_j w_j x_j + b \quad (1.2)$$

Una red simple con una única capa, es capaz de representar relaciones lineales entre el valor del elemento de procesado de la salida y el valor de los elementos de procesado de la entrada. Umbralizando el valor de la salida se puede construir un clasificador. Pero, también se pueden realizar otras tareas como aproximación de funciones. En espacios de entrada multidimensionales la red representa un hiperplano de decisión y es trivial asumir que se pueden definir varias unidades de salida.

Supongamos que vamos a entrenar una red para localizar el hiperplano de decisión más idóneo para un conjunto de muestras de entrada consistente en valores de entrada y valores de salida deseada. Para cada muestra dada del conjunto de entrada, la salida real de la red difiere de la salida deseada en δ , donde y es la salida real para el patrón de entrada. La regla delta una función de error (o función de costo) basada en dicha diferencia para realizar el ajuste de los pesos.

La función de error, dada por el método del menor error cuadrático medio, es la suma de los cuadrados de los errores individuales de cada patrón. Es decir, el error total E viene dado por:

$$E = \sum_p E^p = \frac{1}{2} \sum_p (d^p - y^p)^2 \quad (1.3)$$

En donde el índice p recorre el conjunto de patrones de entrada y representa el error del patrón p-ésimo. Los valores de los pesos de conexión se buscan de forma tal que se minimice la función y este proceso se realiza por un método conocido gradiente descendiente. La idea del método, es realizar un cambio en los pesos inversamente proporcional a la derivada del error respecto al peso para cada patrón:

$$\Delta_p w_j = -\gamma \frac{\partial E^p}{\partial w_j} \quad (1.4)$$

En donde Y es una constante de proporcionalidad. La derivada se puede descomponer mediante la siguiente expresión:

$$\frac{\partial E^p}{\partial w_j} = \frac{\partial E^p}{\partial y^p} \frac{\partial y^p}{\partial w_j} \quad (1.5)$$

Dado que la función de activación es lineal, como se comentaba al comienzo,

$$\frac{\partial y^p}{\partial w_j} = x_j \quad (1.6)$$

Y:

$$\frac{\partial E^p}{\partial y^p} = -(d^p - y^p) \quad (1.7)$$

Con lo cual:

$$\Delta_p w_j = \gamma \delta^p x_j \quad (1.8)$$

En donde $\delta^p = d^p - y^p$, es la diferencia entre la salida deseada y la salida real para el patrón de entrada p.

La regla Delta, modifica los pesos como función de las salidas deseada y real para unidades de entrada y salida binarias o continuas. Estas características abrieron un nuevo campo de posibilidades para las redes neuronales.

- **Regla Delta generalizada**

La regla Delta generalizada, amplía la regla delta al caso de usar unidades de procesamiento con funciones de activación no lineal. La activación es una función diferenciable de la entrada total dada por la siguiente ecuación:

$$y_k^p = F(s_k^p) \quad (1.9)$$

En donde:

$$s_k^p = \sum_j w_{jk} y_j^p + \theta_k \quad (1.10)$$

Para realizar la generalización de la regla delta debemos considerar:

$$\Delta_p w_{jk} = -\gamma \frac{\partial E^p}{\partial w_{jk}} \quad (1.11)$$

La medida de error E_p viene definida como el error cuadrático total para el patrón p en las unidades de salida N_0 .

$$E_p = \frac{1}{2} \sum_{o=1}^{N_0} (d_o^p - y_o^p)^2 \quad (1.12)$$

Considerando E, como la suma de los errores cuadráticos podemos ver que la variación del error respecto a los pesos viene dada por:

$$\frac{\partial E_p}{\partial w_{jk}} = \frac{\partial E^p}{\partial s_k^p} \frac{\partial s_k^p}{\partial w_{jk}} \quad (1.13)$$

A partir de la expresión de la salida total y definiendo:

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} \quad (1.14)$$

Se puede obtener una regla de actualización de pesos equivalente a la regla Delta, resultando un descenso en el valor del gradiente sobre la superficie de error si se realizan los cambios de los pesos de acuerdo a la expresión:

$$\Delta_p w_{jk} = \gamma \delta_k^p y_j^p \quad (1.15)$$

La conclusión más importante, es que el cálculo recursivo de los factores Delta puede implementarse propagando las señales de error desde la capa de salida a capas ocultas de la red.

El cálculo de δ_k^p puede realizarse en términos de un producto de dos factores; uno que refleje el cambio del error como función de la salida de las unidades y otro que refleje el cambio de la salida como función de los cambios en la entrada; es decir:

$$\delta_k^p = -\frac{\partial E^p}{\partial s_k^p} = \frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial s_k^p} \quad (1.16)$$

Ahora bien, el segundo factor no es otro que la derivada parcial de la función de activación F. Para calcular el primer factor vamos a considerar dos casos: que la unidad k es una unidad de salida de la red y que la unidad k es una unidad de una capa oculta.

En el primer caso, a partir de la definición del error para el patrón p , podemos obtener que:

$$\frac{\partial E_p}{\partial y_o^p} = -(d_o^p - y_o^p) \quad (1.17)$$

Que es el mismo resultado que en la regla Delta básica. Por tanto, para las unidades de la capa de salida:

$$\delta_o^p = (d_o^p - y_o^p) F'_o(s_o^p) \quad (1.18)$$

En el segundo caso, no se conoce como contribuye la unidad al error de salida de la red, pero la medida de error se puede escribir como función de los pesos de las unidades de la capa oculta a la capa de salida:

$$\frac{\partial E^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial s_o^p}{\partial y_h^p} = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} \frac{\partial}{\partial y_h^p} \sum_{j=1}^{N_b} w_{ko} y_j^p = \sum_{o=1}^{N_o} \frac{\partial E^p}{\partial s_o^p} w_{ho} = - \sum_{o=1}^{N_o} \delta_o^p w_{ho} \quad (1.19)$$

Con lo cual:

$$\delta_h^p = F'_h(s_h^p) \sum_{o=1}^{N_o} \delta_o^p w_{ho} \quad (1.20)$$

Extendiendo este cálculo de forma recursiva se pueden calcular los cambios en los pesos de todas las unidades de la red. Este procedimiento, se denomina regla Delta generalizada para redes feedforward con unidades no lineales.

3.3.2 Aprendizaje No Supervisado

En las redes de Aprendizaje No Supervisado, no existe un agente externo que indique las respuestas del sistema. Es decir, la red no sabe cuál es la respuesta correcta. Estas son utilizadas para encontrar regularidades o

correlaciones, con la información que recopilan a lo largo del uso. Pueden indicar, entre otras cosas, grado de familiaridad de las variables de entrada o hasta clusterizarlas en diferentes categorías.

El proceso de aprendizaje finaliza cuando se alcanza la condición de detención. Esta generalmente indica un valor umbral que debe ser alcanzado por el cálculo del error cuadrado de todos los ejemplos de entrenamiento. Otra condición puede ser completar un cierto número de ciclos de entrenamiento. Luego de alcanzar la condición, los pesos quedan estáticos y la transformación de los datos de entrada a los de salida queda resuelta.

Uno de los métodos más utilizados en la modificación de los pesos es llamado Backpropagation. Como su nombre lo indica, el error se propaga hacia atrás a través de la red neuronal. Inicia su propagación en la capa de salida y atraviesa toda la red hasta llegar a la capa de entrada. Esto permite que los pesos entre las conexiones de las neuronas en las capas ocultas también puedan ser modificados.

A mediados de los 80, el científico Hopfield creó una estructura de redes neuronales capaces de almacenar patrones presentados en forma incompleta o con ruido. Esta consiste en neuronas íntegramente conectadas en forma recurrente. Esto significa que si existe una conexión entre las neuronas N_i y N_j , también se tendrá una entre N_j y N_i y sus pesos serán iguales ($w_{ij}=w_{ji}$). El conjunto de valores de entrada y salida son valores binarios ($\{0, 1\}$ o $\{-1, 1\}$), indicando sólo su estado de actividad o inactividad. La clave detrás de estas redes es que si se cuenta con un patrón conocido, los pesos de las conexiones neuronales pueden ser calculados. En estos casos sólo cambia el estado de las neuronas, manteniendo sus pesos fijos.

3.4 Funciones de Activación y Salida

Además de la regla de propagación es necesario poseer expresiones para las funciones de activación (calculan la activación en función de la entrada total) y funciones de salida (calculan la salida en función de la activación).

La función de activación calcula la activación de la unidad en función de la entrada total y la activación previa, aunque en la mayor parte de los casos es simplemente una función no decreciente de la entrada total. Los tipos de función más empleados son: la función escalón, función lineal y la función sigmoideal.

La función de salida empleada usualmente es la función identidad y así la salida de la unidad de procesamiento es idéntica a su nivel de activación.

Las redes neuronales están formadas por un conjunto de neuronas artificiales interconectadas. Las neuronas de la red se encuentran distribuidas en diferentes capas de neuronas, de manera que las neuronas de una capa están conectadas con las neuronas de la capa siguiente, a las que pueden enviar información.

3.4.1 Función Escalón (Hardlim)

En la figura de abajo se puede observar como esta función de transferencia acerca la salida de la red a cero, si el argumento de la función es menor que cero y la lleva a uno si este argumento es mayor que uno. Esta función crea neuronas que clasifican las entradas en dos categorías diferentes, característica que le permite ser empleada en la red tipo Perceptrón.

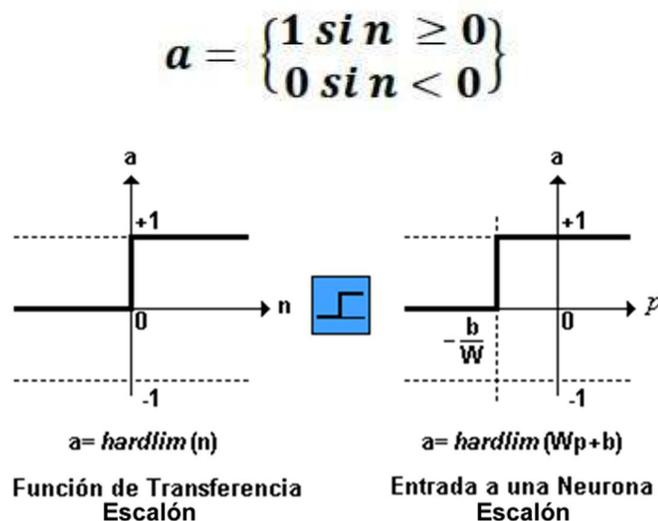


Figura 3.4.1 Función Escalón

3.4.2 Función de Transferencia lineal (Purelin)

La salida de una función de transferencia lineal es igual a su entrada, la cual se representa en la Figura 3.4.2.

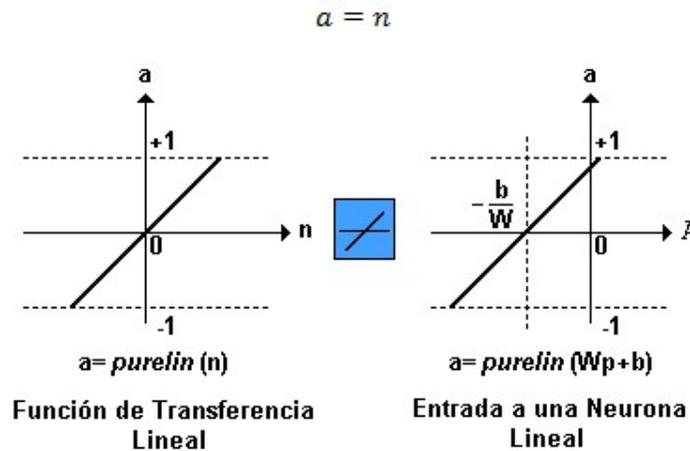


Figura 3.4.2 Función Lineal

3.4.3 Función de Transferencia sigmoideal (logsig)

Esta función toma los valores de entrada, los cuales pueden oscilar entre más y menos infinito, y restringe la salida a valores entre cero y uno, de acuerdo a la expresión:

$$a = \frac{1}{1 + e^{-n}}$$

Esta función es comúnmente usada en redes multicapa, como la Backpropagation, en parte porque la función logsig es diferenciable, como se muestra en la Figura 3.4.1.

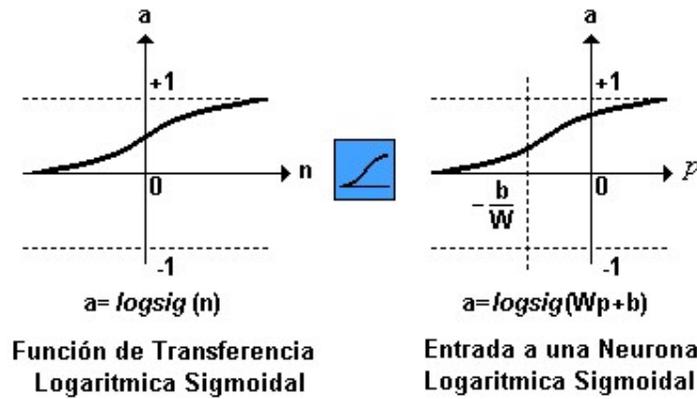


Figura 3.4.1 Función Sigmoidal

3.4.4 Principales funciones de transferencia

Nombre	Relación Entrada /Salida	icono	Función en Matlab
Limitador Fuerte	$a = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases}$		hardlim
Limitador Fuerte Simétrico	$a = \begin{cases} 1 & \text{si } n \geq 0 \\ -1 & \text{si } n < 0 \end{cases}$		hardlims
Lineal Positiva	$a = \begin{cases} 0 & \text{si } n < 0 \\ n & 0 \leq n \end{cases}$		poslin
Lineal	$a = n$		purelin
Lineal Saturado	$a = \begin{cases} 0 & \text{si } n < 0 \\ n & \text{si } 0 \leq n \leq 1 \\ 1 & n > 1 \end{cases}$		satlin
Lineal Saturado Simétrico	$a = \begin{cases} -1 & \text{si } n < -1 \\ n & \text{si } -1 \leq n \leq 1 \\ 1 & \text{si } n > 1 \end{cases}$		satlins
Sigmoidal Logarítmico	$a = \frac{1}{1 + e^{-n}}$		logsig
Tangente Sigmoidal Hiperbólica	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$		tansig
Competitiva	a=1 Neurona con n max a=0 El resto de las neuronas		Compet

Tabla 3.4.1 Principales Funciones de Transferencia

3.5 Usos y Aplicaciones

Cada uno de los modelos de redes trabaja de manera diferente y sus resultados pueden diferir en gran medida. Esto las hace aptas para ser utilizadas en diferentes campos. Algunas aplicaciones utilizadas en la actualidad son:

- **Análisis Financiero**
 - Pronóstico financiero
 - Valoración de riesgo de préstamos
 - Previsión de evolución de precios
 - Interpretación de firmas
- **Procesado de Imágenes**
 - Clasificación
 - Reconocimiento
 - Reconstrucción de imágenes
 - Inspección industrial
- **Medicina**
 - Diagnóstico médico por imágenes y lectura de sensores
 - Monitorización de cirugías
 - Entendimiento de causa de enfermedades
 - Obtención de modelos retinales
 - Predicción de reacciones adversas a medicamentos
- **Control y Robótica**
 - Balance
 - Comandos de brazos robóticos
- **Militares**
 - Clasificación de señales de radar
 - Creación de armas inteligentes
 - Optimización del uso de recursos escasos
 - Reconocimiento y seguimiento en el tiro al blanco

- Procesado de Voz
 - Reconocimiento de fonemas
 - Identificación de palabras escritas o habladas
 - Pronunciación de un texto (text to speech)
- Manufacturación
 - Control de producción de líneas de proceso
 - Inspección de calidad
- Modelo Ambiental
 - Previsión de tiempo
 - Análisis de patrones ambientales
- Empresas
 - Evaluación de factibilidad de hallar yacimientos petrolíferos
 - Identificación de candidatos para posiciones específicas
 - Modelado de sistemas para automatización y control

4. Datos Recopilados

La RNA que se utiliza en el proyecto es de Aprendizaje Supervisado. Esto quiere decir que al sistema se le otorga un set de variables de entrada y un valor de salida esperado para cada ciclo de la etapa de entrenamiento. Luego de un determinado número de ciclos, se da por finalizada esta etapa y se continúa a la de prueba.

