Presented at the FIG Working Week 2017,
May 29 - June 2, 2017 in Helsinki, Finland

House Price Estimation in Hanoi using Artificial Neural Network and Support Vector Machine: in Considering Effects of Status and House Quality

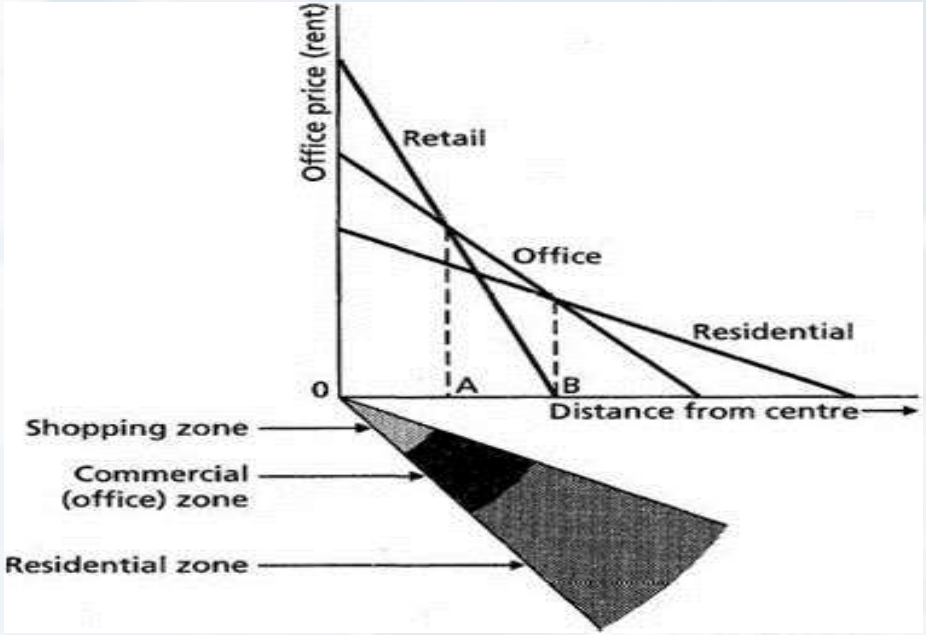
Quang-Thanh BUI, Nhu-Hiep DO, Huu-Phe HOANG

Organised by FIG

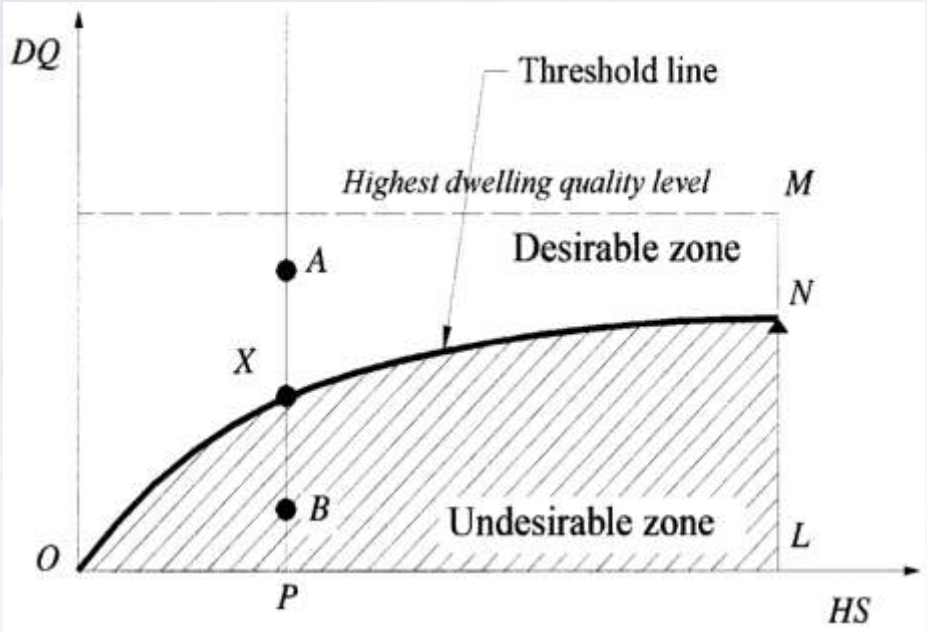
Platinum Sponsors: esri Trimble.

