

# Sistema de recomendación de inmuebles Bancolombia basado en simulación bayesiana y redes neuronales

**Juan Camilo Díaz Zapata**  
*Científico de Datos líder*

Gerencia de Modelos Analíticos  
Centro de Excelencia en Analítica  
Bancolombia

Meetup - Machine Learning Colombia

# Contenido

- 1 Introducción
- 2 Experiencia
- 3 Implementación
  - Preguntas tipo filtro
  - Preguntas tipo clasificación: gusto y afinidades
  - Preguntas tipo clasificación: composición familiar
  - Algoritmo Aplicado
- 4 Resultados
  - Resultado Simulación
  - Resultado Modelación
  - Puesta en producción
- 5 Trabajos en desarrollo
- 6 Referencias

## Motivación

Crear un sistema de recomendación de inmuebles para Bancolombia donde la experiencia del cliente y no cliente (usuario) al conseguir su próxima vivienda sea memorable, ya que te permita acompañar desde la etapa de búsqueda de la vivienda hasta la legalización del inmueble

# Motivación

Tu360 > Tu360Inmobiliario

## Tu360Inmobiliario Asesoría y compañía en el mundo de la vivienda

**Conoce las viviendas recomendadas para ti**

Según tu perfil y lo que conocemos de ti,  
podemos sugerirte algunas viviendas

[Ir a recomendados](#)



# Motivación

Se planteó crear dos sistemas de recomendación para abarcar todo el mercado posible:

# Motivación

Se planteó crear dos sistemas de recomendación para abarcar todo el mercado posible:

- ✓ Sistema de recomendación implícito.

# Motivación

Se planteó crear dos sistemas de recomendación para abarcar todo el mercado posible:

- ✓ Sistema de recomendación implícito.
- ✓ Sistema de recomendación explícito.

## Recomendador implícito

Es un sistema de recomendación pensado para los clientes bancolombia, donde es un sistema de recomendación contextual utilizando la mayor información transaccional histórica de cada cliente, este se basa en un sistema de scores de 4 capas:



## Recomendador implícito

Es un sistema de recomendación pensado para los clientes bancolombia, donde es un sistema de recomendación contextual utilizando la mayor información transaccional histórica de cada cliente, este se basa en un sistema de scores de 4 capas:

1. Capa geográfica.

## Recomendador implícito

Es un sistema de recomendación pensado para los clientes bancolombia, donde es un sistema de recomendación contextual utilizando la mayor información transaccional histórica de cada cliente, este se basa en un sistema de scores de 4 capas:

1. Capa geográfica.
2. Capa de similaridad.

## Recomendador implícito

Es un sistema de recomendación pensado para los clientes bancolombia, donde es un sistema de recomendación contextual utilizando la mayor información transaccional histórica de cada cliente, este se basa en un sistema de scores de 4 capas:

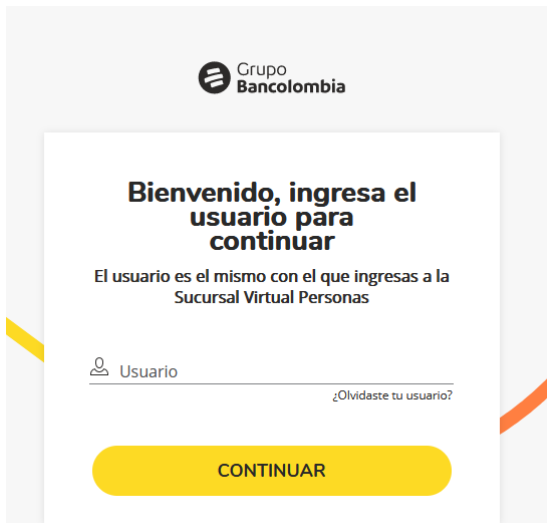
1. Capa geográfica.
2. Capa de similaridad.
3. Capa de afinidad de gustos.