La función principal de esta red neuronal es predecir el rendimiento de las campañas publicitarias digitales a nivel geográfico. La Ad Network que presta los datos para llevar a cabo la investigación, se especializa en la distribución de publicidad destinada a hispanos en Estados Unidos. Por ende, el rendimiento de las campañas publicitarias analizadas, cumplirá con la particularidad de depender del número de hispanos en el lugar geográfico en donde se distribuyan. Este y muchos otros factores deberán ser tenidos en cuenta al elegir las entradas de la RNA en cuestión. Para simplificar la elección de dichos datos, se han dividido las variables de entrada en tres grupos:

- Data Ad Network: Datos recopilados por la Ad Network, a través de campañas anteriores.
- Data Censo: Datos del censo de los Estados Unidos del año 2010
- Data Campaña: Datos recopilados de las campañas que se estarán analizando actualmente.

4.1 Data Ad Network

Los datos referentes a la Ad Network son los recopilados a lo largo de la existencia de la agencia. Entre los datos necesarios, se encuentra el alcance a nivel geográfico de la empresa. Más precisamente, se requiere la cantidad de usuarios alcanzados diariamente en los diferentes counties estadounidenses. Otro factor influyente, es la cantidad de usuarios con dispositivos móviles, dado que el

rendimiento de campañas destinadas al segmento de dispositivos móviles suele ser más alto.

Hay que tener en cuenta que tanto la población hispana, como el número de usuarios móviles, están incrementando de manera mantenida. Esto los convierte en mercados atractivos para cualquier empresa que quiera posicionarse correctamente en la etapa de crecimiento.

4.2 Data Campaña

El proyecto analizará un total de 5 campañas publicitarias, con volumen de entrega de banners similar y con una duración no mayor al mes. Los tópicos referentes a las campañas serán heterogéneos, para poder identificar la variación del rendimiento de cada una de ellas, en las diferentes ubicaciones. Los tópicos de las campañas serán los siguientes:

1. Correo
2. Supermercado
3. Mascotas
4. Elementos de baño
5. Consultoría financiera

La diversidad en los temas abordados, determinará las variables elegidas en el tercer grupo de inputs.

A pesar de las diferencias entre campañas, los datos que se obtendrán serán los mismos para todas. La Ad Network tiene una apertura de datos por usuario, pero para simplificar el procedimiento, todas las variables que se nombrarán a continuación tendrán una apertura por county.

- Sistema Operativo
- Browser
- Idioma del Browser
- Fecha y Hora

- Impresiones por Usuario
- Canales de entrega
- Impresiones y Clicks

El Sistema Operativo (SO) de los usuarios, además de ayudar a identificar el hardware más utilizado por los individuos, también permite dividir a los usuarios por “Dispositivos Móviles” y “Dispositivos Fijos”. Como se indica en el capítulo 4.1, los dispositivos móviles suelen tener mejor rendimiento en las campañas publicitarias.

El Browser utilizado no es determinante en el valor final del CTR, aunque el Idioma del Browser ayuda a confirmar si el segmento alcanzado es primariamente hispano. Teniendo en cuenta que la empresa se especializa en entregar publicidad destinada a hispanos en Estados Unidos, se supone que el rendimiento óptimo se encontrará si se alcanza al segmento deseado. Igualmente, como se indica en el capítulo 3, la gran mayoría de los hispanos consideran como idioma materno al inglés. Esto indica que no es necesario encontrar un porcentaje alto de browsers configurados en español para asegurarse que el segmento alcanzado sea el correcto.

La información que se podrá obtener gracias a la fecha en la cual cada usuario vio la publicidad, servirá principalmente para identificar si la mayoría de las entregas se realizaron en los fines de semana o en los días laboralmente activos. En muchos casos, el valor del CTR varía ampliamente dependiendo del día en el que se envíen la mayoría de las impresiones.

La cantidad de veces que el usuario ve la misma publicidad, también llamado “frecuencia de entrega”, no es un dato menor. Si se muestra más una determinada cantidad de veces a lo largo de la campaña, el aviso puede tener un efecto negativo en el individuo y si se muestra muy pocas, puede que este no la note. Para cada campaña, hay una frecuencia óptima que minimiza la cantidad de impresiones necesarias para alcanzar la cantidad de Clicks deseados.

Los canales de entrega hacen referencia a la temática de las páginas web en donde la publicidad fue divisada. Entre las más importantes se encuentran sitios de Business, Entertainment, Sports, etc. Es evidente que mientras más

correlación exista entre la temática de la publicidad y la del sitio web, mayor será el CTR obtenido por la campaña.

Por último, es necesario contar con la cantidad total de Impresiones y Clicks por county para poder calcular el CTR obtenido.

4.3 Data Censo

Gran parte de la información recopilada se extrae del censo de Estados Unidos. Esta se representará gráficamente para aportar un impacto visual que ayude a establecer cuáles son los potenciales target de las diferentes campañas.

La información del censo de Estados Unidos está dividida en las siguientes secciones.

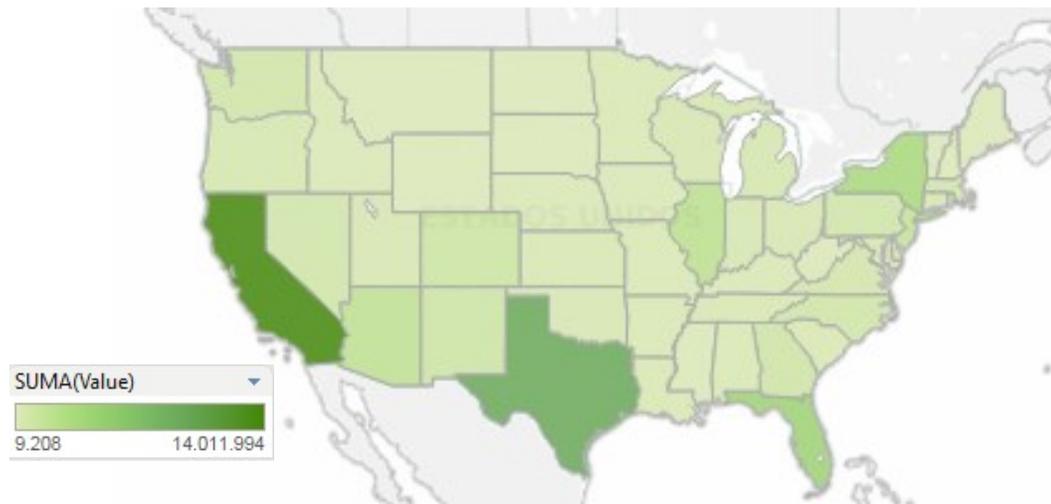
- Población Hispana
- Origen de población hispana
- Edad
- Sexo
- Educación
- Acculturation
- Ingreso
- Generación

4.4 Visualización de las Variables de Entrada

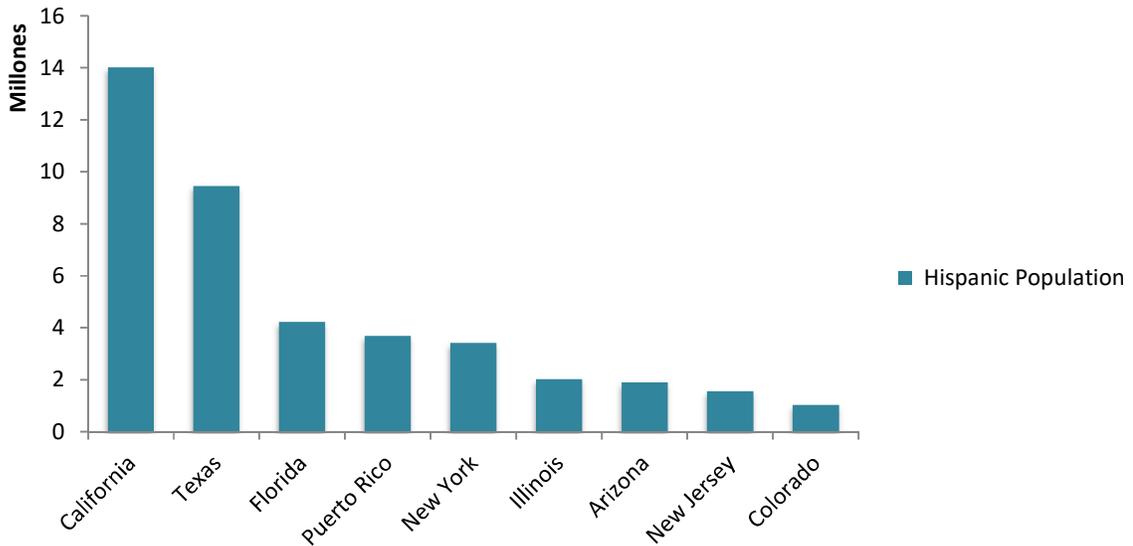
La visualización de las variables de entrada se realizará con una apertura por estado, con el fin de no comprometer el acuerdo de confidencialidad firmado con la empresa que prestó los datos para realizar el análisis. El entrenamiento de la red se realizará con una apertura a nivel county.

4.4.1 Población Hispana

En este grafico se puede observar la concentración de hispanos en Estados Unidos, con fuerte presencia de los mismos en California, Florida y Texas.



Población Hispana



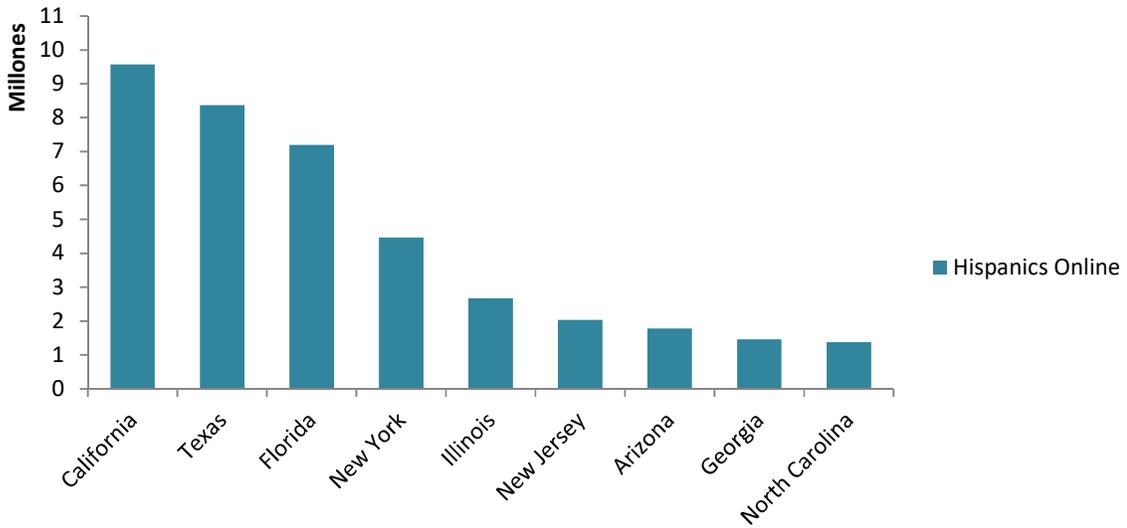
Se graficaron los primeros 9 estados con mayor población hispana, como se podía ver en el grafico anterior, los de mayor población son California y Texas y luego los siguientes están en niveles de población hispana muy similares. Florida un poco arriba del resto con más de 4 millones.

Esta información es relevante para el proyecto dado que focaliza su target publicitario en la comunidad hispana. Viendo así, cuales son los estados con mayor cantidad de hispanos dando una mayor cantidad de Clickers potenciales.

4.4.2 Población Hispana Online

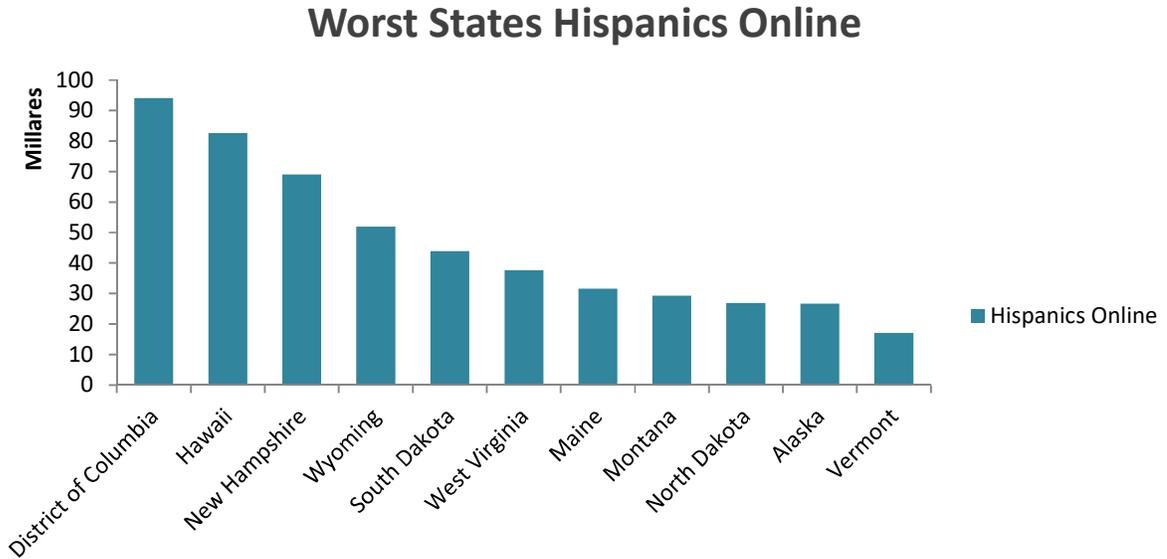
Otro dato que se consideró importante es la cantidad de hispanos online promedio por estado, claro está que este dato está relacionado a la cantidad de hispanos que viven en cada uno. La información representada gráficamente es la siguiente:

Top States Hispanics Online



A pesar de que no hay grandes diferencias entre población hispana e hispanos online en los primeros puestos, se pueden apreciar cambios en los siguientes. New Jersey tiene menos hispanos que Arizona, pero más hispanos online. Lo mismo sucede con Colorado y Georgia y North Carolina. Lo que se puede inferir es que la proporción de hispanos online sobre el total de hispanos varía entre estados y que la población por sí sola, no es un buen indicador del alcance que tendrá una campaña.

El mismo análisis se puede realizar para saber cuáles son los estados con menor cantidad de hispanos online. En la siguiente tabla se encuentran los peores 11.



