

Predicción de precios inmobiliarios:

Estado del arte.

Lo primero de todo, diferenciaría entre las viviendas de segunda mano y las recién construidas, ya que en las primeras el precio podría llegar a regatearse, mientras que en las segundas el precio suele ser completamente cerrado.

Una vez diferenciadas las viviendas en esos dos grupos, haría un análisis de mercado, preguntando en inmobiliarias e incluso a los vecinos de la zona (que suelen tener un criterio más cercano a la realidad. Todos somos marujas hasta cierto punto, y su valoración será sincera y con detalles que la inmobiliaria podría no dar).

Como punto final, miraría que factores afectan al precio de las mismas y los aunaría con los mencionados anteriormente:

- Ubicación
- Tamaño
- Materiales
- Orientación
- Instalaciones y servicios comunes (Cine comunitario? Piscina?)
- Viviendas denominadas objetos de deseo (viviendas que todo el mundo quiere por ciertas características, como ser un ático, ser una planta baja con jardín...)

Adquisición de datos.

Miraría datos referentes a “estadística registral sobre la inmobiliaria”. Este tipo de datos hacen referencia a temas como:

- Índice de Precio de la Vivienda de Ventas Repetidas (IPVVR).
- Precio medio registrado de locales comerciales y naves industriales.
- Número y distribución de compraventas de vivienda registradas.
- Superficie media de vivienda registrada
- Nacionalidad en las compras de vivienda.
- Periodo medio de posesión de la vivienda transmitida.
- Distribución del volumen de nuevo crédito hipotecario por tipo de entidad concesionaria, bien inmueble, grado de protección (libre vs. protegida) y superficie.
- Importe y duración de nuevo crédito hipotecario contratado por compraventa de vivienda.
- Tipos de Interés.
- Cuota hipotecaria mensual media y su relación con el Coste Salarial Medio.
- Impagos hipotecarios.

Problema / Solución.

El problema se basa en la predicción del precio de venta de uno o varios inmuebles.

Para poder conseguir algo como esto, procuraría crear un dataset con todos los valores de los datos recogidos gracias a la estadística registral mencionada antes.

En ese dataset pondría los datos en columnas, agrupados según viviendas nuevas o de segunda mano. Dentro de esos grupos, lo separaría según los factores mencionados en el estado del arte (ubicación, tamaño...).

La columna final serían los precios de cada uno de los inmuebles, de forma que cuando quiera entrenar mi modelo predictivo, pueda fácilmente separar todos los datos de cada vivienda, de su respectivo precio.

De este modo, cuando quiera entrenar la red, el dataset quedaría tal que así:

Dataset = [ubicación	tamaño (m2)	orientación...	precio (€)	
	Bilbao	80	Norte	150.000	
	Baracaldo	106	Sur	200.000	
]

Habría que tener en cuenta que todos los valores que están dentro del dataset, deben ser numéricos y no categóricos como el caso de la columna de “ubicación”, de forma que habría que transformar esa columna tal que así:

Dataset = [ubic. (Bilbao)	ubic.(Baracaldo)	tamaño (m2)	orientación...	
	1	0	80	Norte	
	0	1	106	Sur	
]

De este modo todos los valores categóricos pasan a ser numéricos, y la red puede ser entrenada correctamente. Con la columna “orientación” y todas las que tengan valores categóricos habría que hacer lo mismo.

Una vez organizado todo el dataset, entrenaría una red neuronal (el modelo predictivo a usar debería decidirse dependiendo del resultado esperado, en este caso, un valor en €).

Una vez entrenado, el modelo debería ser capaz de predecir valores a un porcentaje alto. Si no es así, habría que ir ajustando los hiperparámetros que componen el modelo predictivo (los hiperparámetros son parámetros que ajustan mejor o peor el resultado de predicción del modelo predictivo. Si no se consigue una buena predicción, quizá no se estén usando los valores necesarios para este caso).

Fiabilidad de posible cliente (banca):

Estado del arte.

Al igual que en la tarea anterior, el primer planteamiento que haría sería identificar que factores afectan directa e indirectamente en la fiabilidad de los clientes para una banca. Entre ellos se encuentran:

- Marketing
- Vinculación del cliente con la entidad financiera
- Retención del cliente con la entidad financiera
- Calidad percibida
- Satisfacción
- Expectativas
- Personalización hacia las necesidades del cliente
- ...

Adquisición de datos.

En este caso, los datos necesarios para poder crear un dataset en condiciones, se basarían en formularios hechos a los clientes (premiándolos con algo por el esfuerzo de hacer el cuestionario. Un poquito de motivación siempre hace que la gente se explaye más) y en datos registrados de sus transacciones (aquí es importante saber hasta que punto se pueden manejar ese tipo de datos, ya que son personales e intransferibles).

Problema / Solución.

Al igual que en la tarea anterior, se da la definición del problema a resolver: fiabilidad de los clientes de una banca.

Este tema es un tanto abstracto dado que también influyen las experiencias de vida de cada cliente, además de sus criterios de aceptación y su capital en el banco. Por esto se han planteado los formularios antes (con intención de entender la mentalidad de cada individuo).

Una vez entrenado el modelo elegido, probaría los resultados dentro de una banca. Hablaría con el responsable de fidelización de clientes, y le propondría usar mi código para ver si realmente se ajusta a la realidad o necesita mejorar.