## Recomendador implícito


Es un sistema de recomendación pensado para los clientes bancolombia, donde es un sistema de recomendación contextual utilizando la mayor información transaccional histórica de cada cliente, este se basa en un sistema de scores de 4 capas:

1. Capa geográfica.
2. Capa de similaridad.
3. Capa de afinidad de gustos.
4. Capa financiera.

## Recomendador implícito




The screenshot shows a login interface for 'Grupo Bancolombia'. At the top is the logo. The main heading is 'Bienvenido, ingresa el usuario para continuar'. Below it, a subtitle states 'El usuario es el mismo con el que ingresas a la Sucursal Virtual Personas'. There is a text input field with a person icon and the label 'Usuario'. To the right of the field is a link that says '¿Olvidaste tu usuario?'. At the bottom is a large yellow button labeled 'CONTINUAR'.

 Grupo  
Bancolombia

**Bienvenido, ingresa el  
usuario para  
continuar**

El usuario es el mismo con el que ingresas a la  
Sucursal Virtual Personas


 Usuario

[¿Olvidaste tu usuario?](#)

**CONTINUAR**

## Recomendador implícito

Te recomendamos estos inmuebles:




Nuevo

Dublín (MI Ciudad) Apto 90m2  
**Desde \$383.640.000**  
Sabaneta, Sabaneta, Antioquia  
Tipo D: 90 m<sup>2</sup> | 3 Hab | 3 Baños

Comparar Conocer más





Nuevo

Florencia Apartamento de 98.14  
**Desde \$617.428.500**  
Poblado, Medellín, Antioquia  
Tipo H: 98.14 m<sup>2</sup> | 3 Hab | 2 Bañ.

Comparar Conocer más



## Planteamiento del Problema

Bancolombia no cuenta con la información suficiente para recomendar viviendas de acuerdo a las a las características definidas como afinidades y composición familiar de los clientes y los no clientes

## Planteamiento del Problema

Bancolombia no cuenta con la información suficiente para recomendar viviendas de acuerdo a las características definidas como afinidades y composición familiar de los clientes y los no clientes





## Solución planteada

Se desarrolló un sistema de recomendación en tiempo real basado en cinco preguntas fundamentales hacia cualquier persona:

## Solución planteada

Se desarrolló un sistema de recomendación en tiempo real basado en cinco preguntas fundamentales hacia cualquier persona:

1. Zona geográfica de interés

## Solución planteada

Se desarrolló un sistema de recomendación en tiempo real basado en cinco preguntas fundamentales hacia cualquier persona:

1. Zona geográfica de interés
2. Rango de precio de búsqueda

## Solución planteada

Se desarrolló un sistema de recomendación en tiempo real basado en cinco preguntas fundamentales hacia cualquier persona:

1. Zona geográfica de interés
2. Rango de precio de búsqueda
3. Presupuesto para cuota inicial

## Solución planteada

Se desarrolló un sistema de recomendación en tiempo real basado en cinco preguntas fundamentales hacia cualquier persona:

1. Zona geográfica de interés
2. Rango de precio de búsqueda
3. Presupuesto para cuota inicial
4. Gustos y afinidades

## Solución planteada

Se desarrolló un sistema de recomendación en tiempo real basado en cinco preguntas fundamentales hacia cualquier persona:

1. Zona geográfica de interés
2. Rango de precio de búsqueda
3. Presupuesto para cuota inicial
4. Gustos y afinidades
5. la composición familiar del usuario

# Cuestionario

[< Atrás](#)**Bancolombia**

1

2

3

4

5

## Zonas de interés

Escoge las zonas donde te gustaría encontrar vivienda, máximo 3.

**Diámetro alrededor**

500m



1Km



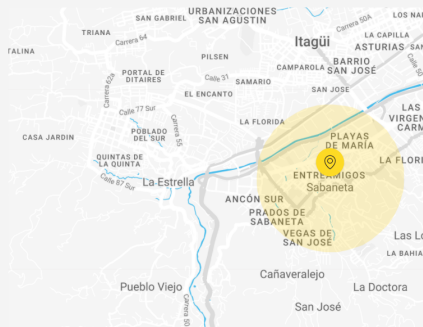
2Km



3Km



5Km



# Cuestionario

✓

2

3

4

5

## Rango de precios

Puedes ver los que estan disponibles en las zonas que seleccionaste.