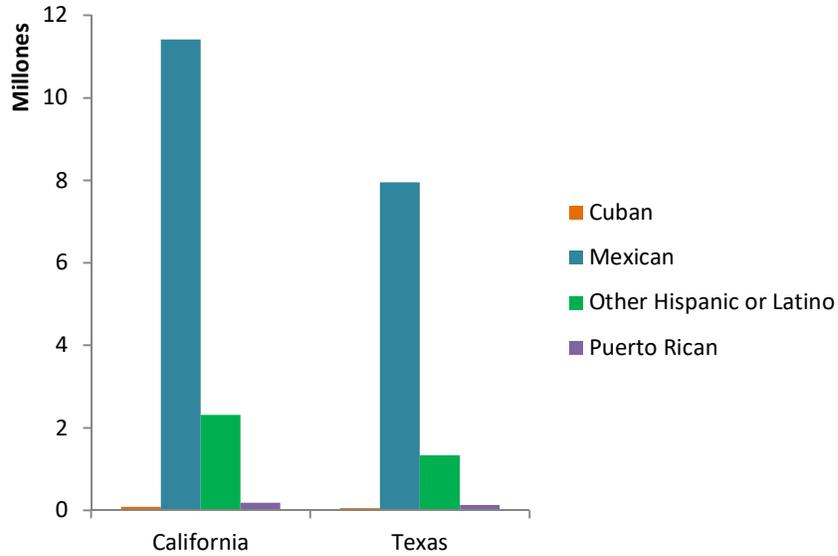
Estos estados cuentan con una menor probabilidad de éxito de campaña si se analiza la cantidad online. De todas formas, esto no es determinante ya que hay otras variables que pueden influir en el alcance de la campaña.

4.4.3 Origen de la Población Hispana

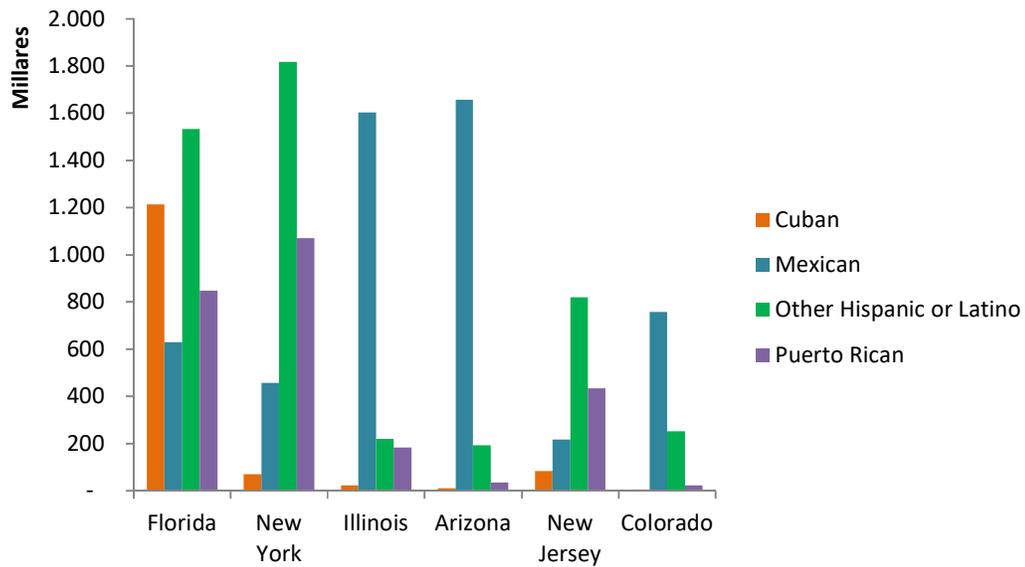
Dentro de la población hispana vamos a analizar de qué origen es cada hispano. Esto resulta útil para poder discriminar acorde al tipo de campaña va a ser más acorde al origen de cada hispano.

Hay 4 orígenes posibles: México, Puerto Rico, Cuba y Otros Hispanos. Se ejemplifica la distribución de los datos en los dos estados con mayor población hispana: California y Texas.

Redes Neuronales Artificiales – Optimización de Campañas Publicitarias Digitales



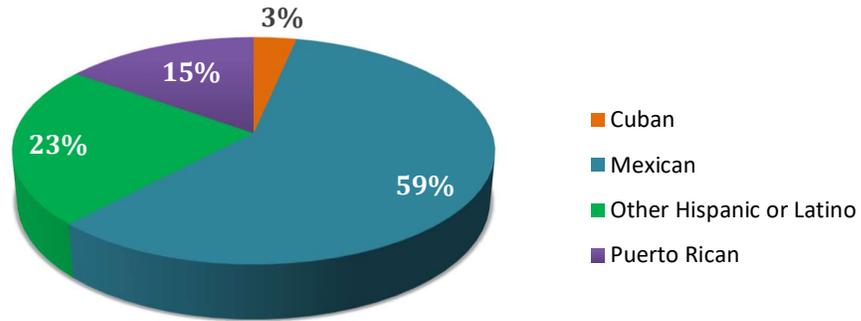
Las distribuciones en ambos estados no son equitativas, indicando la superioridad numérica de los mexicanos. En los estados que suceden a California y Texas en población hispana, se puede observar una distribución más pareja.



Predominando los latinos en los primeros 2 estados, siguiendo con gran número de mexicanos. Puede observarse que el estado más heterogéneo es florida, esto lo hace interesante dada su diversidad.

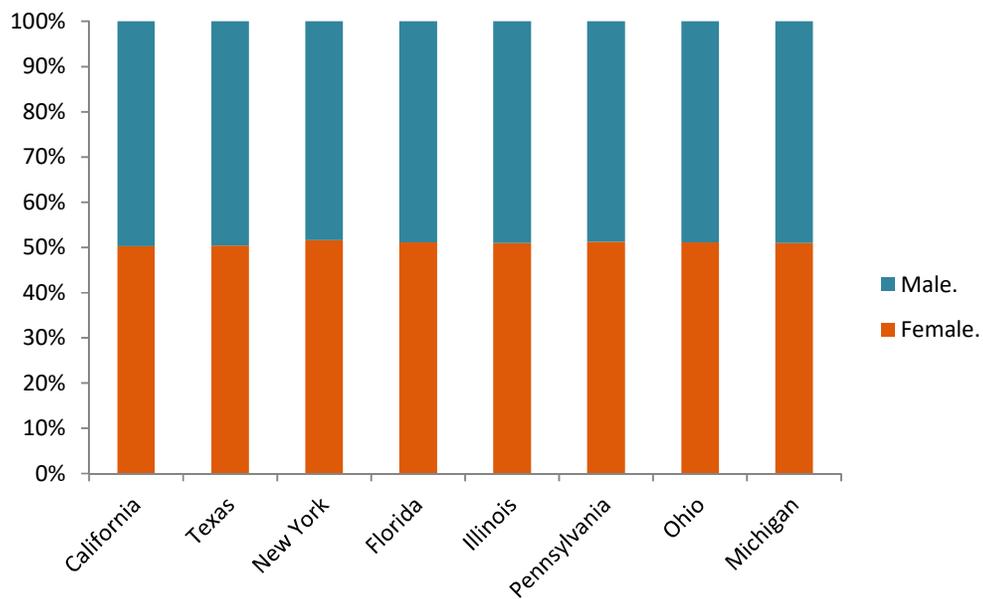
Por último se graficó el origen de los hispanos en Estados Unidos, predominando con más del 50% mexicanos. De un total de 50 millones.

Hispanics in United States

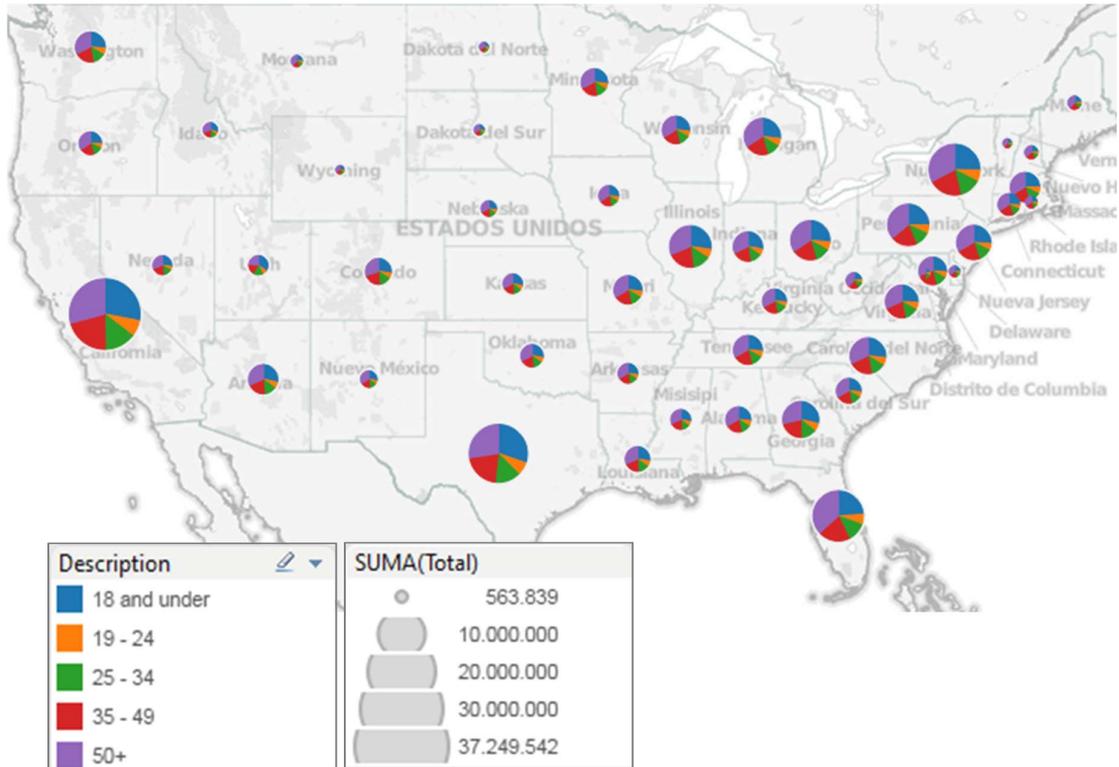


4.4.4 Edad y Sexo

En la siguiente tabla se muestra la relación entre cantidad de Hombres y Mujeres por estado. En este no se encuentran grandes diferencias, aunque si se abre por county, se pueden apreciar algunas variaciones.



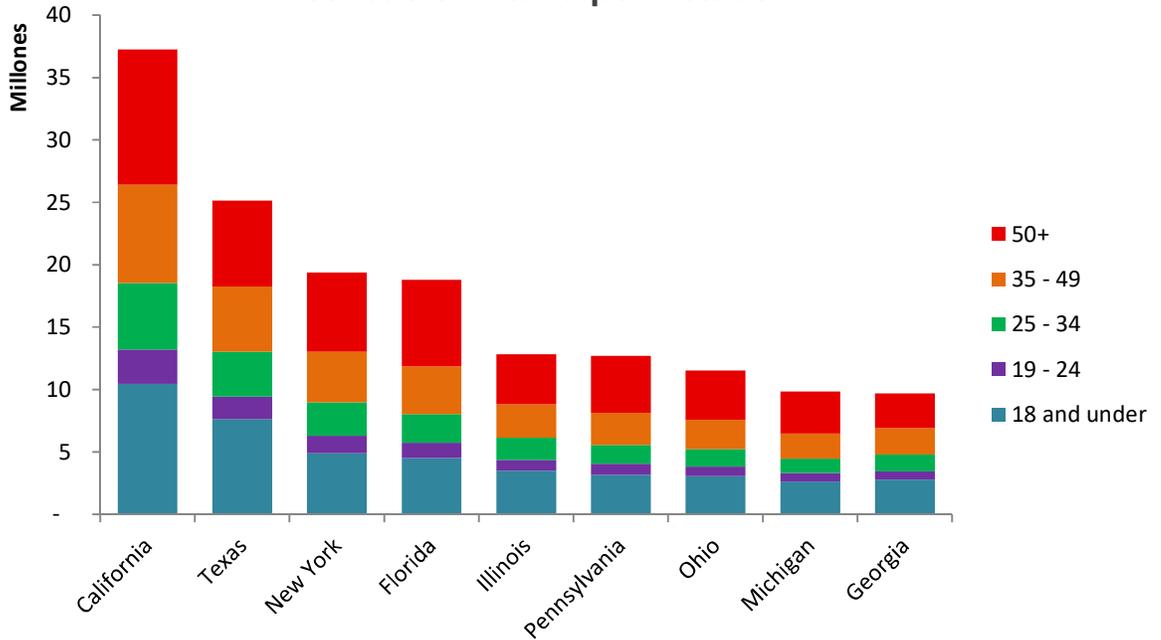
En el siguiente Mapa, se puede observar la distribución de edades de todos los estados. En el siguiente grafico se verán los principales estados a mayor detalle.



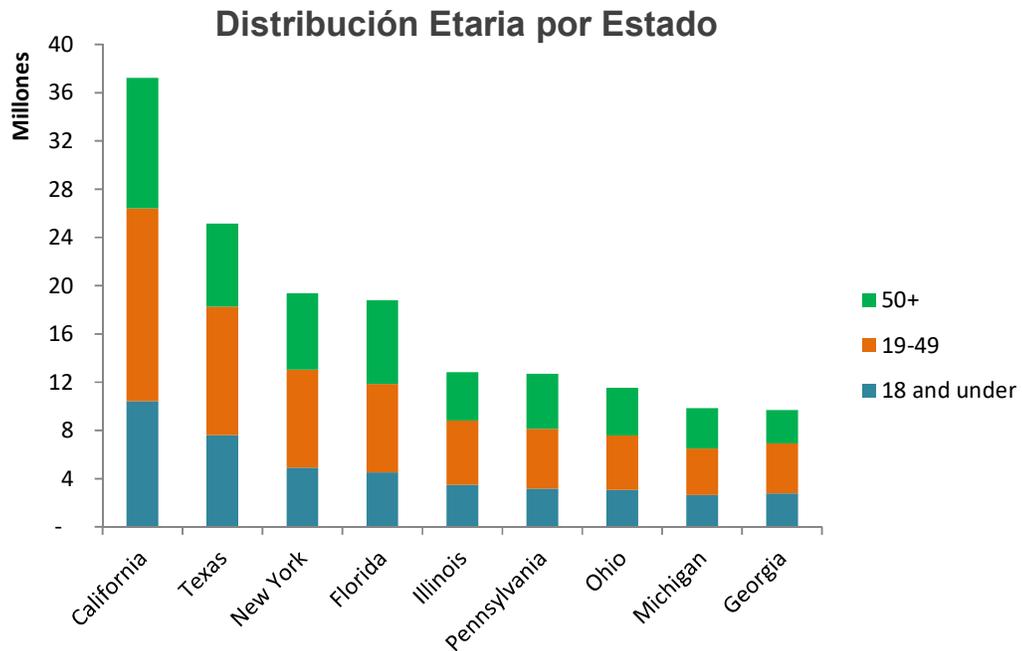
Más en detalle, para los principales estados, se cuenta con la siguiente distribución.

