Sistema de recomendación de inmuebles Bancolombia basado en Simulación Bayesiana y redes neuronales

Resumen

Introducción

Un sistema de recomendación de inmuebles para Bancolombia hace que la experiencia del cliente al conseguir su próxima vivienda sea memorable, ya que con este nuevo sistema el banco te facilita y acompaña desde la etapa de búsqueda de la vivienda hasta la legalización del inmueble. Bancolombia día a día cuenta con millones de datos, sin embargo, no cuenta con la información suficiente para recomendar viviendas de acuerdo con las afinidades y composición familiar de los usuarios, por esto 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 y (5) la composición familiar del usuario. El sistema de recomendación inicialmente se basa en como integrar las respuestas de estas cinco preguntas a un conjunto de inmuebles que satisfaga de la mejor forma las respuestas de cada usuario.

Imagen de cómo utilizar la solución

La forma en cómo se manipulan las respuestas a estas cinco preguntas es clave para obtener una buena recomendación a los usuarios, por esto las tres primeras preguntas se tratan como un conjunto de filtros sobre la base total de inmuebles en la oferta existente actualmente, las dos últimas preguntas se tratarán como variables de clasificación que permitirán darle un orden a la base de ofertas resultante de las tres primeras preguntas, con el sentido que en primer lugar aparezca la oferta más recomendada de compra.

La composición familiar es una de las variables que tienen mayor preponderancia a la hora de la búsqueda de una vivienda según grandes portales inmobiliarios como lo describe tatata () y por supuesto es la variable de mayor influencia sobre el tipo de configuración del inmueble (tamaño del inmueble, número de habitaciones y de baños, etc…) con llevando a que el tratamiento de esta respuesta sea ideal para el éxito del sistema de recomendación propuesto. Sin embargo, dado la importancia de esta variable y la ausencia de información de parte del banco que contengan una relación entre el numero de personas a vivir en un inmueble y las características de un inmueble, se opto por simular una base de datos de modelado que permitiera el ajuste una red neuronal a partir de la elicitación de una distribución a priori basada en la distribución beta de acuerdo con el conocimiento experto de varias personas.

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 (De finnetti 1937). Para la extracción de este conocimiento se diseña un cuestionario de elicitación tomando consideraciones desde el campo.

Este proceso de la obtención de una base de datos simulada a partir de una elicitación, para la estimación ha sido estudiada y aplicada por distintos autores como Tangarife y correa 2017 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 2021 estimo los parámetros de un modelo lineal múltiple a partir de una elicitación bayesiana a varios expertos, el algoritmo propuesto en este articulo para la obtención de la base de datos simulada de la cual se estimaran los parámetros de la red neuronal se basó en la metodología desarrollada por Tangarife y correa 2017

A continuación, se dará la explicación del algoritmo aplicado para la obtención de una base simulada que dará pie a la estimación de los parámetros de la red neuronal.

1. Se deben fijar los niveles de las covariables de forma adecuada donde X\_1, X\_2, …X\_k son puntos de diseño que se eligen en consenso con el experto en este caso, avaluadores de viviendas y asesores comerciales de viviendas. Donde:

X\_i = Vector con x\_i1, x\_i2, x\_i3, xi\_4 y xi1:nro personas habitar inmueble 1, 2, … 5 xi2 : es el metraje del inmueble, xi3: numero habitaciones de la vivienda, xi4: número de baños de la vivienda

Estos puntos de diseño deben tener al menos el numero de parámetros que se quieren estimar dentro del modelo en este caso se tienen tres variables a estimar se escogieron siete puntos de diseño.

|  |  |  |  |
| --- | --- | --- | --- |
| Nro. diseño | Tamaño inmueble (mts) | Nro. habitaciones | Nro. Baños |
| 1 | 46 | 1 | 1 |
| 2 | 63 | 2 | 2 |
| 3 | 70 | 3 | 2 |
| 4 | 85 | 3 | 2 |
| 5 | 100 | 3 | 2 |
| 6 | 120 | 4 | 3 |
| 7 | 180 | 4 | 3 |

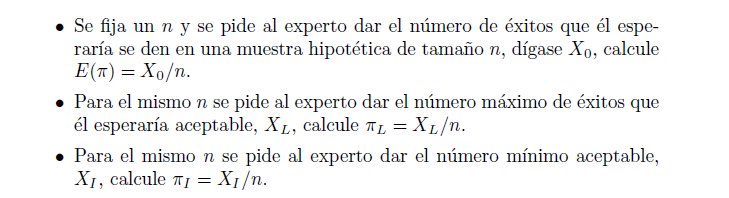
Cada uno de Estos diseños de inmuebles se combinan con los valores de xi1 obteniendo 35 pubntos de diseño

y deben corresponder y deben estar distribuidos en los percentiles 25, 50 y 75 de la probabilidad de éxito de la variable respuesta que este caso es:

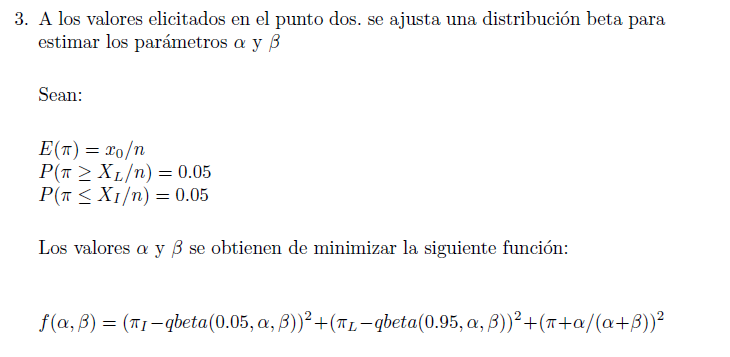
Y: 1 si el la persona P con una composición familiar r compra la casa:

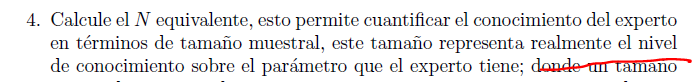
Donde r esta entre 1 y 5

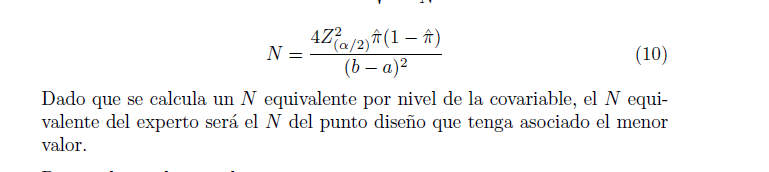
1. para cada punto de diseño y cada valor de r se procede a

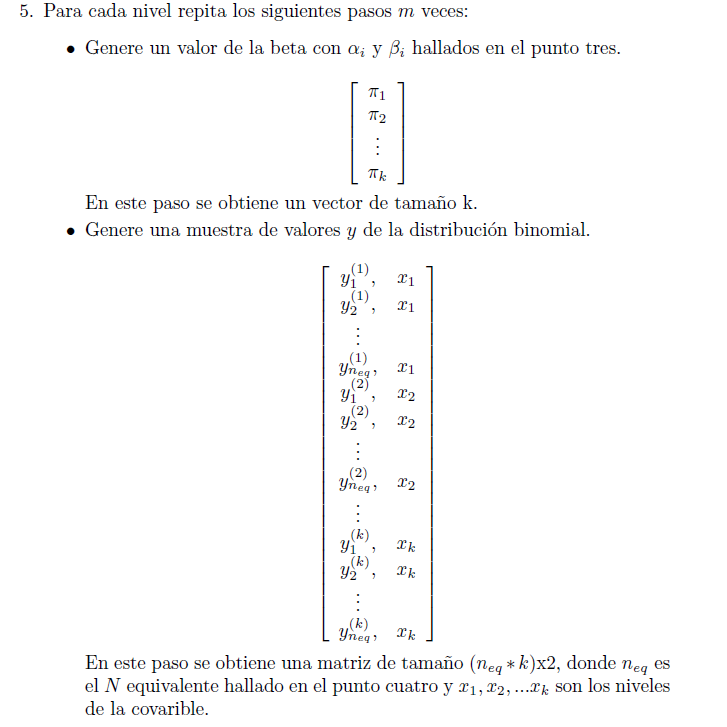


Este proceso se hizo con un n de tamaño = 100 y se hizo para una muestra de 16 expertos en promedio para cada punto del diseño y valor de r









En nuestro caso 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.

A partir de esta base datos generada se procedió a estimador una red neuronal con 4 neuronas de entrada (nro\_personas, tamaño inmueble, nro habitaciones, nro de baños) y una capa oculta de 6 neuronas con función de activación sigmo id, obteniendo un ajuste del 80% en base de test

Con este modelo puesto en producción le ha permitido a los clientes Bancolombia obtener una mejor experiencia a la hora de adquirir una vivienda nueva y al modelo ir obteniendo información real de los mismos clientes q le permitirá retroalimentar el modelo inicial de recomendación basado en una base de datos simulada.

* Estado del arte casos similares, elicitaciones, distribución beta
* Algoritmo propuesto basado en el articulo
* Resultados de la base de simulación
* Estimación y resultados de la red neuronal
* Conclusión