📍

CI 71Sur #45a-25 Sabaneta

Valor mínimo:	Valor maximo:
\$273,412,704	\$706,466,811

### ¿Cuanto estás dispuesto a pagar?

Selecciona el rango que consideras invertir en tu próxima vivienda

\$273,412,704

\$475,111,441

Siguiente →

Figura 2: Pregunta 2: Rango de precios



# Cuestionario

✓

✓


3

4

5

## Ahorros para tu vivienda

¿Con cuánto dinero cuentas para la compra? Incluye primas, cesantías y/o ahorro grupal.



¿Tienes ahorros disponibles?

☒ Si tengo

☐ No tengo

Valor de tus ahorros\*

\$ 50,000,000

Consulta el saldo en tu fondo de ahorros y/o cesantías.

Si eres cliente [Protección](#) [ingresa aquí](#)

Siguiente →

Figura 3: Pregunta 3: Ahorros para tu vivienda

# Cuestionario

✓

✓

✓

4

5

## Lugares de interés

Selecciona los lugares que te gustaría tener cerca de tu próxima vivienda.

<input checked="" type="checkbox"/> <b>Salud</b>  <ul style="list-style-type: none"><li>Clinicas</li><li>Hospitales</li><li>Farmacias</li><li>Veterinarias</li></ul>	<input type="checkbox"/> <b>Consumo</b>  <ul style="list-style-type: none"><li>Mercados</li><li>Super mercados</li><li>Almacenes de cadena</li><li>Restaurantes</li></ul>	<input checked="" type="checkbox"/> <b>Transporte</b>  <ul style="list-style-type: none"><li>Estaciones de transporte</li><li>Estaciones de trenes</li><li>Talleres de Vehículos</li><li>Bombas de gasolina</li></ul>	<input checked="" type="checkbox"/> <b>Entretenimiento</b>  <ul style="list-style-type: none"><li>Centros comerciales</li><li>Malls</li><li>Teatros</li><li>Bares</li><li>Playas</li></ul>
<input type="checkbox"/> <b>Deporte</b>	<input checked="" type="checkbox"/> <b>Educación</b>	<input type="checkbox"/> <b>Seguridad</b>	<input type="checkbox"/> <b>Otros</b>

Figura 4: Pregunta 4: Lugares de interés

# Cuestionario

Figura 5: Pregunta 5: Cantidad de personas a vivir

# Resultado

8 inmuebles cumplen con tus criterios de selección:

Productos por página: 12 ▾

Ordenar por: Seleccionar ▾



NUEVO



Sakura Apartamentos 55.43m2  
Desde **\$273.412.704**

📍 Aves maria, Sabaneta, Antioq...  
Tipo A: 55.43m<sup>2</sup> | 3 Hab | 2 Baños

Comparar

Conocer más



NUEVO



Sakura Apartamentos 63.56m2  
Desde **\$348.937.603**

📍 Aves maria, Sabaneta, Antioq...  
Tipo B: 63.56m<sup>2</sup> | 3 Hab | 2 Baños

Comparar

Conocer más



NUEVO



Sakura Apartamentos 66.86m2  
Desde **\$323.166.181**

📍 Aves maria, Sabaneta, Antioq...  
Tipo C: 66.86m<sup>2</sup> | 3 Hab | 2 Baños

Comparar

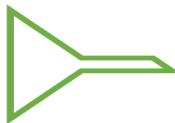
Conocer más



Figura 6: Resultado modelo de recomendación

# Preguntas tipo filtro

	clave	nombre_urbanizacion	area	price
1	3#40480396	Paseo Viscaya	88	290326000
2	1#324749778	Parma Apartamentos	111	899500000
3	2#81262589	Venecia Central	67	194900000
4	0.5#1296798118	Senior Suites Guaymaral	60	344177000
5	0.5#1298999773	Catalana	77	398691000
6	0.5#1296698808	Habitat Infinity	37	173858000
7	1#325030943	Biel	69	327120000
8	1#324208360	Fontana Park	67	253465000
9	2#812577733	Simplicity	56	304105964
10	5#20262423	Turó	102	614411260
11	0.5#1296794778	Alameda de San Diego	77	335000000