Redes Neuronales Artificiales – Optimización de Campañas Publicitarias Digitales

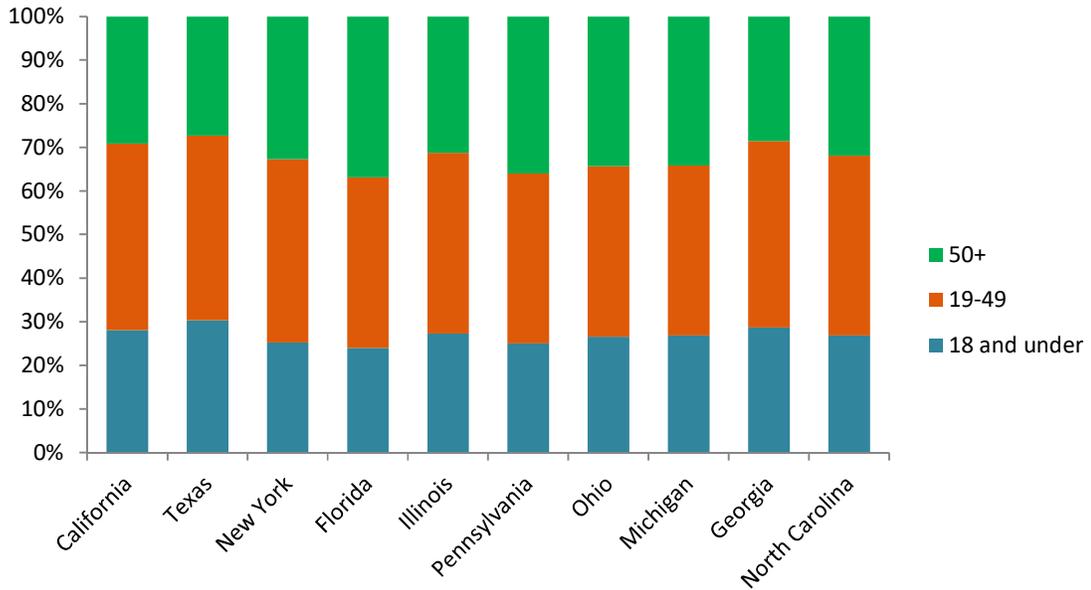
Distribución Etaria por Estado



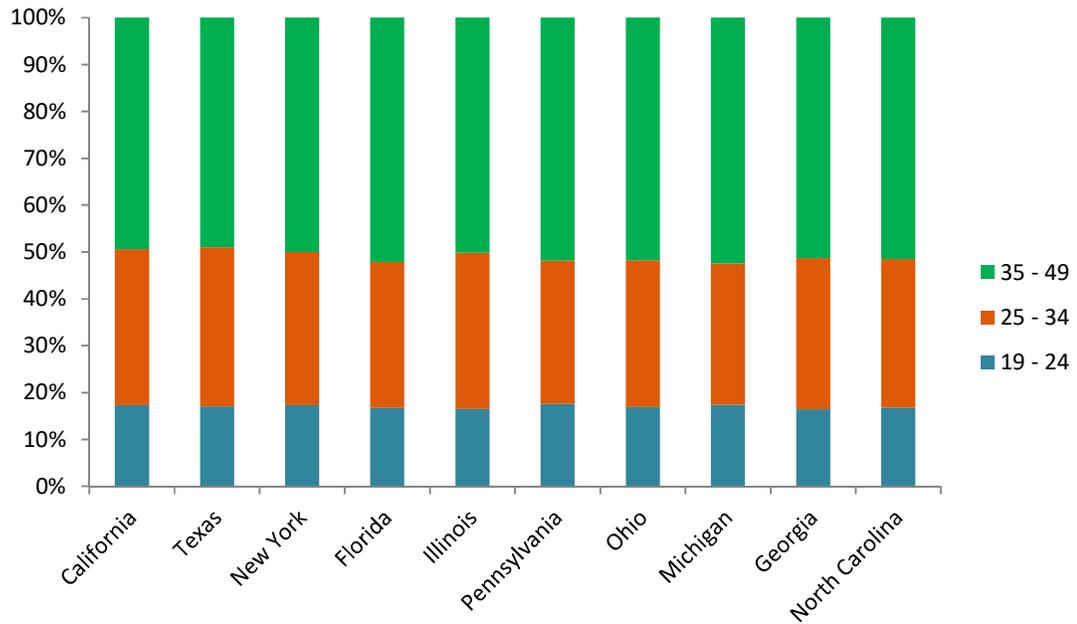
Se puede observar que los grupos con mayor población son "18 and under" o 50+. Por lo que se consideró agrupar las edades de 19 a 49 años. Dando los siguientes resultados.



Con los siguientes valores porcentuales:

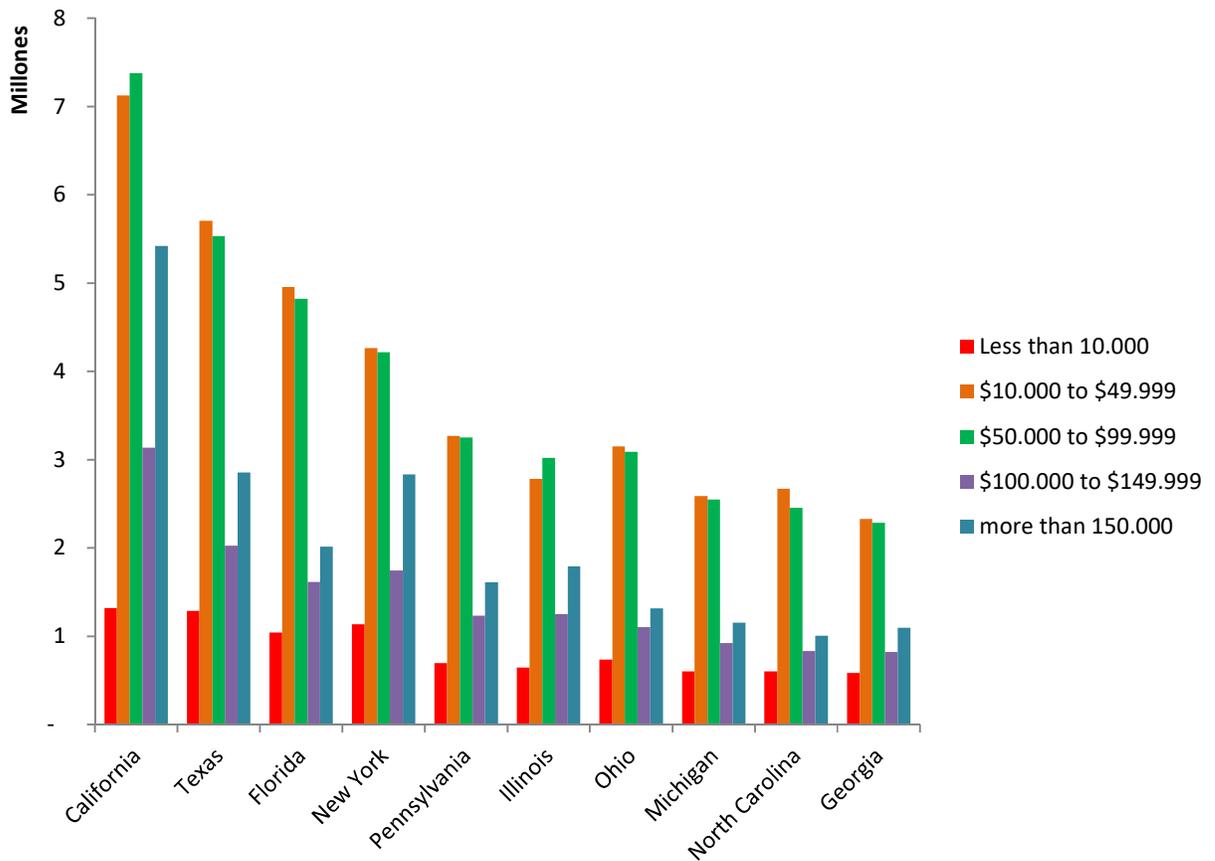


En cada estado se puede observar que para cada clasificación se tiene casi 1/3 de la porción. Particularmente para el rango de 19-49 años queda así:

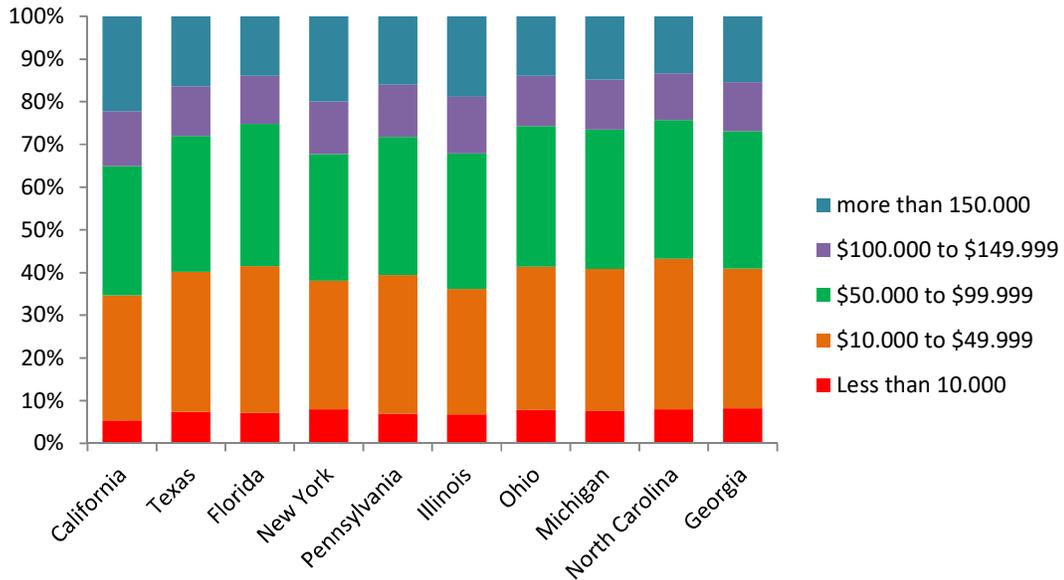


4.4.5 Household Income

Otro dato importante para determinar el tipo de campaña que se destinará a cada estado es el ingreso per cápita. En el gráfico siguiente, al estar en valor absoluto, no se puede apreciar si un estado tiene un ingreso promedio mayor a otro.



Es por esta razón que se toman los valores relativos.



La mayoría de la población estadounidense gana entre 10.000 y 100.000 USD por año. Y los estados con mayor % de población por encima de 150.000 USD son California y New York.

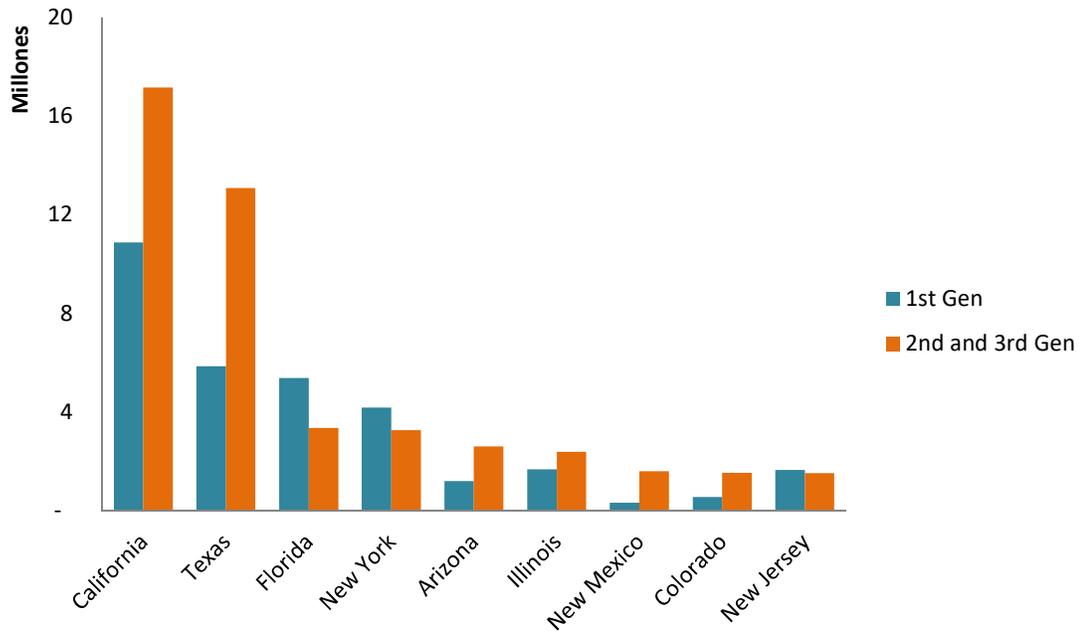
4.4.6 Generaciones Hispánicas

Se estudian las distintas generaciones de hispanos en Estados Unidos dadas sus características:

- 1era generación: Hispanos que llegaron de un país extranjero a los Estados Unidos.
- 2da y 3era Generación: Hispanos que nacieron en los Estados Unidos.

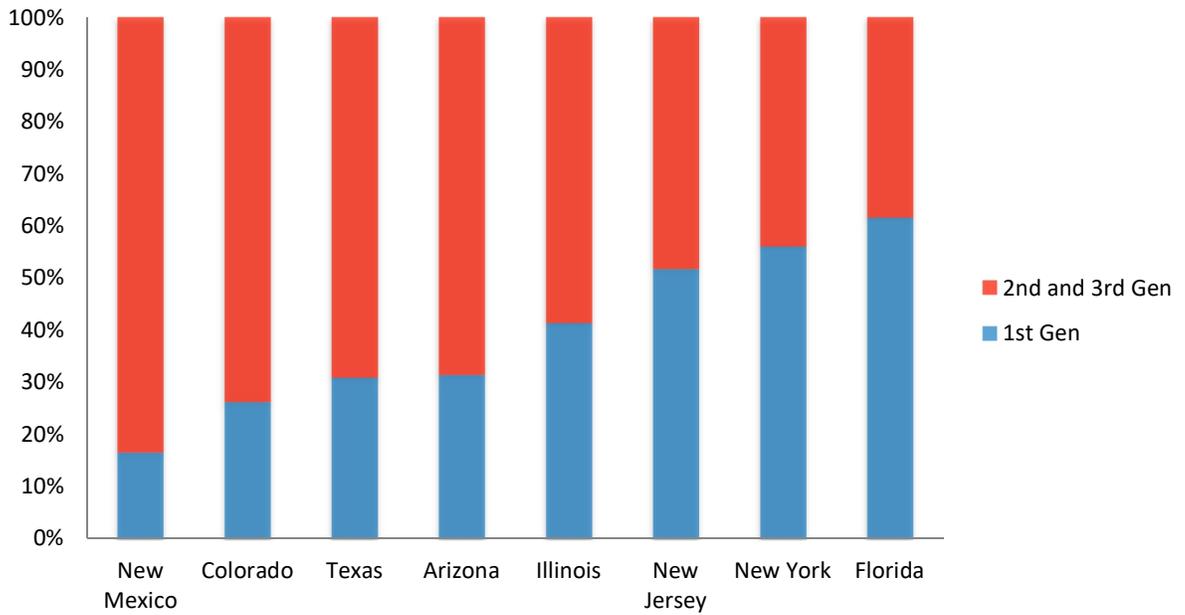
Este dato es de gran utilidad debido a las diferencias que son de interés para el proyecto. Esto es así porque, se está explotando hoy en día el Advertising a hispanos de habla inglesa, estos son los hispanos de 2da y 3era generación.

Top hispanic state's generations



Y a nivel porcentual la distribución es la siguiente:

Top States 2nd & 3rd Gen



Hay un gran número de segunda y tercera generación de hispanos, tiene un nivel de crecimiento alto y son un potencial mercado muy importante para las Ad-Network. Las publicidades en ingles a estos hispanos de primer lengua inglesa es algo todavía no explotado y muy atractivo.