- In shaping the structure of housing market, several theories have been proposed, aiming at discussing the relationship between location and physical quality of houses as in (Alonso, 1964; Fujita, 1989; Kim, Pagliara, & Preston, 2005; Smith, 1987)

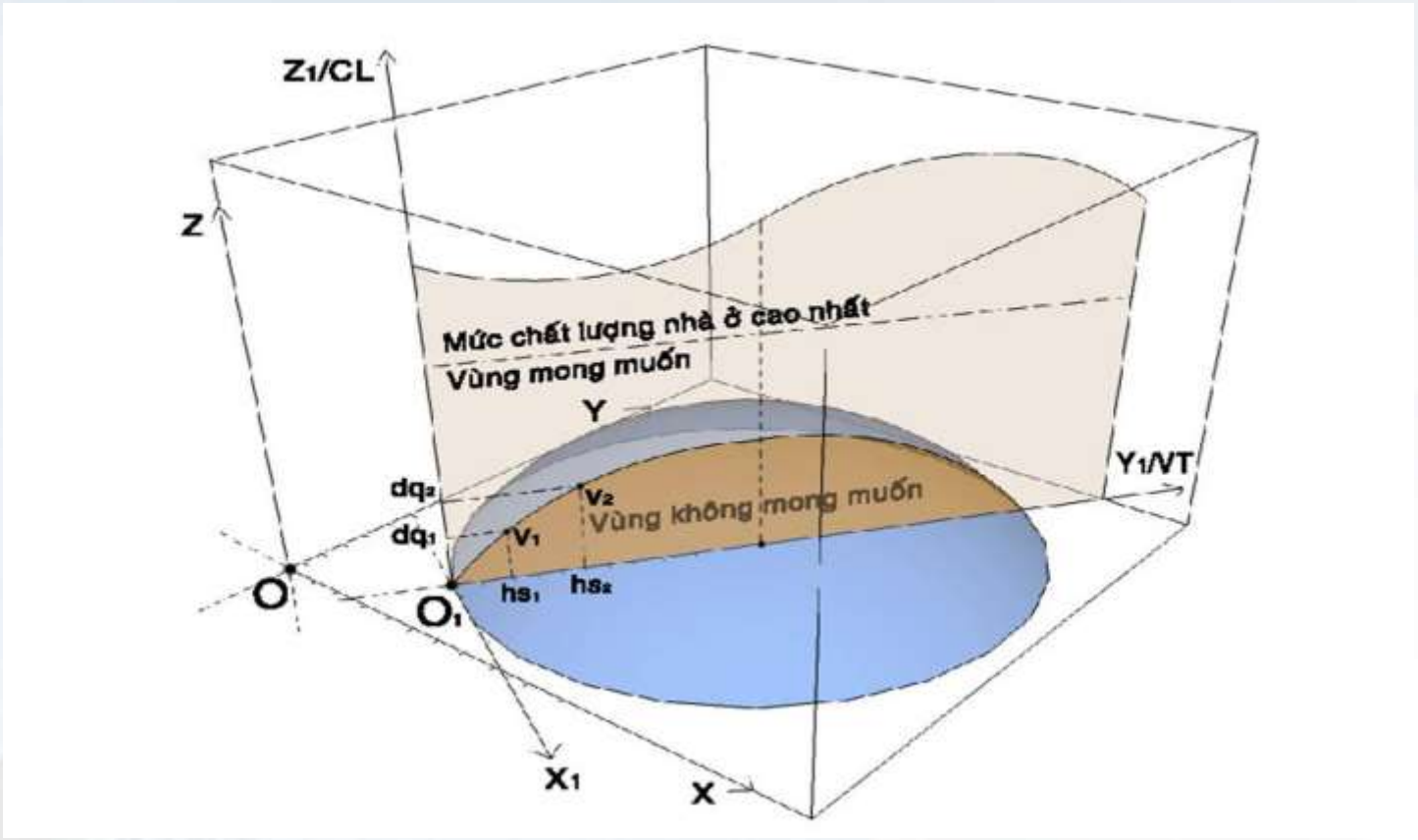
- (Huu Phe & Wakely, 2000) explained the dynamics of urban transformation and development, or in particular the dynamics of choice of residential houses (Status Quality Trade Off - SQTO) theory to house prices