1. Zona de interés
2. Rango precio
3. Presupuesto cuota inicial

	clave	nombre_urbanizacion	area	price
1	2#81262589	WJ Towers	25	160320000
2	2#81262589	WJ Towers	30	203814000
3	2#81262589	WJ Towers	36	231232000
4	2#81262589	WJ Towers	40	271484000
5	2#81262589	Venecia Central	62	133000000
6	2#81262589	Venecia Central	67	194900000
7	2#81262589	Venecia Central	67	219900000
8	2#81262589	Edificio El Vigia	101	385820000
9	2#81262589	Edificio El Vigia	137	523340000
10	2#81262589	Edificio El Vigia	187	731170000

Figura 7: Filtrado de ofertas existentes

## Pregunta gusto y afinidades

Dentro de esta pregunta el usuario puede optar por seleccionar cada una de ocho posibles categorías, donde cada variable es dicótoma:

$$x_i = \begin{cases} 1 & \text{si 1-ésima característica es seleccionada} \\ 0 & \text{en otros casos} \end{cases}$$

$$i = 1, 2, \dots, 8$$

## Pregunta gusto y afinidades

Dentro de esta pregunta el usuario puede optar por seleccionar cada una de ocho posibles categorías, donde cada variable es dicótoma:

$$x_i = \begin{cases} 1 & \text{si 1-ésima característica es seleccionada} \\ 0 & \text{en otros casos} \end{cases}$$

$$i = 1, 2, \dots, 8$$

generando un vector binario del tipo:

$$X = [x_1, x_2, \dots, x_8]$$

## Pregunta gusto y afinidades

Dentro de esta pregunta el usuario puede optar por seleccionar cada una de ocho posibles categorías, donde cada variable es dicótoma:

$$x_i = \begin{cases} 1 & \text{si 1-ésima característica es seleccionada} \\ 0 & \text{en otros casos} \end{cases}$$

$$i = 1, 2, \dots, 8$$

generando un vector binario del tipo:

$$X = [x_1, x_2, \dots, x_8]$$



## Pregunta gusto y afinidades

adicional cada inmueble seleccionado tiene un vector de características previamente calculado que califica la influencia de cada categoría en cada inmueble:

$$A = [a_1, a_2, \dots, a_8]$$

donde  $0 \leq a_i \leq 1$

## Pregunta gusto y afinidades

adicional cada inmueble seleccionado tiene un vector de características previamente calculado que califica la influencia de cada categoría en cada inmueble:

$$A = [a_1, a_2, \dots, a_8]$$

donde  $0 \leq a_i \leq 1$

Al final cada inmueble tendrá un score de gusto que permitirá clasificar los inmuebles de acuerdo a los gustos y afinidades:

$$S_1 = \frac{A^\top \cdot X}{\sum_{i=1}^8 x_i} \quad (1)$$

## ¿Que es una elicitación?

Una elicitación es el proceso de hallar una distribución de probabilidad de una variable aleatoria donde dichas probabilidades son basadas en el conocimiento de uno o varios expertos acerca de un punto de interés [2].

## ¿Que es una elicitación?

Una elicitación es el proceso de hallar una distribución de probabilidad de una variable aleatoria donde dichas probabilidades son basadas en el conocimiento de uno o varios expertos acerca de un punto de interés [2].

### Estado del Arte

- ✓ Tangarife y Barrera [5] quienes estimaron los parámetros de un modelo logístico a partir de un elicitación de una distribución a priori beta y normal truncada.