5. Complementos de las RNA

Una red neuronal, a pesar de ser una herramienta predictiva avanzada, no alcanza su máximo potencial si no se la complementa correctamente. Los complementos que se pueden utilizar son varios, pero se pueden dividir en 2 grupos: de entrada y de salida. Los primeros son los que preparan los datos a ser analizados para su mejor aprovechamiento. Los segundos son los encargados de analizar los datos de salida para facilitar la configuración de la red que mejores resultados otorgue.

5.1 Complementos de Entrada

Las redes neuronales son capaces de incorporar una gran cantidad de variables de entrada y al mismo tiempo, de identificar cuáles son las que tienen mayor peso en el cálculo de la variable objetivo. El problema es que, a pesar de que la red minimiza el papel de las variables menos importantes, estas siguen originando pequeños errores en la predicción y es conveniente eliminarlas.

En primer paso es poder identificar cuáles son los inputs que no otorgan información nueva. Por ejemplo, si una variable de entrada fuese la población de una ciudad determinada y otras dos variables la población masculina y la femenina, se puede ver como sumando las dos últimas, se puede obtener la primera. Esto genera información redundante que la red debe procesar y por lo tanto, debe ser eliminada. Una solución posible es utilizar la variable población total y sólo la proporción de hombres o mujeres en la ciudad.

Otra manera de encarar esta problemática es utilizar herramientas estadísticas que faciliten la identificación de correlaciones entre variables y permitan seleccionar las entradas de mayor importancia. Una de ellas es el Análisis de Componentes Principales (PCA por sus siglas en inglés). Esta es una técnica estadística con la finalidad de reducir la dimensión o el número de

variables de una base de datos, perdiendo la menor cantidad de información posible. Luego de pasar los datos de entrada a través de la PCA, los nuevos componentes de la base serán una combinación lineal de los anteriores y además, serán independientes entre sí.

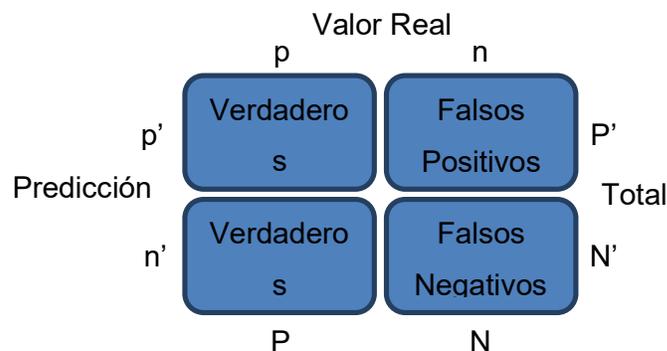
El análisis de componentes principales se puede dividir en 5 etapas:

1. Análisis de la matriz de correlaciones
2. Selección de factores
3. Análisis de matriz factorial
4. Interpretación de factores
5. Cálculo de puntuaciones factoriales

5.2 Complementos de Salida

Luego de analizar los datos con la red neuronal elegida, es necesario realizar un análisis para corroborar que el modelo elegido representa correctamente el caso de estudio. Una manera de hacerlo es calcular el porcentaje de aciertos sobre el número total de predicciones, aunque este enfoque deja de lado varios elementos que deben ser tenidos en cuenta.

Dos de las herramientas más utilizadas en el análisis de modelos predictivos son las "Curvas ROC" y los "Lift Charts". Las primeras, llamadas Receiver Operating Characteristic Curve, son una representación gráfica del ratio de verdaderos positivos, frente al ratio de falsos positivos que pueda devolver el modelo. Las 4 diferentes soluciones que puede devolver un modelo predictivo se visualizan en el Figura 5.2.1 a continuación.



Donde el Ratio de Verdaderos Positivos (VPR) y el de Falsos Positivos (FPR) se calcula como:

- $VPR = \frac{VP}{P}$
- $FPR = \frac{FP}{N}$

Luego, los valores obtenidos se representan en un espacio ROC, que consta del VPR en el eje 'Y' y del FPR en el 'X'. El rendimiento del modelo predictivo será mejor si su curva se acerca al extremo izquierdo superior y empeorará si se asemeja a la recta que representa a un modelo predictivo aleatorio. El espacio ROC se puede ver en la Figura 5.2.2.

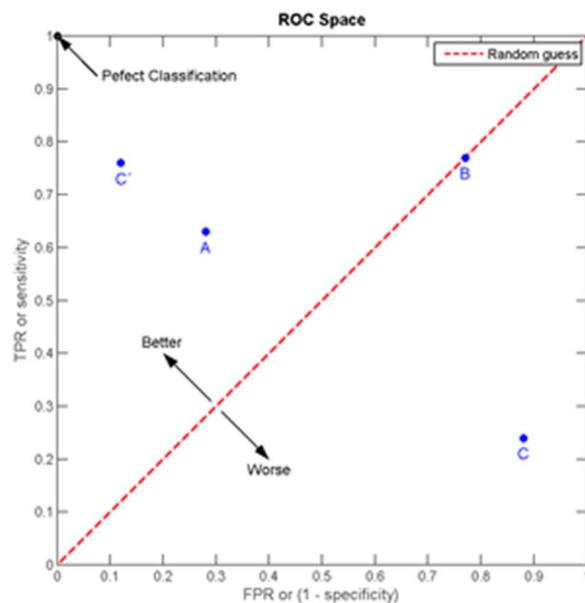


Figura 5.2.2 Espacio ROC

Una vez que se obtiene la curva característica del modelo, esta se puede clasificar según el área que tenga bajo la curva. La clasificación que se suele utilizar es:

- 0.5 – 0.6 → Modelo Malo
- 0.6 – 0.75 → Modelo Regular
- 0.75 – 0.9 → Modelo Bueno
- 0.9 – 0.97 → Modelo Muy Bueno
- 0.97 – 1 → Modelo Excelente

La segunda herramienta de análisis mencionada, Lift Charts, consiste en el armado de gráfico que indique cual es el porcentaje de aciertos del modelo, dependiendo de la confianza con la que predice cada valor. Por ejemplo, el modelo representado en el Gráfico 5.2.3 predice que una variable tiene más del 96% de probabilidad de tener un valor determinado, estará en lo correcto 349 de cada 355 veces.

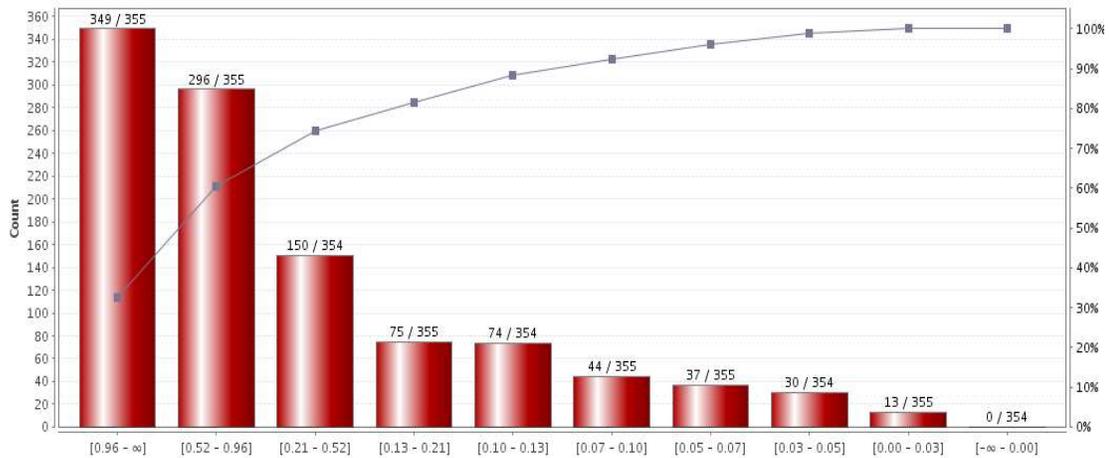


Gráfico 5.2.3 Lift Chart

Este tipo de gráfico muestra de manera discreta como se ve afectado el rendimiento de la red, a medida que baja la confianza con la que toma decisiones.

6. Armado de la Red Neuronal Artificial y sus Complementos

El armado de la red neuronal se divide en tres pasos. Primero se realizan diferentes procesos para mejorar los datos que ingresarán a la red. Esto se realiza con los Complementos de Entrada. Luego se entrena la REN y finalmente se utilizan los Complementos de Salida para mejorar el rendimiento de la herramienta predictiva.

6.1 Complementos de Entrada

El primer paso para comenzar con la RNA consiste en la elección y correcta manipulación de los datos de entrada. Como se explicó en el Capítulo 4, los datos de entrada estarán formados por la Data de la Ad Network, la Data de las Campañas y la Data del Censo. Dentro de estos grupos, se pueden encontrar variables que alcanzan valores considerablemente grandes (como la población por estado) y variables con valores porcentuales, que no superan la unidad (como el porcentaje de hombres o mujeres por estado). Las variaciones abruptas entre la amplitud de los valores que puede alcanzar cada variable, influye sobre la capacidad predictiva de la red y debe ser minimizado o eliminado. Para lograrlo, se deben Normalizar todas las variables de entrada.

La técnica de normalización está pensada para lograr que todos los valores de los atributos entren dentro de un rango específico. El método elegido en el proyecto final es el de Transformación-Z. También llamado Normalización Estadística, tiene como finalidad convertir la data en una distribución normal con media cero y varianza unitaria. Esto quiere decir que el 99.9% de los valores que esta transformación devuelva, estarán contenidos entre -3 y 3.

Una vez normalizada la data, se debe intentar eliminar toda la información redundante que la base de datos contenga. Esto se logra con la técnica PCA que se explica en el Capítulo 5.2. Al realizar un Análisis de Componentes Principales

en los datos normalizados de entrada, se reduce la cantidad de variables considerablemente, conservando la mayor parte de la información posible. En el Gráfico 6.1 se grafica la varianza acumulada de los componentes principales calculados por el análisis PCA. Aquí se puede observar como los primeros componentes logran retener la mayor parte de la información. Esto hace posible eliminar a los últimos, sin perder datos importantes. Si sólo se utilizan los primeros 8 componentes, se tendrá una varianza acumulada del 95,9%, pudiendo eliminar los últimos 3 componentes por su poco parte de información y haciendo la base de datos más fácil de procesar para la RNA.

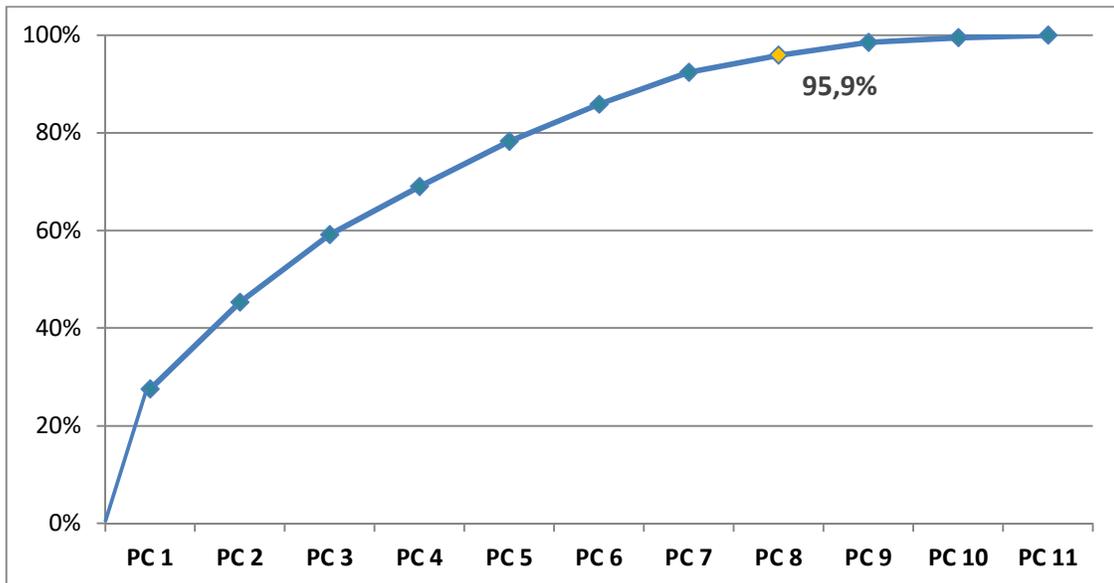


Gráfico 6.1 Varianza Acumulada de los Componentes Principales Calculados

Cada uno de estos componentes está formado por una combinación lineal de las variables de entrada normalizadas. El análisis de componente principal, devuelve una tabla con cada uno de los pesos de las variables, en los diferentes componentes. Se puede observar como cada una de las variables influye de mayor manera en un componente en particular (*Los valores de la tabla han sido*

modificados y los nombres de las variables utilizadas permanecerán ocultos, debido a la confidencialidad de los datos).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Variable 1	0.017	0.33	0.735	0.025	0.075	0.334	-0.32	0.057	0.236	0.046	0.042
Variable 2	-0.12	-0.21	-0.36	-0.32	0.125	0.788	-0.18	-0.15	0.135	0.061	0.023
Variable 3	0.048	0.041	-0.02	0.398	-0.83	0.345	0.183	-0.01	-0.03	-0	-0
Variable 4	0.128	-0.01	-0.21	0.735	0.258	0.021	-0.54	-0.19	-0.05	0.004	-0.01
Variable 5	-0.18	-0.64	0.286	0.011	0.031	0.161	-0.1	-0.04	-0.17	0.107	0.033
Variable 6	-0.051	0.296	-0.03	0.092	0.028	-0.02	0.039	-0.08	-0.01	0.205	0.771
Variable 7	-0.23	-0.43	0.134	0.089	-0.07	-0.1	-0.07	-0.36	0.642	0.033	-0.02
Variable 8	-0.052	0.241	-0.06	0.095	0.029	-0.02	0.036	0.05	-0.01	0.753	-0.29
Variable 9	-0.73	0.156	0.026	0.068	0.036	0.069	-0.03	0.029	-0.05	-0.08	-0.2
Variable 10	0.2	0.278	0.269	0.232	0.303	0.324	0.585	-0.38	-0.01	-0.03	-0.06
Variable 11	0.131	0.473	0.189	-0.34	-0.24	-0.05	-0.32	-0.59	-0.14	-0.01	-0.08

Gráfico 6.1 Varianza Acumulada de los Componentes Principales Calculados

Al eliminar los últimos tres componentes, dado que sólo aportan el 4.1% varianza total, se están eliminando los principales aportes de las Variables 6, 7 y 8. De esta manera se concluye en que estas tres variables, aportaban información redundante.

6.1.1 Realización de la Normalización y el PCA

El programa utilizado para realizar los procesos de Normalización de variables y Análisis de Componentes Principales es llamado RapidMiner. Es un programa gratuito destinado a la minería de datos. Cada uno de los procesos se visualiza en bloques, haciendo más simple la visualización y el entendimiento del proceso.

La Figura 6.1.1 muestra los pasos procesos elegidos para realizar la normalización y el PCA.