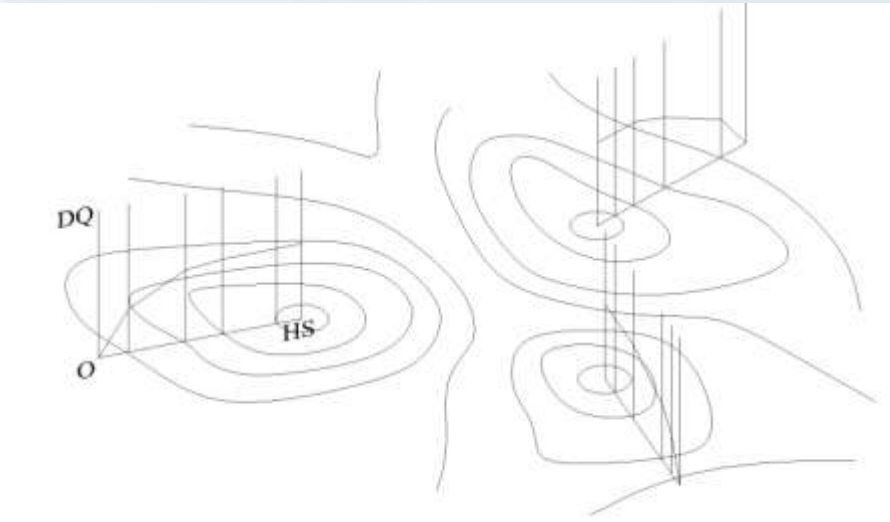
Trade-off between travel cost with house prices

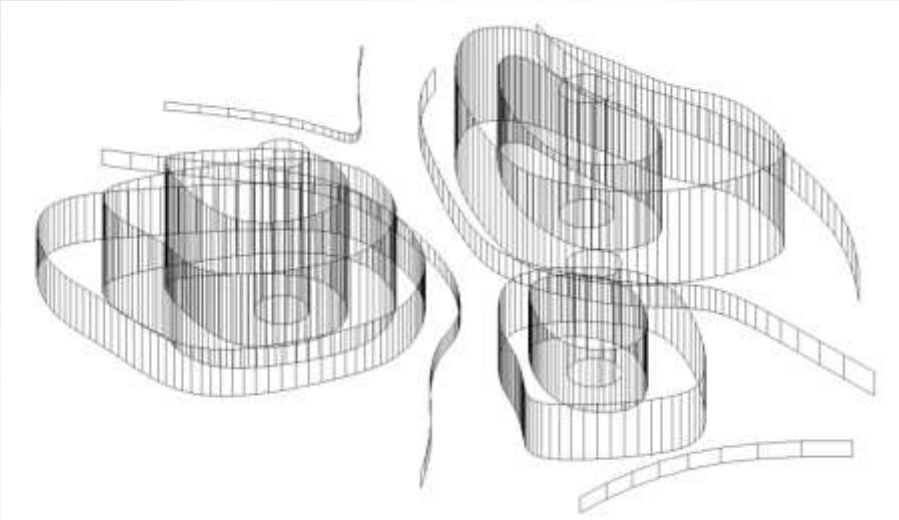


Trade-off with predetermined level of housing status


Z1/CL
Z
Mức chất lượng nhà ở cao nhất
Vùng mong muốn
Y
Y1/VT
dq2
dq1
V2
V1
Vùng không mong muốn
O
O1
hs1
hs2
X1
X