# ¿Que es una elicitación?

Una elicitación es el proceso de hallar una distribución de probabilidad de una variable aleatoria donde dichas probabilidades son basadas en el conocimiento de uno o varios expertos acerca de un punto de interés [2].

## Estado del Arte

- ✓ Tangarife y Barrera [5] quienes estimaron los parámetros de un modelo logístico a partir de un elicitación de una distribución a priori beta y normal truncada.
- ✓ Correa y Barrera [1] estimaron los parámetros de un modelo lineal múltiple a partir de una elicitación bayesiana a varios expertos.

# ¿Que es una elicitación?

Una elicitación es el proceso de hallar una distribución de probabilidad de una variable aleatoria donde dichas probabilidades son basadas en el conocimiento de uno o varios expertos acerca de un punto de interés [2].

## Estado del Arte

- ✓ Tangarife y Barrera [5] quienes estimaron los parámetros de un modelo logístico a partir de un elicitación de una distribución a priori beta y normal truncada.
- ✓ Correa y Barrera [1] estimaron los parámetros de un modelo lineal múltiple a partir de una elicitación bayesiana a varios expertos.

## Planteamiento Solución

Dado que Bancolombia no cuenta con información que permita concluir de acuerdo a la composición familiar cuales son los inmuebles más acordes y dada la relevancia de esta pregunta en la elección de un inmueble [3], se aplicó una **elicitación** basada en el algoritmo de Tangarife y Barrera [5] para la obtención de una **base de datos simulada** de la cual se estimaron los parámetros de una **red neuronal** que permitió dar un score de clasificación para cada inmueble de las ofertas filtradas por las tres primeras preguntas.

## Paso 1

Se deben fijar los niveles de las covariables

$$X_1, X_2, \dots, X_k$$



# Paso 1

Se deben fijar los niveles de las covariables

$$X_1, X_2, \dots, X_k$$

donde:

$$X_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}]$$

## Paso 1

Se deben fijar los niveles de las covariables

$$X_1, X_2, \dots, X_k$$

donde:

$$X_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}]$$

donde:

$x_{i1}$  : Número de personas en habitar el inmueble.

$x_{i2}$  : Tamaño del inmueble (mts).

$x_{i3}$  : Número de habitaciones del inmueble.

$x_{i4}$  : Número de baños del inmueble.

## Paso 1

En la tabla 1 se describe los puntos de diseño seleccionados para las variables  $x_{i2}, x_{i3}, x_{i4}$ .

$x_{i2}$	$x_{i3}$	$x_{i4}$
46	1	1
63	2	2
70	3	2
85	3	2
100	3	2
120	4	3
180	4	3

Cuadro 1: Puntos de diseño para las variables  $x_{i2}, x_{i3}, x_{i4}$

## Paso 1

En la tabla 1 se describe los puntos de diseño seleccionados para las variables  $x_{i2}, x_{i3}, x_{i4}$ .

$x_{i2}$	$x_{i3}$	$x_{i4}$
46	1	1
63	2	2
70	3	2
85	3	2
100	3	2
120	4	3
180	4	3

Cuadro 1: Puntos de diseño para las variables  $x_{i2}, x_{i3}, x_{i4}$

Cada uno de los puntos descritos en la tabla 1 es combinado con los valores de la variable  $x_{i1} = 1, 2, 3, 4, 5$ ; obteniendo un total de 35 ( $k = 35$ ) puntos de diseño.

# Paso 1

La variable respuesta esta definida como:

$$y_j = \begin{cases} 1 & \text{si } j\text{-ésimo usuario con una composición familiar } x_{i1} \text{ adquiere el inmueble} \\ 0 & \text{en otros casos} \end{cases}$$

$$j = 1, 2, \dots, n$$

## Paso 2

Se fija un  $n$  y se pide al experto dar el número personas con alguna composición familiar  $x_{i1}$  que comprarían en una muestra hipotética de tamaño  $n$ , dígase  $Y_0$ , calcule  $E(\pi) = Y_0/n$ .