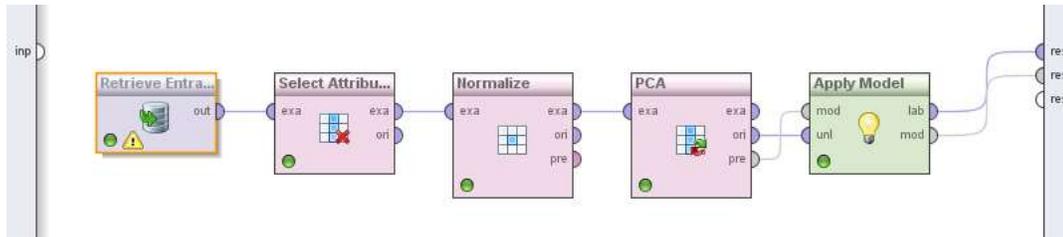


Figura 6.1.1 Proceso de Normalización y PCA en RapidMiner

El primer proceso se denomina Retrieve y es el encargado de leer los archivos que conforman la base de datos. En este caso, busca los archivos destinados a la entrada y salida de la RNA.

El segundo proceso es el de Select Attribute. Este permite seleccionar los datos que serán utilizados en el Análisis de Componentes Principales. Dado que la base de datos cuenta con variables tales como CountyId (un número que identifica a los diferentes counties), el nombre de la campaña publicitaria que se mostró o hasta la salida de la red (el valor de CTR por county), sería erróneo incluir a todas las variables en el proceso de PCA. Las variables que se destinan para este proceso son únicamente las de entrada, con atributos numéricos.

El tercer proceso es el de Normalización. Este permite elegir entre varios métodos diferentes de normalización, entre ellos Z-Transformation, Range-Transformation, Interquartile-Range, etc. Como se explicó en el capítulo anterior, se elige el método Z-Transformation.

El cuarto elemento es el de PCA, es el que crea el modelo de análisis de Componentes Principales necesario para la base de datos determinada. Este permite elegir cual será al varianza acumulada umbral a partir de la cual se dejarán de tener en cuenta los Componentes Principales. En este caso, la elección fue del 95%, dejando fuera a los últimos 3 componentes.

El último elemento es el encargado de aplicar el modelo de PCA a la base de datos y devolver la base nueva, con los componentes principales resultantes.

6.2 Red Neuronal Artificial

Una vez que los datos de entrada fueron manipulados correctamente, es momento de armar la RNA. El primer paso es volver a juntar los datos que se dejaron fuera del PCA con los resultados de ese análisis. Una vez que la base de datos se completa de vuelta (datos de entrada y salida), se puede proceder a la configuración de la red. Los diferentes factores que se pueden modificar en la red, teniendo en cuenta que se trabajará con una Feed-Forward Neural Network entrenada mediante un algoritmo de Back-Propagation, son los siguientes:

- Cantidad de Capas Ocultas
- Cantidad de Neuronas por Capa
- Cantidad de Ciclos de Entrenamiento
- Tasa de Aprendizaje
- Error Épsilon

La tasa de aprendizaje indica cuanto se modifican los pesos en cada ciclo de entrenamiento. El error épsilon determina que si el error durante los ciclos de entrenamiento es menor a un valor determinado, el entrenamiento de la red finaliza.

Cada vez que se modifique uno de los factores que caracterizan a la red, su poder predictivo cambiará. Para poder elegir la configuración óptima, hace falta poder analizar los datos con los complementos de salida debidos.

6.3 Complementos de Salida

Finalmente, luego de que el modelo de RNA ha sido validado, se procede al análisis de resultados. Este se lleva a cabo con los complementos de salida mencionados en el Capítulo 5.2. El proceso de análisis se realizará sucesivas veces, con el fin de calcular la variación en la capacidad predictiva de la red, cada vez que se modifican los factores que la caracterizan. Los factores que sufrirán cambios en la red serán:

- Cantidad de Capas Ocultas
- Cantidad de Neuronas por Capa
- Cantidad de Ciclos de Entrenamiento
- La tasa de Aprendizaje

También hay que tener en cuenta el hecho de que la red no calculará el valor del CTR en sí, sino que intentará predecir si este es Malo, Bueno o Muy Bueno. Esto genera una nueva incógnita en el problema: ¿Cuáles son los valores que se utilizarán para definir la calidad del CTR? Estos valores deberán ser razonables dentro del mercado de la publicidad digital y también tendrán que lograr resultados buenos en la RNA. Por ejemplo, decir que un CTR debe ser menor a 0,05% para ser catalogado como Malo, no sería práctico dado que esto ocurre en muy pocos casos y aunque la red fuera buena prediciéndolo, no tendría un uso aplicable. Por otro lado, indicar que para ser muy bueno, el CTR debe ser mayor a 0,13% puede ser correcto, aunque si la red predijera mejor cuales son los valores por encima de 0,135%, se debería adoptar este valor.

6.4 Armado de la RNA

Para armar la RNA se utilizará el mismo programa que para realizar el análisis PCA y la Normalización. El RapiMiner cuenta con todas las herramientas necesarias para crear y modificar todos los factores identificados en el Capítulo 6.3. En la Figura 6.4.1 se pueden observar cuales son la herramientas necesarias para el armado de una Feed-Forward Neural Network entrenada mediante un algoritmo de Back-Propagation.

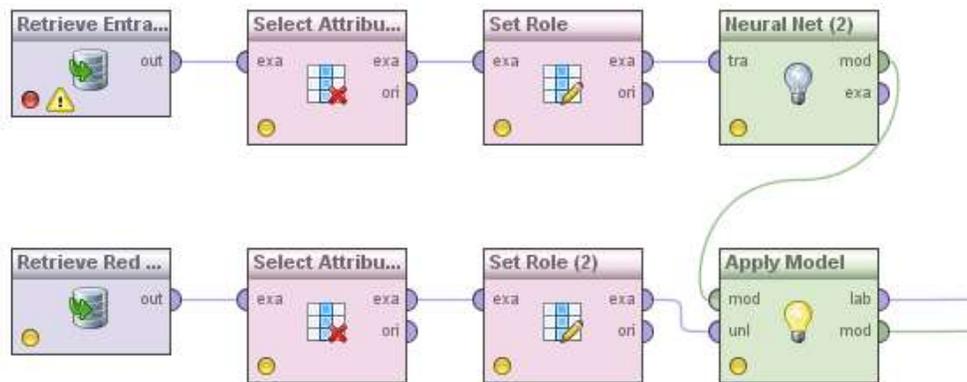


Figura 6.4.1 Armado de RNA en RapidMiner

El primer paso consiste en separar todo el set de datos en dos grupos:

- Set de entrenamiento
- Set de testeo

El set de entrenamiento será utilizado para formar la herramienta predictiva y el de testeo para validarla.

Las primeras dos herramientas utilizadas en el armado de la RNA fueron explicadas en el Capítulo 6.1.1. En la primera, Retrieve, se ingresa la base de datos con los datos de entrada y de salida. En la fila superior se ingresará el set de entrenamiento y en la inferior el de testeo. Luego, Select Attribute selecciona los datos que formarán parte del proceso. El tercer paso, Set Role, sirve para indicarle a la red, cuál será la variable objetivo o la salida.

En este caso se entrenarán dos redes diferentes con los mismos set de entrada pero con distintas salidas. La primera tendrá una salida binaria que indique con el valor unitario, si el CTR ha sido Bueno y con un cero, si ha sido Malo. La segunda red tendrá una salida similar, aunque indicará con un uno, si el valor de CTR es Muy Bueno y con un cero, si el valor del CTR es Bueno o Malo.

Por último, a Apply Model le ingresa el modelo de red entrenado con un set de datos de entrada y un set de datos de validación para contrastar los resultados.

6.5 Armado de los Complementos de Salida

Los complementos de salida son importantes para la interpretación de los resultados de la RNA. Los mismos se anexan al modelo creado en el Capítulo 6.4. En la Figura 6.5.1 se pueden ver sumados a la RNA, las herramientas necesarias para realizar el análisis de RC y el Lift Chart.

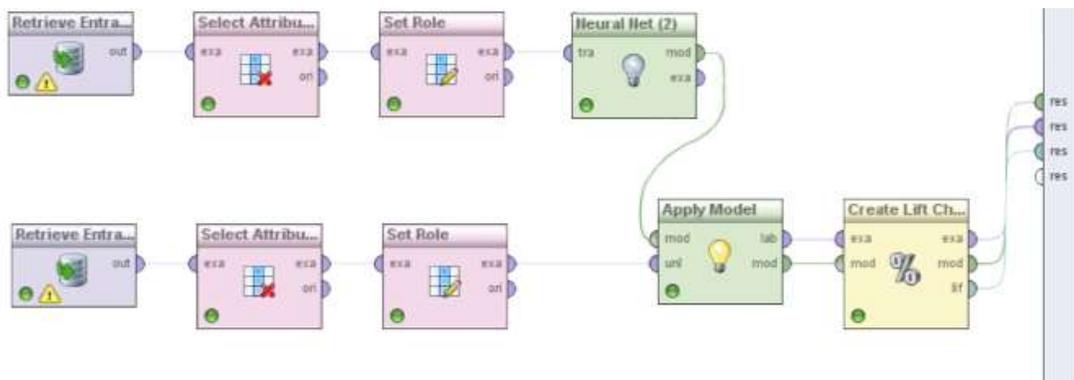


Figura 6.5.1 Anexo de ROC y Lift Chart a RNA en RapidMiner

La herramienta correspondiente al Lift Chart es la que se encuentran a la derecha, con el símbolo de porcentaje en el centro.

7. Optimización de Red Neuronal Artificial

En el siguiente capítulo, se modificarán las variables que influyen sobre el rendimiento de la red, con el fin de elegir la configuración óptima.

7.1 Ciclos de entrenamiento

Se evalúan los distintos resultados que se obtienen acorde al número de ciclos de entrenamiento. Se consideran los siguientes ciclos:

- 100
- 500
- 750
- 1000
- 1500

1000 Ciclos de Entrenamiento

	Total	Aciertos	%
<u>1</u>	1172	833	71.08%
<u>0</u>	2374	2326	97.98%

Total	Aciertos	%
3546	3159	89.09%

La red neuronal da como resultado una predicción del 89,09%, con un total de 3159 aciertos de 3546 posibles. Se obtiene una predicción del 71,08% de los 1, significando que el CTR es superior a 0,12%.

	<u>0</u>	<u>1</u>
<u>0</u>	2326	48
<u>1</u>	339	833

%	<u>0</u>	<u>1</u>
<u>0</u>	97.98%	2.02%
<u>1</u>	28.92%	71.08%

En los cuadros de doble entrada, se pueden apreciar los falsos positivos (predicciones positivas erróneas) y la cantidad y el porcentaje de valores positivos que fueron ignorados. Estos factores serán cruciales a la hora de elegir la configuración del modelo predictivo.

En la Figura 7.1.1 se puede observar el gráfico Lift Chart generado para la configuración actual. En él se aprecia que de las predicciones realizadas con una confianza mayor a 94%, la cantidad de predicciones positivas correctas, alcanzan el 60% de las predicciones positivas correctas totales.

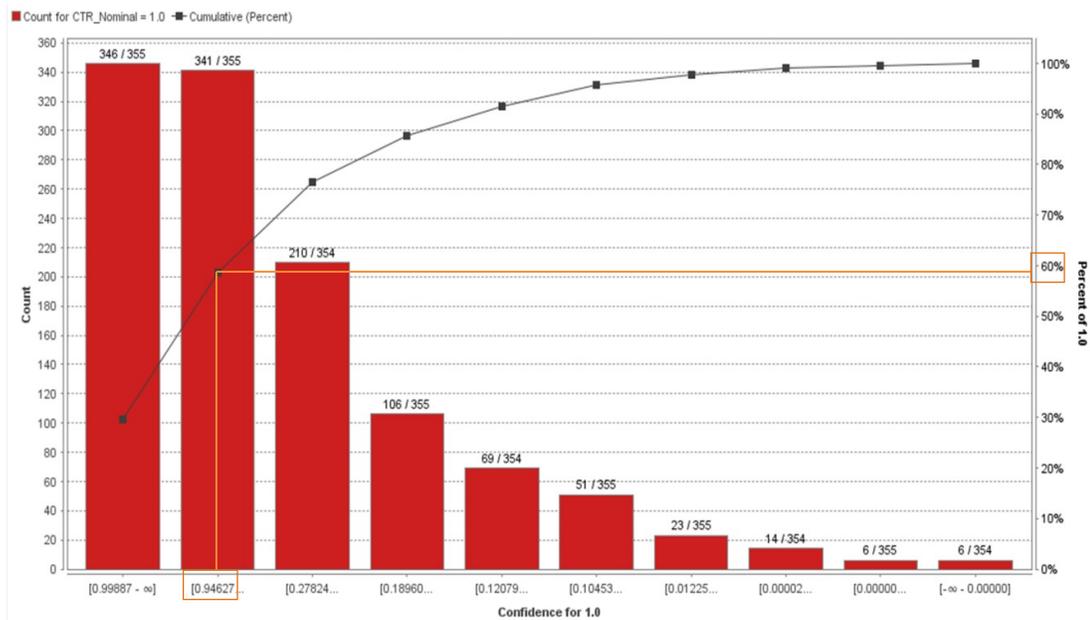


Figura 7.1.1 Lift Chart generado para 1000 Ciclos de Entrenamiento en RapidMiner

El segundo gráfico que se obtiene de los complementos de salida es el de ROC. En este se comparan 3 técnicas predictivas populares:

- Redes Neuronales (Azul)
- Árboles de Decisión (Verde)
- Bayes (Rojo)

Se puede ver como Bayes es peor que la red neuronal en todo momento y como el Árbol de Decisión no es confiable en todo el espectro del espacio ROC. Es por eso que se decide seguir utilizando Redes Neuronales a lo largo del trabajo.

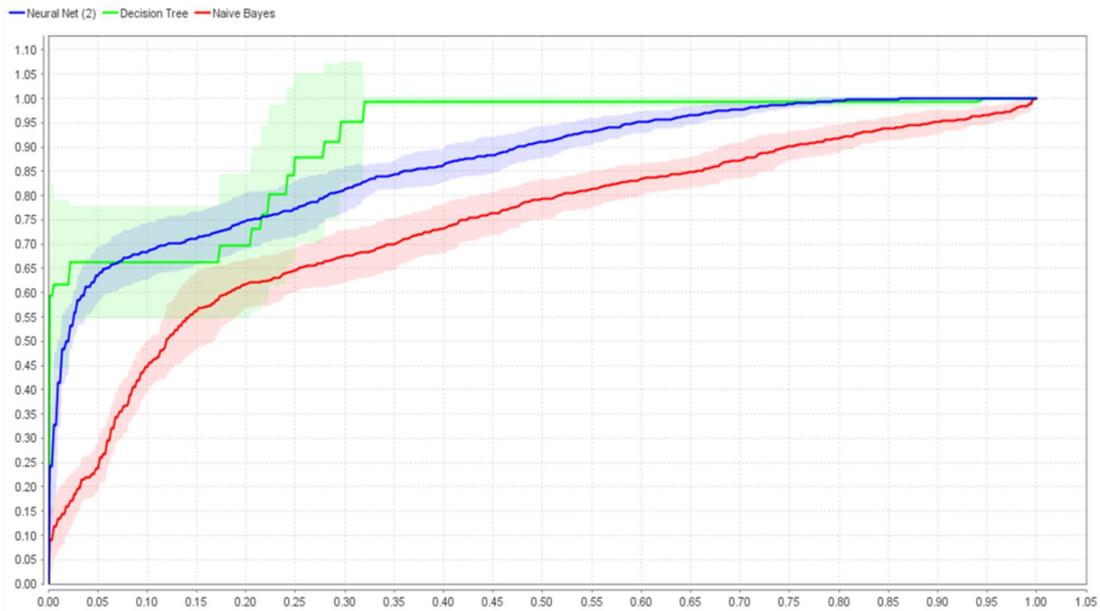


Figura 7.1.2 ROC generado para Red Neuronal, Árbol de Decisión y Bayes en RapidMiner

750 Ciclos de Entrenamiento

	Total	Aciertos	%
<u>1</u>	1172	825	70.39%
<u>0</u>	2374	2333	98.27%

Total	Aciertos	%
3546	3158	89.06%

La red neuronal da como resultado una predicción del 89,06%, con un total de 3158 aciertos de 3546 posibles. La cantidad de aciertos en los valores positivos disminuye, aunque no de manera apreciable.

