



FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

Matemáticas Aplicadas y Computación

# Proyecto Final

---

Integrantes:

Antonio Rojas Alexis David  
Galván Sandoval Diego Alexis  
Garzón Hernández José Emmanuel  
Valtierra Iparrea Daniel  
Vázquez Montoya Víctor Hugo.

# Minería de Datos

Grupo: 2801

# Índice

1. Resumen Ejecutivo	1
2. Planteamiento del problema	1
3. Objetivos	1
3.1. Objetivo General . . . . .	1
3.2. Objetivo Particular . . . . .	2
4. Marco Teórico	2
5. Análisis Exploratorio de Datos	2
6. Desarrollo	2
6.1. Estructura OLTP . . . . .	2
6.2. Estructura OLAP . . . . .	2
7. Modelo Supervisado	3
8. Modelo No Supervisado	3
9. Referencias	3

## 1. Resumen Ejecutivo

Existe una gran variedad de metodologías de predicción que han sido aplicadas exitosamente en los mercados de negocios tales como las acciones, los bonos, etc. El objetivo de este estudio es predecir el precio de los bienes raíces de una empresa líder en el mercado durante los últimos años en México, particularmente en la zona metropolitana, en la Ciudad de México, dicha empresa es conocida bajo su nombre comercial *Properaty*. Es un nuevo portal web de propiedades que busca mejorar la experiencia de compra, venta y alquiler de inmuebles en Argentina. Properati proviene del Latin “*Properatus*” que significa “moverse de una forma rápida” lo cual refleja la idea de un sitio ágil que le permita al usuario encontrar rápidamente su futura propiedad[9]. Para dicha predicción será necesario hacer previamente un análisis preeliminar de la información que se tenga presente; los datos correspondientes a dos conjuntos principales: el de rentas y ventas respectivamente. La importancia de hacer el presente estudio es para hallar modelos que puedan ser predictivos para llegar a la fase final del sistema de información, es decir, donde se tenga conocimiento y aquí mismo tomar una decisión importante para un esquema de negocios.

## 2. Planteamiento del problema

La problemática a manejar dentro del presente estudio surge de la necesidad de que no se conoce un buen modelo hasta el momento que pueda predecir precios.

## 3. Objetivos

### 3.1. Objetivo General

*Encontrar regiones en la zona metropolitana, las cuales sean de oportunidad para las constructoras.*

T_Property	
♦ Id	
◊ Created_On	
◊ Property_Type	
◊ Place_Name	
◊ State_Name	
◊ Latitude	
◊ Longitude	
◊ Price	
◊ Surface_Total_In_m2	
◊ Surface_Covered_In_m2	
◊ Price_USD_Per_m2	
◊ Floor	
◊ Rooms	
◊ Expenses	
◊ Property_URL	
◊ Description	
◊ Title	
◊ Image_Thumbnail	

Figura 1: OnLine Transactional Process

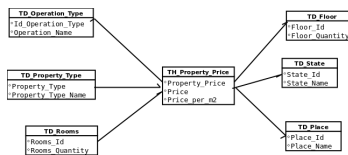


Figura 2: On Line Analytical Processing

### 3.2. Objetivo Particular

## 4. Marco Teórico

## 5. Análisis Exploratorio de Datos

## 6. Desarrollo

### 6.1. Estructura OLTP

A continuación se muestra la estructura OLTP

### 6.2. Estructura OLAP

Como se puede apreciar en la imagen de la estructura OLAP presentada a continuación...

## 7. Modelo Supervisado

## 8. Modelo No Supervisado

## 9. Referencias

### Referencias

- [1] Okmyung Bin. A prediction comparison of housing sales prices by parametric versus semi-parametric regressions. *Journal of Housing Economics*, 13(1):68–84, 2004.
- [2] Jieh-Haur Chen, Chuan Fan Ong, Linzi Zheng, and Shu-Chien Hsu. Forecasting spatial dynamics of the housing market using support vector machine. *International Journal of Strategic Property Management*, 21(3):273–283, 2017.
- [3] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [4] Robin A Dubin. Predicting house prices using multiple listings data. *The Journal of Real Estate Finance and Economics*, 17(1):35–59, 1998.
- [5] Jirong Gu, Mingcang Zhu, and Liuguangyan Jiang. Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications*, 38(4):3383–3386, 2011.
- [6] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489–501, 2006.
- [7] Byeonghwa Park and Jae Kwon Bae. Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data. *Expert Systems with Applications*, 42(6):2928–2934, 2015.
- [8] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [9] Properaty. <https://www.properati.com.mx/data>, 22 de Abril de 2018.
- [10] Christian Robert. Machine learning, a probabilistic perspective, 2014.
- [11] Jooyong Shim, Okmyung Bin, and Changha Hwang. Semiparametric spatial effects kernel minimum squared error model for predicting housing sales prices. *Neurocomputing*, 124:81–88, 2014.
- [12] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.