## Paso 2

Se fija un  $n$  y se pide al experto dar el número personas con alguna composición familiar  $x_{i1}$  que comprarían en una muestra hipotética de tamaño  $n$ , dígase  $Y_0$ , calcule  $E(\pi) = Y_0/n$ .

Para el mismo  $n$  se pide al experto dar el número mínimo de éxitos que él esperaría aceptable,  $Y_I$ , calcule  $\hat{\pi}_I = Y_I/n$ .

## Paso 2

Se fija un  $n$  y se pide al experto dar el número personas con alguna composición familiar  $x_{i1}$  que comprarían en una muestra hipotética de tamaño  $n$ , dígase  $Y_0$ , calcule  $E(\pi) = Y_0/n$ .

Para el mismo  $n$  se pide al experto dar el número mínimo de éxitos que él esperaría aceptable,  $Y_I$ , calcule  $\hat{\pi}_I = Y_I/n$ .

Para el mismo  $n$  se pide al experto dar el número máximo de éxitos que él esperaría aceptable,  $Y_S$ , calcule  $\hat{\pi}_S = Y_S/n$ .



## Paso 2

Se fija un  $n$  y se pide al experto dar el número personas con alguna composición familiar  $x_{i1}$  que comprarían en una muestra hipotética de tamaño  $n$ , dígase  $Y_0$ , calcule  $E(\pi) = Y_0/n$ .

Para el mismo  $n$  se pide al experto dar el número mínimo de éxitos que él esperaría aceptable,  $Y_I$ , calcule  $\hat{\pi}_I = Y_I/n$ .

Para el mismo  $n$  se pide al experto dar el número máximo de éxitos que él esperaría aceptable,  $Y_S$ , calcule  $\hat{\pi}_S = Y_S/n$ .

## Paso 2

Se fija un  $n$  y se pide al experto dar el número personas con alguna composición familiar  $x_{i1}$  que comprarían en una muestra hipotética de tamaño  $n$ , dígase  $Y_0$ , calcule  $E(\pi) = Y_0/n$ .

Para el mismo  $n$  se pide al experto dar el número mínimo de éxitos que él esperaría aceptable,  $Y_I$ , calcule  $\hat{\pi}_I = Y_I/n$ .

Para el mismo  $n$  se pide al experto dar el número máximo de éxitos que él esperaría aceptable,  $Y_S$ , calcule  $\hat{\pi}_S = Y_S/n$ .

Este paso se hizo con un tamaño  $n$  de 100 y una muestra de 16 expertos para cada punto de diseño.



## Paso 3

A los valores elicitados en el paso dos, se ajusta una distribución beta para estimar los parámetros  $\alpha$  y  $\beta$ , sean:

$$E(\pi) = y_0/n$$

$$P(\pi \leq Y_I/n) = 0.05$$

$$P(\pi \geq Y_S/n) = 0.05$$

## Paso 3

A los valores elicitados en el paso dos, se ajusta una distribución beta para estimar los parámetros  $\alpha$  y  $\beta$ , sean:

$$E(\pi) = y_0/n$$

$$P(\pi \leq Y_I/n) = 0.05$$

$$P(\pi \geq Y_S/n) = 0.05$$

los valores de  $\alpha$  y  $\beta$  se obtienen de minimizar la siguiente función:

$$f(\alpha, \beta) = (\pi_I - q\text{beta}(0.05, \alpha, \beta))^2 + (\pi_S - q\text{beta}(0.95, \alpha, \beta))^2 + (\pi + (\alpha/(\alpha + \beta)))^2$$

## Paso 4

Calcule el  $N$  equivalente, esto permite cuantificar el conocimiento del experto en términos de tamaño muestral, este tamaño representa realmente el nivel de conocimiento que los expertos tienen en la clasificación adecuada de inmuebles a una composición familiar determinada.

$$N = \frac{4Z_{\alpha/2}^2 \hat{\pi}(1 - \hat{\pi})}{(\hat{\pi}_S - \hat{\pi}_I)^2}$$

## Paso 5

Para cada nivel se repiten los siguientes pasos  $m$  veces:

- ♣ Genere un valor de la beta con  $\alpha_i$  y  $\beta_i$  hallados en el punto tres.