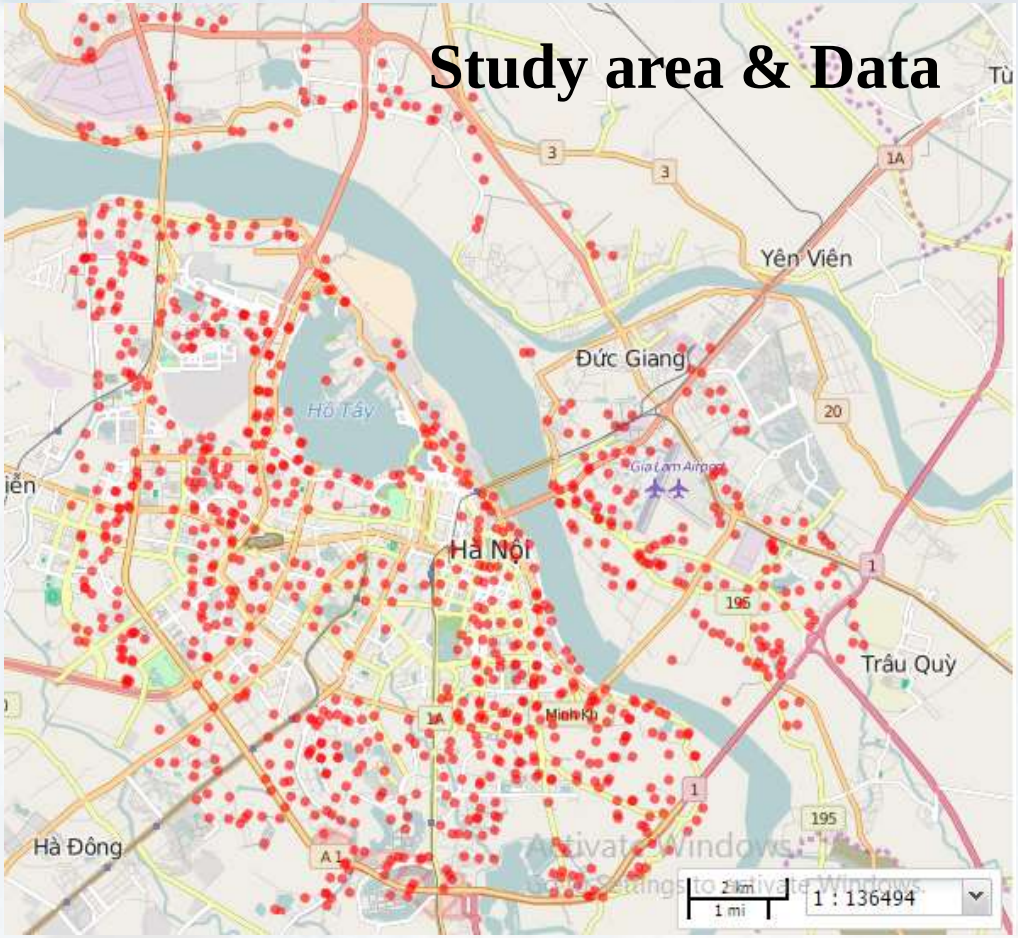
SQTO in 3D

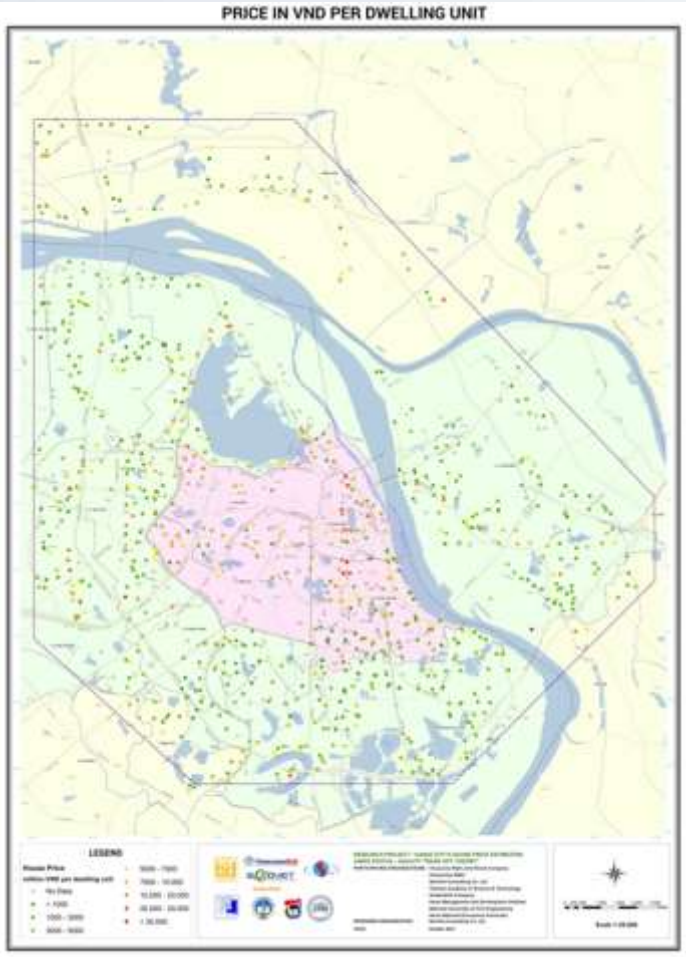




SQTO in 3D

Study area & Data





Activate Windows
Go to Settings to activate Windows.