Dado que el objetivo de esta red es identificar las campañas Muy Buenas o Buenas, se considera de mayor utilidad a la red con mayor capacidad predictiva de valores positivos.

	<u>0</u>	<u>1</u>
<u>0</u>	2333	41
<u>1</u>	347	825

%	<u>0</u>	<u>1</u>
<u>0</u>	98,27%	1,73%
<u>1</u>	29,61%	70,39%

1500 Ciclos de Entrenamiento

	Total	Aciertos	%
<u>1</u>	1172	824	70.307%
<u>0</u>	2374	2339	98.526%

Total	Aciertos	%
3546	3163	89.20%

Se obtienen los mismos resultados que en las anteriores corridas, con una pequeña diferencia en los aciertos positivos. No existen diferencias apreciables con respecto a los resultados obtenidos con 1000 Ciclos de Entrenamiento, y por ende, se decide seguir con la primera configuración.

	<u>0</u>	<u>1</u>
<u>0</u>	2339	35
<u>1</u>	348	824

%	<u>0</u>	<u>1</u>
<u>0</u>	98,53%	1,47%
<u>1</u>	29,69%	70,31%

Incluso la RNA tiene un número menor de aciertos positivos, con una cantidad de 824 contra 833 obtenidos en la Red entrenada con 1000 ciclos.

500 Ciclos de Entrenamiento

	Total	Aciertos	%
<u>1</u>	1172	816	69.62%
<u>0</u>	2374	2329	98.10%

Total	Aciertos	%
3546	3145	88.69%

Se obtienen resultados muy similares a las anteriores corridas, con leves diferencias, en especial en el acierto de valores positivos. Por lo que no hay

diferencias apreciables con respecto al número de ciclos de entrenamiento, tomando como óptimo el de 1000 ciclos.

	<u>0</u>	<u>1</u>		<u>0</u>	<u>1</u>
<u>0</u>	2329	45	%	98,10%	1,90%
<u>1</u>	356	816	%	30,38%	69,62%

100 Ciclos de Entrenamiento

	Unos	Aciertos	%	Total	Aciertos	%
<u>1</u>	1172	801	68.34%	3546	3082	86.91%
<u>0</u>	2374	2281	96.08%			

Se obtienen los mismos resultados que en los casos mencionados con anterioridad, con una pequeña diferencia en el acierto de valores positivos. No hay diferencias apreciables y por ende, se continúa tomando como el óptimo de 1000 ciclos.

Sin embargo, aumento la cantidad de predicciones falsas positivas. Este valor prácticamente se duplica.

	<u>0</u>	<u>1</u>	%	<u>0</u>	<u>1</u>
<u>0</u>	2281	93	96.08%	96.08%	3.92%
<u>1</u>	371	801	31.66%	68.34%	

7.2 Cantidad de Capas Ocultas y de Neuronas Por Capa

Una vez que se elige trabajar con 1000 Ciclo de Entrenamiento, se procede a realizar el cálculo de la cantidad de Capas Ocultas y de Neuronas por Capa que se deben utilizar en la red. Para lograrlo se realizan pruebas con una sola capa oculta, variando la cantidad de neuronas que la componen, y luego con dos capas ocultas, variando el número de neuronas totales.

Para el primer paso, se utilizan en la capa oculta un total de:

- 12 Neuronas
- 16 Neuronas
- 20 Neuronas

12 Neuronas

En el primer caso, se plantea una red con una capa oculta de 12 neuronas. La misma se configura con un total de 1000 ciclos de entrenamiento, dado que este es el número óptimo de corridas de entrenamiento que se encontró en el capítulo 7.1.

Los resultados son los siguientes:

	<u>0</u>	<u>1</u>
<u>0</u>	2301	73
<u>1</u>	356	816

%	<u>0</u>	<u>1</u>
<u>0</u>	96.93%	3.07%
<u>1</u>	30.38%	69.62%

16 Neuronas

Al entrenar la red con 16 neuronas en la capa oculta, se puede observar una mejora considerable. El incremento en la capacidad predictiva de los valores positivos es mayor al 10%.

	<u>0</u>	<u>1</u>
<u>0</u>	2203	171
<u>1</u>	226	946

%	<u>0</u>	<u>1</u>
<u>0</u>	92.80%	7.20%
<u>1</u>	19.28%	80.72%

20 Neuronas

Al agregar un mayor número de neuronas, la red empieza a perder efectividad. Esto puede indicar que el modelo no es lo suficientemente complejo

como para utilizar más de 16 neuronas o que el modelo es de una complejidad tal, que no puede ser resuelto por una sola capa de neuronas.

	<u>0</u>	<u>1</u>		%	<u>0</u>	<u>1</u>
<u>0</u>	2318	56		<u>0</u>	97.64%	2.36%
<u>1</u>	289	883		<u>1</u>	24.66%	75.34%

En resumen, los aciertos de los valores positivos, obtenidos por una sola capa oculta son:

	Unos	Aciertos	%
<u>12</u>	1172	816	69.62%
<u>16</u>	1172	946	80.72%
<u>20</u>	1172	883	75.34%

El segundo paso consiste en corroborar si la complejidad de la red es tal, que la utilización de una segunda capa, incrementa la capacidad predictiva.

16 & 12 Neuronas por capa oculta

Con la aplicación de una segunda capa de neuronas ocultas, la capacidad predictiva del modelo no empeora. Se acierta prácticamente el 79% de los valores positivos, es decir, con CTR mayor a 0,012%.

	Total	Aciertos	%
<u>1</u>	1172	922	78,67%
<u>0</u>	2374	2292	96,55%

Total	Aciertos	%
3546	3214	90,64%

En cuanto a las campañas con un rendimiento inferior a 0,012%, la red tiene una certeza el 96,55%. En valores absolutos predice 2292 valores nulos de 2374 totales.

Los resultados que se obtienen positivos y se puede considerar exitoso el entrenamiento de la RNA. Es por eso que se decide incrementar el número de neuronas en las capas.

	<u>0</u>	<u>1</u>
<u>0</u>	2292	82
<u>1</u>	250	922

%	<u>0</u>	<u>1</u>
<u>0</u>	96,55%	3,45%
<u>1</u>	21,33%	78,67%

20 & 12 Neuronas por capa oculta

La capacidad de predecir correctamente se incrementa al agregar nuevas neuronas al modelo. Los aciertos alcanzan valores de 91.79%, siendo este el valor más alto obtenido.

	Total	Aciertos	%
<u>1</u>	1172	1015	86,60%
<u>0</u>	2374	2240	94,36%

Total	Aciertos	%
3546	3255	91,79%

También se incrementa la capacidad de la red de diferenciar las campañas por encima del 0,012%.

	<u>0</u>	<u>1</u>
<u>0</u>	2240	134
<u>1</u>	157	1015

%	<u>0</u>	<u>1</u>
<u>0</u>	94,36%	5,64%
<u>1</u>	13,40%	86,60%

Se puede observar que al aumentar el número de neuronas en la primera capa oculta, se obtiene un aumento en el número de predicciones positivas correctas. Al ser la variable objetivo del sistema, se seguirán incrementando hasta alcanzar valores óptimos.

24 & 20 Neuronas por capa oculta

La red neuronal dio como resultado una predicción del 95,32%, un total de 3380 aciertos de 3546 posibles. Este es el valor tope de predicciones correctas, dado que si se siguiera incrementando al cantidad de neuronas, esta cifra comenzaría a caer.

	Total	Aciertos	%
<u>1</u>	1172	1036	88,40%
<u>0</u>	2374	2344	98,74%

Total	Aciertos	%
3546	3380	95,32%

Se obtuvo una predicción del 88,40% de los counties con CTR mayor a 0,012%. En cuanto a los de rendimientos inferiores, la red predijo correctamente el 98,74%.

	<u>0</u>	<u>1</u>
<u>0</u>	2344	30
<u>1</u>	136	1036

%	<u>0</u>	<u>1</u>
<u>0</u>	98,74%	1,26%
<u>1</u>	11,60%	88,40%

Se observa que aumentando el número de neuronas de ambas capas ocultas, da un aumento en la precisión de las predicciones. Y un incremento en el nivel de confianza de la red, que puede visualizarse en el Lift Chart de la Figura 7.2.1.

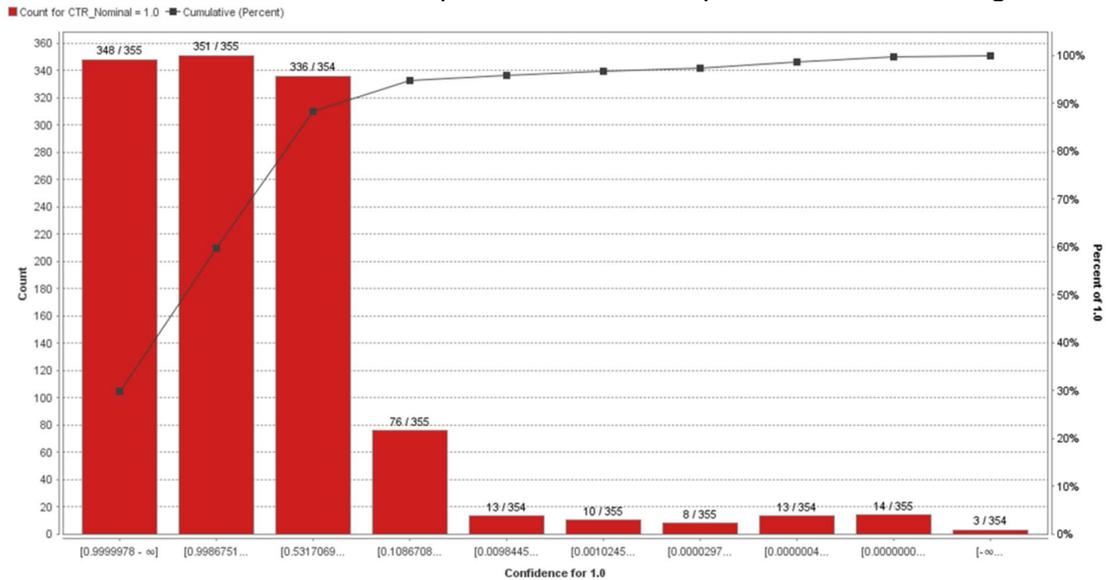


Figura 7.2.1 Lift Chart de RNA con 20 y 24 neuronas en sus capas ocultas

7.3 Constante de Aprendizaje

Aquí se evalúa como varía el rendimiento de la red al modificar la variable de aprendizaje. La misma es la que determina cuanto se modificarán los pesos de las neuronas que componen a la red, en cada ciclo de entrenamiento.

Por default, en los capítulos anteriores se estaba trabajando con una constante de valor igual a 0,3. Al incrementar esta variable, la RNA aprenderá más rápidamente, pero es error que tendrá al final del entrenamiento será mayor. Si se decrementa, la red aprenderá de manera más paulatinamente, pero el resultado final, si se realizan suficientes ciclos de entrenamiento, será más exacto.

7.3.1 Constante de aprendizaje 0,3

Los valores para esta constante son los mismos que los plantados en el capítulo anterior y solamente se utilizarán para contrastar los resultados con las redes entrenadas con las otras dos constantes de aprendizaje (una mayor y una menor).

	Total	Aciertos	%
<u>1</u>	1172	1036	88,40%
<u>0</u>	2374	2344	98,74%

Total	Aciertos	%
3546	3380	95,32%

	<u>0</u>	<u>1</u>
<u>0</u>	2344	30
<u>1</u>	136	1036

%	<u>0</u>	<u>1</u>
<u>0</u>	98,74%	1,26%
<u>1</u>	11,60%	88,40%

Como se puede ver, la cantidad total de aciertos fue del 95,32% y la cantidad de valores positivos bien predichos fueron del 88.40%. Se verá como varía esto a medida que se modifica la constante en cuestión.

7.3.2 Constante de aprendizaje 0,2

Los valores obtenidos para este análisis fueron alentadores. Hubo un decremento leve en la cantidad de aciertos totales y un incremento pronunciado en la cantidad de aciertos positivos.

	Total	Aciertos	%
<u>1</u>	1172	1059	90.36%
<u>0</u>	2374	2319	97.68%

Total	Aciertos	%
3546	3378	95.26%

	<u>0</u>	<u>1</u>
<u>0</u>	2319	55
<u>1</u>	113	1059

%	<u>0</u>	<u>1</u>
<u>0</u>	97.68%	2.32%
<u>1</u>	9.64%	90.36%

Lo más destacable de esta modificación es el incremento de aciertos positivos en casi 2%.

7.3.3 Constante de aprendizaje 0,5

Aquí los resultados de la red empeoran, indicando que la con la cantidad de ciclos de entrenamiento que se utilizan, no hace falta incrementar la variable de aprendizaje. Como se muestra en las tablas referentes a este último análisis, todas las variables analizadas decrecen.

	Total	Aciertos	%
<u>1</u>	1172	999	85.24%
<u>0</u>	2374	2332	98.23%

Total	Aciertos	%
3546	3331	93.94%

	<u>0</u>	<u>1</u>
<u>0</u>	2332	42
<u>1</u>	173	999

%	<u>0</u>	<u>1</u>
<u>0</u>	98.23%	1.77%
<u>1</u>	14.76%	85.24%

Se concluye que, con un número elevado de ciclos de entrenamiento, la constante de aprendizaje se puede disminuir para poder predecir con mayor precisión. En este caso, al aplicar constantes menores a 0,2, los resultados no mejoraban en gran medida y por eso, no se los incluyo en el trabajo.

7.4 Exigencia al CTR

En este capítulo se evalúa la influencia de la variación del valor CTR al definir la calidad de un county. En los casos anteriores se estaba tomando como valor de referencia un CTR mayor a 0,012%, dado que este es un valor muy cercano a la media de los CTRs analizados en este caso.

Variando el nivel de CTR no sólo se podrá ver la variación de la capacidad predictiva de la red, sino también la capacidad de la red de armar segmentos de diferentes cualidades. La idea inicial de este proyecto era poder no sólo predecir el rendimiento de una campaña publicitaria en un county, sino también poder clasificarlo como Malo, Bueno o Muy Bueno y así poder comercialarlo como inventario regular o Premium.

7.4.1 Clasificación Bueno-Muy Bueno:

CTR superior a 13

En este caso se evalúa a la red neuronal, modificando el valor de la variable de salida. Anteriormente, si el CTR era mayor a 0,012% se definía a la variable de salida con un valor unitario y de ser menor, con valor nulo. Ahora se impone que este valor debe superar el 0,013%. Se mantienen los 1000 ciclos de entrenamiento y 2 capas ocultas, con 24 y 20 neuronas respectivamente. De esta manera se le exige más al valor del CTR para ser calificado como Muy Bueno.