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_k \end{bmatrix}$$

En este paso se obtiene un vector de tamaño  $k$ , en nuestro caso  $k = 35$

## Paso 5

- ❖ Genere una muestra de valores  $y$  a partir de la distribución conjugada de la beta, la distribución binomial.

$$\begin{bmatrix} y_1^{(1)}, & X_1 \\ y_2^{(1)}, & X_1 \\ \vdots & \vdots \\ y_{n_{eq}}^{(1)}, & X_1 \\ \vdots & \vdots \\ y_1^{(k)}, & X_k \\ y_2^{(k)}, & X_k \\ \vdots & \vdots \\ y_{n_{eq}}^{(k)}, & X_k \end{bmatrix}$$



## Paso 5

- ❖ Genere una muestra de valores  $y$  a partir de la distribución conjugada de la beta, la distribución binomial.

$$\begin{bmatrix} y_1^{(1)}, & X_1 \\ y_2^{(1)}, & X_1 \\ \vdots & \vdots \\ y_{n_{eq}}^{(1)}, & X_1 \\ \vdots & \vdots \\ y_1^{(k)}, & X_k \\ y_2^{(k)}, & X_k \\ \vdots & \vdots \\ y_{n_{eq}}^{(k)}, & X_k \end{bmatrix}$$

En este paso se obtiene una matriz de tamaño  $(n_{eq} * k) \times 5$ , donde  $n_{eq}$  es el  $N$  equivalente hallado en el paso cuatro y  $X_k$  es una variable que representa los niveles definidos en el paso uno de tamaño cuatro.

## Resultado Simulación

El algoritmo fue construido en el sistema computacional R [4] y obtuvimos una base de datos con un total de 770 mil registros entre todas las combinaciones de puntos de diseño y composiciones familiares con su respectiva variable respuesta. [Respositorio GitHub](#)

$i$	$y_1$	$x_1$	$x_2$	$x_3$	$x_4$
1	1	1	46	1	1
2	1	1	46	1	1
3	0	2	46	1	1
4	0	4	63	2	2
5	1	2	63	2	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
768.000	1	3	70	3	2
769.000	0	1	85	3	2
770.000	1	5	180	3	2

Cuadro 2: Ejemplo de base de modelación resultante

## Resultado Modelación

A partir de esta base de datos generada, se procedió a estimar una red neuronal por medio del paquete **Keras** en el sistema computacional Python [6] con cuatro neuronas de entrada dadas por  $x_{i1}$ ,  $x_{i2}$ ,  $x_{i3}$  y  $x_{i4}$  y una capa oculta de seis neuronas con función de activación sigmoide, obteniendo un  $AUC = 80\%$  en base de validación y prueba.

```
def train(self, X_train, y_train):
    self.logger.Info("Training model")
    self.model = keras.Sequential([
        keras.layers.Dense(6, activation='sigmoid',
                            input_shape=(X_train.shape[-1],)),
        keras.layers.Dense(1, activation='sigmoid')])
    self.model.compile(optimizer=tf.keras.optimizers.Adam(),
                       loss=keras.losses.BinaryCrossentropy())
    self.model.fit(X_train, y_train, batch_size=self.batch_size,
                   epochs=self.n_epochs, verbose=2)
    self.logger.Info("Computing metrics")
    y_train_hat = self.model.predict(X_train)
    auc_score = roc_auc_score(y_train, y_train_hat)
    self.logger.Info("Train AUC {}".format(auc_score))

def test(self, X_test, y_test):
    self.check_model()
    y_test_hat = self.model.predict(X_test)
    auc_score = roc_auc_score(y_test, y_test_hat)
    self.logger.Info("Test AUC {}".format(auc_score))
```