Hedonic regression and geographically weighted regression are two common approaches in dealing with spatial data and with house prices in particular as in works of (Gollini, Lu, Charlton, Brunsdon, & Harris, 2015; Lu, Harris, Charlton, & Brunsdon, 2014) (Chen, Clapp, & Tirtiroglu, 2011; Dorsey, Hu, Mayer, & Wang, 2010; Liao & Wang, 2012; Selim, 2009)

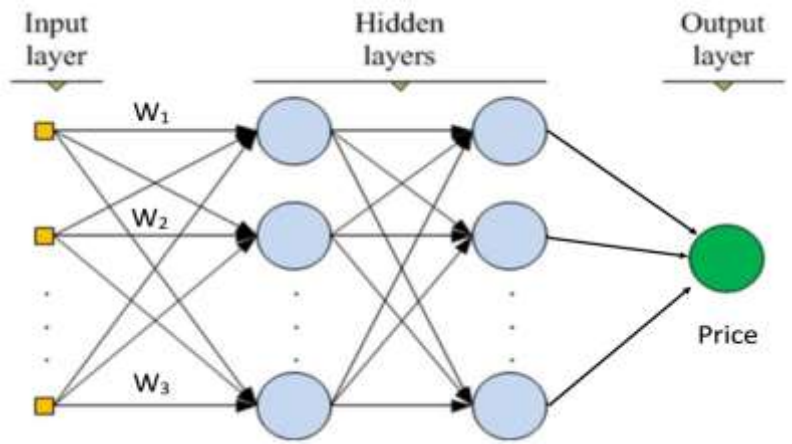
Recent researches have proved non-linear machine learning and the use of fuzzy logics (Kuşan, Aytekin, & Özdemir, 2010) can be a suitable solution to predict the variation of house prices. (Chiarazzo, Caggiani, Marinelli, & Ottomanelli, 2014; Steven P. Peterson, 2009; Vo et al., 2014; Wong, So, & Hung, 2002)

Dependent variable	Description	Data type
HPRICVND	Price of house in Millions of Vietnamese Dong	Ratio
Tangible Independent variables	Description	
AIRCON	Air-Conditioner (Yes, No)	Nominal
GFA	Total floor area (incl. mezzanine) (m^2)	Ratio
PLOTAREA	Total plot area (m^2)	Ratio
SHOPFRNT	Shop Front (Yes, No)	Nominal
PLUMBING	Plumbing Quality (Good, Other)	Nominal
HOUSEGRADE	Permanent, Other	Nominal
Intangible Independent variables	Description	
CAR	Car ownership (Yes, No)	Nominal
CENTDISR	Measured distance to Centre District	Ratio
DISCENDI	Perceived travel time to the Centre District	Ratio
EDYEARS	Time in education of the interviewee (years)	Ratio
OCCUP_PRIVBIZ	Private Business owner (1=Yes, 0=No)	Nominal
SCHOOQLT	School Quality (Good, Other)	Nominal

STRTYPE	Type of street (Business, Residential)	Nominal
BACTULIEM*	Located in Bac Tu Liem District	Nominal
BADINH*	Located in Ba Dinh District	Nominal
CAUGIAY*	Located in Cau Giay District	Nominal
DONGANH*	Located in Dong Anh District	Nominal
DONGDA*	Located in Dong Da District	Nominal
GIALAM*	Located in Gia Lam District	Nominal
HADONG*	Located in Ha Dong District	Nominal
HAIBATRUNG*	Located in Hai Ba Trung District	Nominal
HOANGMAI*	Located in Hoang Mai District	Nominal
HOANKIEM*	Located in Hoan Kiem District	Nominal
LONGBIEN*	Located in Long Bien District	Nominal
NAMTULIEM*	Located in Nam Tu Liem District	Nominal
SOCSON*	Located in Sóc Soḥ District	Nominal
TAYHO*	Located in Tay Ho District	Nominal
THANHTRI*	Located in Thanh Tri District	Nominal
THANHXUAN*	Located in Thanh Xuan District	Nominal

