

An ensemble learning based model for real estate project classification

Harvey Rodriguez Gil

Universidad Nacional de Colombia

Enero 27, 2022



Descripción del problema

Uno de los principales problemas de las personas cuando deciden adquirir un lugar para vivir (aunque en alguno casos es para invertir) es el ¿cuál será el monto de la hipoteca y de las cuotas mensuales? Y es que este tema aunque parece meramente de la persona que quiere cumplir sus sueños, no lo es del todo cierto, dado que los bancos influyen mucho en este punto. La razón: dependiendo de la vivienda escogida, los bancos dentro de sus protocolos o lineamientos establecidos le dará una calificación a ella, afectando los montos, las tasas de interés de un crédito para la compra de la misma. Por lo tanto se propone un modelo de clasificación de la propiedad, que le ayude a las personas a tomar una decisión en la etapa de seleccionar un préstamo.



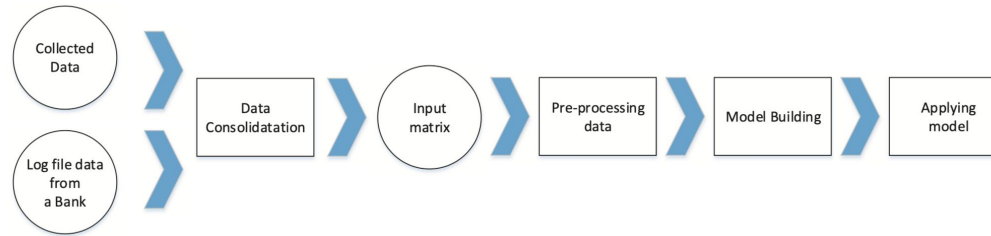
Variables de Interés

Las viviendas actualmente no solo cuentan con las características propias del apartamento casa que se desee comprar, si no que también al ser parte estas de una unidad o urbanización, son las amenidades (variables externas) las que influyen en esta clasificación. Las variables utilizadas en el modelo son las que se visualizan en la tabla. El proceso de selección de ellas ha sido mediante el uso de una red neuronal, y el peso dado a cada una de ellas a partir de un algoritmo genético.

Attribute	Coefficient
Swimming Pool (Salt)	1.000
Swimming Pool (Chlorine)	0.966
Fitness	0.888
Experience	0.784
Wireless	0.706
Sauna	0.570
Park	0.396
Unit	0.394
Listed in Stock Market	0.267
Security	0.170
CCTV	0.120



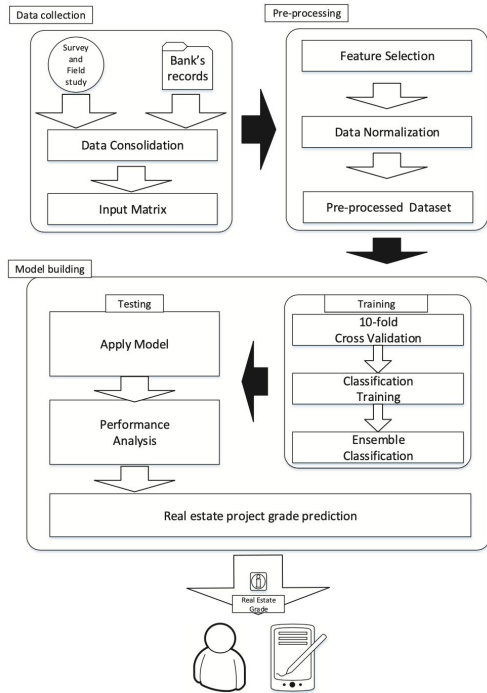
Problema de Entrenamiento



Los datos utilizados en el modelo, como se puede ver en la imagen vienen de dos fuentes diferentes que deben ser consolidados para una mejor respuesta del modelo. La variable de salida que nos entregará el modelo es una clasificación: AA, A, B, C¹; de la propiedad a partir de las variables de entrada indicadas en el anterior slide.



Implementación del modelo



Data Collection: Se cuenta con dos fuentes que deben ser consolidadas de tal forma que se cuente con información completa de las propiedades.

Pre-Processing: Se valida valores faltantes y se reemplazan. Se realiza un proceso de normalización sobre los datos. En este paso también se realiza el análisis de características para reducirlas a únicamente las más importantes.

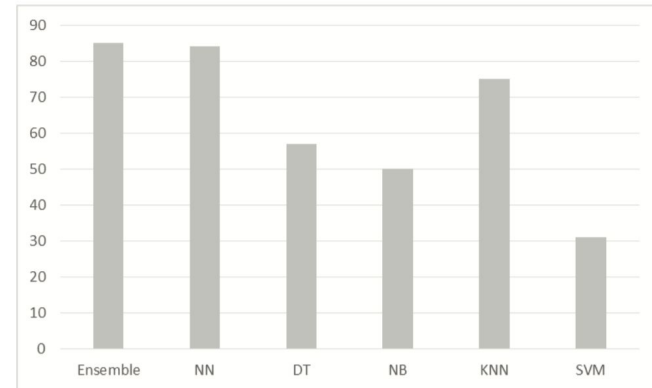
Training: Una vez completado el paso anterior, se separa el dataset en entrenamiento (307 registros) y pruebas (100 registros). El entrenamiento se realiza por medio de la técnica de validación cruzada, y se hace uso de diferentes modelos (DT, NB, NN y SVM)



Implementación del modelo

Testing: El resultado de los diferentes modelos aplicados en la etapa de entrenamiento es utilizado para mejorar la predicción en la etapa de testeo. Por medio de un algoritmo de votación se construye el modelo de clasificación final. En la tabla izquierda podemos ver la distribución de los datos de entrada. Y en la gráfica de la derecha el tasa de exactitud (accuracy) de cada uno de los diferentes modelos aplicados.

Grade	Instances
AA	107
A	100
B	100
C	100
Total	407



Lecciones aprendidas

- En diferentes oportunidades se ha intentado implementar modelos de clasificación de propiedades, sin embargo la desventaja que presentan frente a este modelo propuesto es que solo involucran un método de clasificación.
- El éxito de los modelos de clasificación de propiedades dependen mucho de la selección de características, por lo que futuros trabajos deben estar orientados en la mejor selección de estas. Esto ayuda a reducir el tiempo computacional en la construcción del modelo.



Referencias

- [1] What is Class A, Class B, or Class C property?. Recuperado de <https://www.realtymogul.com/knowledge-center/article/what-is-class-a-class-b-or-class-c-property>
- Worapat Paireekreng, Worawat Choensawat (2015) [An ensemble learning based model for real estate project classification](#)