## Puesta en producción

Al tener un modelo donde a partir de las características principales del inmueble (área, numero habitaciones, numero de baños) y la respuesta a la pregunta cinco (número integrantes a vivir en la inmueble) es posible dar un score o probabilidad de gusto sobre cada inmueble en oferta.

$$S_2 = f(x_{i1}, x_{i2}, x_{i3}, x_{i4}) \quad (2)$$

donde  $f(\cdot)$  es el modelo de red neuronal entrenado en 1.

## Puesta en producción

Al tener un modelo donde a partir de las características principales del inmueble (área, numero habitaciones, numero de baños) y la respuesta a la pregunta cinco (número integrantes a vivir en la inmueble) es posible dar un score o probabilidad de gusto sobre cada inmueble en oferta.

$$S_2 = f(x_{i1}, x_{i2}, x_{i3}, x_{i4}) \quad (2)$$

donde  $f(\cdot)$  es el modelo de red neuronal entrenado en 1.

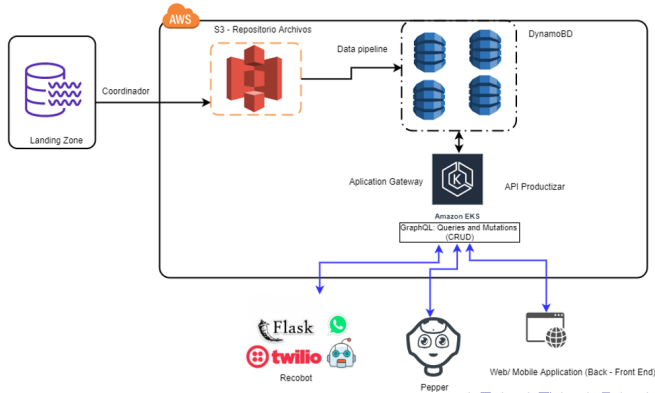
Dado las ecuaciones definidas en (1) y (2) podemos obtener un score final de clasificación que definirá el orden de recomendación de las ofertas que cumplen las tres primeras preguntas al cliente.

$$S_f = W_1 S_1 + W_2 S_2$$

donde  $0 \leq W_1 + W_2 \leq 1$

## Puesta en producción

La puesta en producción se realizó a través de una API alojada en **AWS** donde la información será actualizada con una frecuencia estable, a continuación, se muestran la arquitectura de software mostrando los componentes utilizados para el fácil consumo del modelo y de los datos de forma continua.



## Trabajos en desarrollo

1. Obtención automática de respuestas a preguntas y elección de favoritos en producción.
2. rentrenamiento continuo de la red neuronal con datos simulados y datos reales obtenidos en el punto anterior.
3. Explorar nuevas arquitecturas de la red neuronal y/o nuevos algoritmos que capturen de mejor forma el comportamiento de los datos reales y simulados.

## Referencias I



J. Correa and C. Barrera.

Elicitation of the parameters of multiple linear models.

*Revista Colombiana de Estadística - Applied Statistics*, 44(1):159–170, 2021.



B. D. Finerri.

La prevision: ses lois logiques, ses sources subjectives.

*Annal es de l'Institut Henri Poincard*, 7:1–68, 1937.



Planeatubien.com.

La dimensión ideal para tu departamento, 2021.



R Core Team.

*R: A Language and Environment for Statistical Computing*.

R Foundation for Statistical Computing, Vienna, Austria, 2021.



J. Tangarife and J. Correa.

Elicitación de una distribución a priori para el modelo logístico.

*Comunicaciones en Estadística*, 10(2):225–246, 2017.



## Referencias II



G. Van Rossum and F. L. Drake.  
*Python 3 Reference Manual*.  
CreateSpace, Scotts Valley, CA, 2009.

# Muchas Gracias...

*jucadiaz@bancolombia.com.co*