	Total	Aciertos	%
<u>1</u>	1068	956	89,51%
<u>0</u>	2478	2451	98,91%

Total	Aciertos	%
3546	3407	96,08%

La red neuronal dio como resultado una predicción del 96,08%, un total de 3407 aciertos de 3546 posibles. La red predijo correctamente un total de 89,51% de los counties con CTR Muy Bueno.

	<u>0</u>	<u>1</u>
<u>0</u>	2451	27
<u>1</u>	112	956

%	<u>0</u>	<u>1</u>
<u>0</u>	98,91%	1,09%
<u>1</u>	10,49%	89,51%

Si bien el cambio no es significativo con respecto a los resultados de CTR igual a 0,012%, la cantidad total de counties Muy Buenos disminuye en 100 unidades aproximadamente. Es por esta razón que se cree mejor conservar el valor de CTR anterior al momento de definir el segmento Premium y así ampliar la cantidad de counties a los que se les podrá enviar publicidad Premium.

7.4.2 Clasificación Malo-Bueno

El objetivo de esta red neuronal es diferenciar las campañas malas de las buenas, por eso se le exige un CTR mínimo. En esta sección se evalúa los resultados con los distintos CTR mínimos exigidos.

En caso de ser menor al valor establecido la red devuelve el valor unitario y en caso de ser mayor el nulo.

CTR inferior a 10

En este caso se le exige un CTR mínimo de 0,010%, en caso de ser mayor la campaña es Buena y si es menor Mala.

	Total	Aciertos	%
<u>1</u>	2116	2057	97,21%
<u>0</u>	1430	1324	92,59%

Total	Aciertos	%
3546	3381	95,35%

La red neuronal dio como resultado una predicción del 95,35%, un total de 3381 aciertos de 3546 posibles. La red da un 97% de aciertos de counties Malos como puede observarse en la tabla. Los counties con campañas con CTR mayor a 0,010% tienen un 92,59% de aciertos.

	<u>0</u>	<u>1</u>
<u>0</u>	1324	106
<u>1</u>	59	2057

%	<u>0</u>	<u>1</u>
<u>0</u>	92,59%	7,41%
<u>1</u>	2,79%	97,21%

A pesar de que ambos indicadores son muy altos, por arriba del 90%, se evaluará que resultados se obtienen variando la exigencia al CTR.

CTR inferior a 9

En comparación con el entrenamiento anterior, analizando valores de CTR inferiores a 0,009%, se observa que el porcentaje de aciertos sigue en niveles óptimos.

	Total	Aciertos	%
<u>1</u>	1976	1891	95,70%
<u>0</u>	1570	1469	93,57%

Total	Aciertos	%
3546	3360	94,75%

	<u>0</u>	<u>1</u>
<u>0</u>	1469	85
<u>1</u>	101	1891

%	<u>0</u>	<u>1</u>
<u>0</u>	93,57%	6,43%
<u>1</u>	5,11%	95,70%

Cuantitativamente, se puede decir que el resultado anterior es mejor, dada su mayor precisión de unos (campañas Malas). Aun así se considerará que exigirle más al CTR y analizar si los valores de análisis continúan superando el 90%, ya que esto define con ojo más crítico si un CTR es Bueno o Malo.

CTR inferior a 8

Los resultados obtenidos son similares a los anteriores. Variando el CTR aún se mantiene más de un 90% de aciertos en los unos, con un porcentaje total de aciertos de 95,69%.

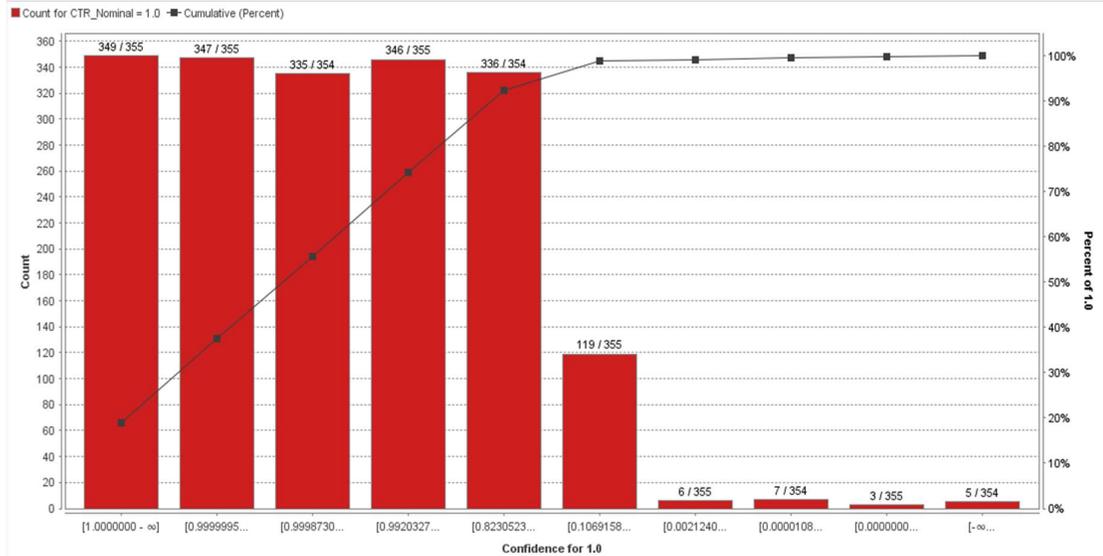
	Total	Aciertos	%
<u>1</u>	1853	1787	96,44%
<u>0</u>	1693	1606	94,86%

Total	Aciertos	%
3546	3393	95,69%

	<u>0</u>	<u>1</u>
<u>0</u>	1606	66
<u>1</u>	87	1787

%	<u>0</u>	<u>1</u>
<u>0</u>	94,86%	5,14%
<u>1</u>	4,70%	96,44%

Teniendo en cuenta sólo las predicciones que se hacen con una confianza mayor al 88%, se cubren más del 90% de los counties Malos como puede observarse en el Lift Chart de abajo.



Finalmente, se aconseja tomar como valores de corte al CTR 0,008% a la hora de definir counties Malos o Buenos. Esto se debe a los valores de las variables de análisis y a la capacidad de la red de clasificar con ojo crítico el rendimiento de un county.

7.5 Configuración Final de la RNA

La elección de la configuración de la RNA, estará basada en su capacidad predictiva sobre la variable objetivo. Esta no es la cantidad de aciertos totales, sino la de verdaderos positivos. La justificación reside en que el número de counties que puede ser clasificado como Bueno o Muy Bueno es mucho menor que el de counties malo, y por ende, más difícil de predecir. También se tendrá en cuenta la cantidad de falsos positivos, dado que si este número es elevado, el segmento que se intentará comercializar no producirá el valor de CTR buscado.

Teniendo en cuenta los resultados obtenidos en el Capítulo 7.1, se elegirá una cantidad de 1000 ciclos de entrenamiento. En el Gráfico 7.5.1 se muestra la evolución de los Verdaderos Positivos y los Falsos Positivos en función a la cantidad de ciclos de entrenamiento.

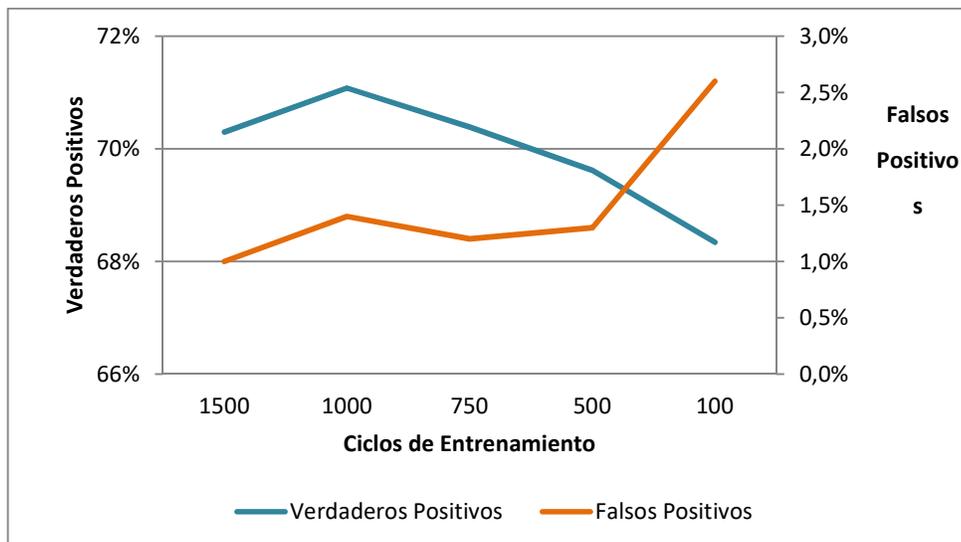


Figura 7.5.1 Proporción de Verdaderos Positivos y Falsos Positivos según cantidad de Ciclos de entrenamiento

Se puede observar que la variación entre los Falsos Positivos de 1500, 1000, 750 y 500 ciclos de entrenamiento es mínima y que la cantidad de Verdaderos Positivos tiene un máximo en los 1000ciclos.

Este mismo análisis se realiza para la elección de la cantidad de neuronas en las 2 capas ocultas del sistema. La cantidad de Verdaderos Positivos incrementa considerablemente al agregar más neuronas a las capas ocultas y,

como se puede apreciar en el Lift Chart del Gráfico 7.2.1, la cantidad gran mayoría de los valores positivos predichos, se realizan con niveles de confianza altos.

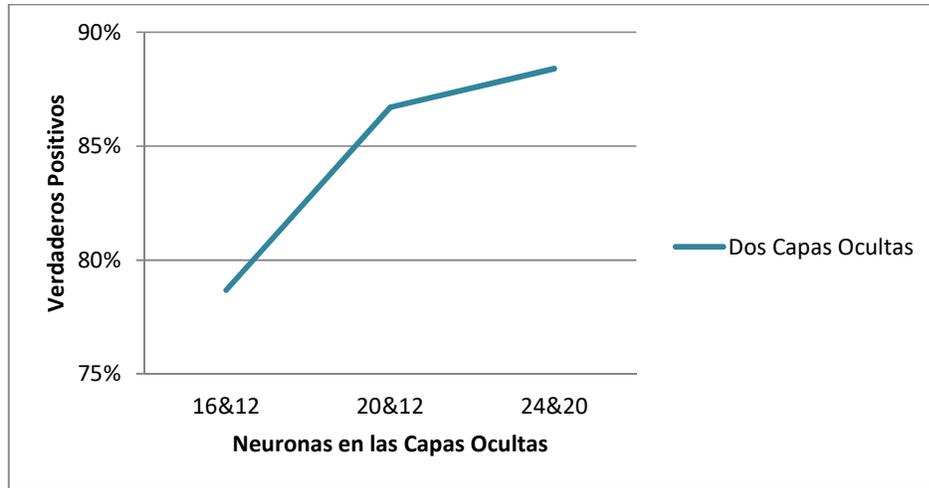


Figura 7.5.2 Proporción de Verdaderos Positivos en función a cantidad de neuronas en las capas ocultas

Finalmente, en función a la constante de aprendizaje utilizada, se observa que el incremento de Verdaderos Positivos, se obtiene a costas de un incremento de Falsos Positivos. A pesar de esto, el incremento del primero es más vertiginoso y por ende, más influyente en los resultados. Mientras que la diferencia entre Verdaderos Positivos de las constantes 0,5 y 0,2 es de 6%, la de Falsos Positivos es de 0,4%. Es por eso que se utiliza la constante 0,2 en el entrenamiento de la red. Esta relación se puede observar en el Gráfico 7.5.3.

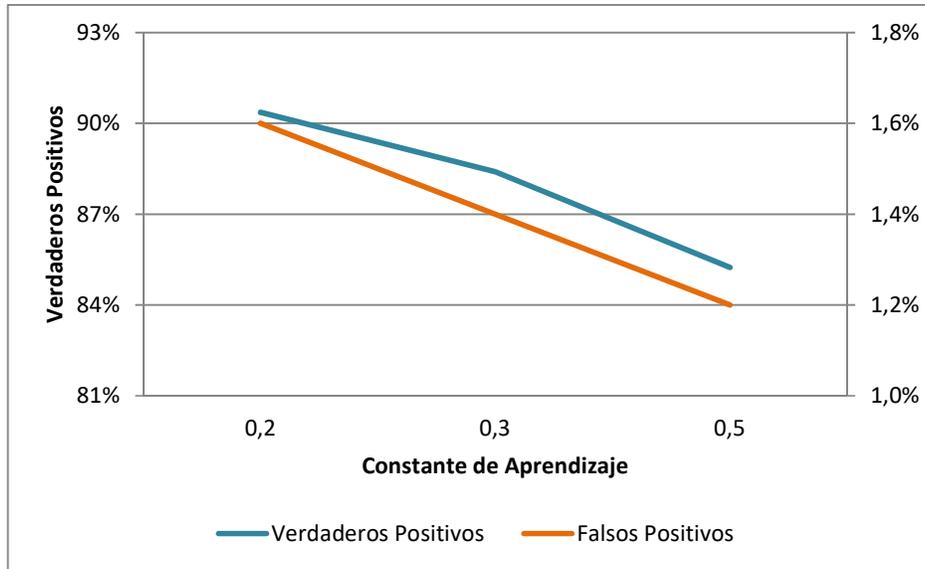


Figura 7.5.3 Proporción de Verdaderos Positivos en función a la Constante de Entrenamiento

8. Conclusión

Más allá de los resultados obtenidos, que indican la factibilidad de la aplicación y comercialización de las RNA, se concluye que esta tecnología tiene un gran poder de procesamiento si se la complementa de manera adecuada.

Durante el trabajo se descubre el efecto negativo de la existencia de información redundante en las variables de entrada. Es por esta razón que deben eliminarse aquellas que no aporten datos nuevos al estudio. Este proceso no sólo debe ser realizado manualmente, sino también con herramientas afines que ayuden a identificar relaciones poco claras entre variables.

Al mismo tiempo, se le debe dar importancia a la elección de la variable de salida. La posibilidad de contar con una salida binaria, no sólo simplifica el análisis posterior de los resultados obtenidos, sino también el entendimiento de los mismos.

Finalmente, se destaca la alta variabilidad de los resultados al modificar las variables que controlan la red. Se hace dificultosa la elección de una configuración óptima y encontrarla es necesariamente el resultado de pruebas empíricas. Al realizar los diferentes entrenamientos, hay que recordar que los resultados obtenidos están ligados a la fidelidad entre el comportamiento de la población estudiada y el de la muestra que conforma los datos de entrada. Representar a esta última fielmente, no necesariamente significa entender la esencia de la problemática planteada. *No se debe sobre-entrenar a la herramienta predictiva utilizada.*

La posibilidad de aplicar una RNA en un proyecto de esta envergadura, demuestra cómo herramientas que permiten el análisis de grandes volúmenes de datos están contribuyendo a la evolución del Marketing Online. La cantidad de información a procesar en la actualidad imposibilita el uso de herramientas tradicionales y obliga a aplicar procesos analíticos avanzados. Cualquier empresa que pretenda triunfar en este campo en crecimiento, debe estar lista para poder

adaptarse a las nuevas tecnologías y así lograr liderar en un mundo globalizado y dinámico.