Methodology

Artificial Neural Network



Support Vector Machine

$$\textbf{Objective Function: } \min_{\mathbf{w},b,\xi_i} \left\{ \frac{\|\mathbf{w}\|^2}{2} + C\sum_{i=1}^{n} (\xi_i)^k \right\}$$

$$\textbf{Linear Constraints: } y_i\,(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall \mathbf{x}_i \in \mathbf{D}$$

$$\xi_i \geq 0 \;\forall \mathbf{x}_i \in \mathbf{D}$$

Performance assessment

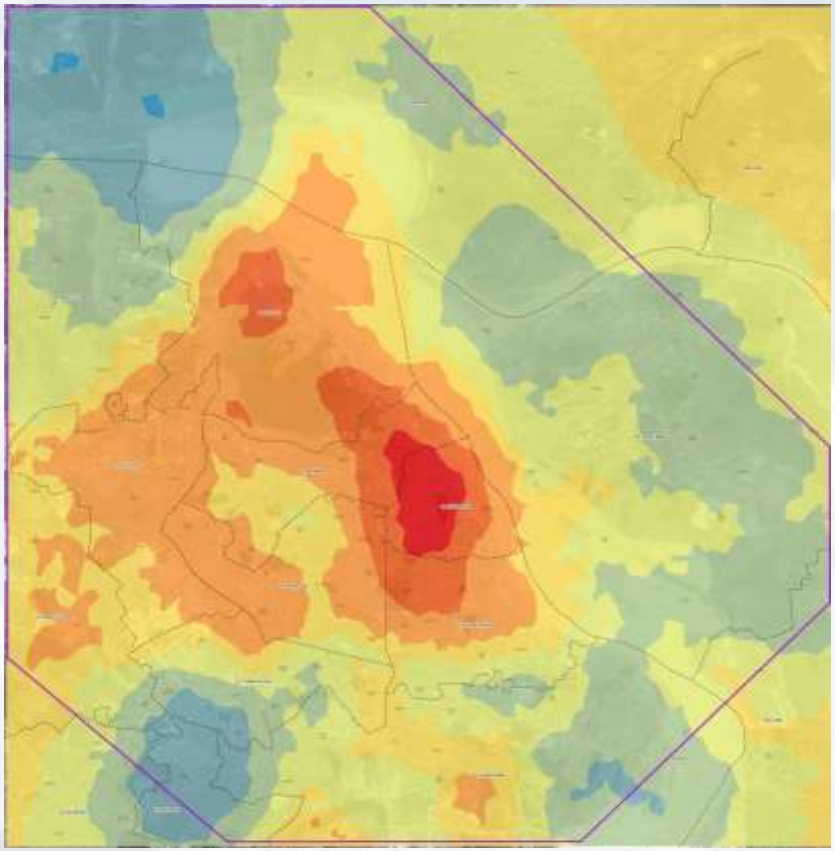
Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were often used to evaluate the performance of house price prediction models (Park & Bae, 2015; Wu, Gyourko, & Deng, 2012)

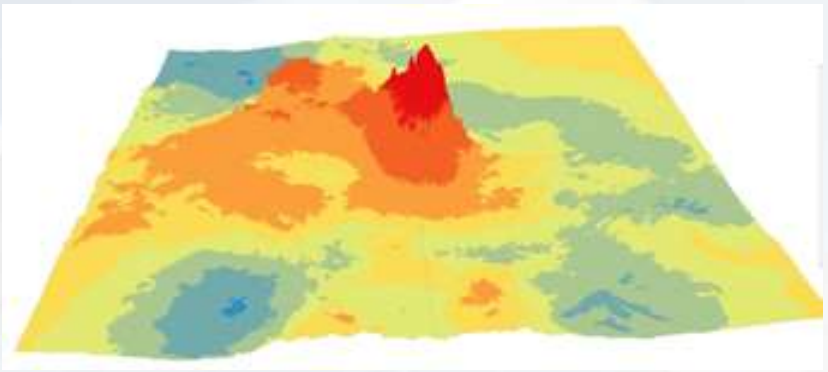
Results

Spatial distribution of house prices

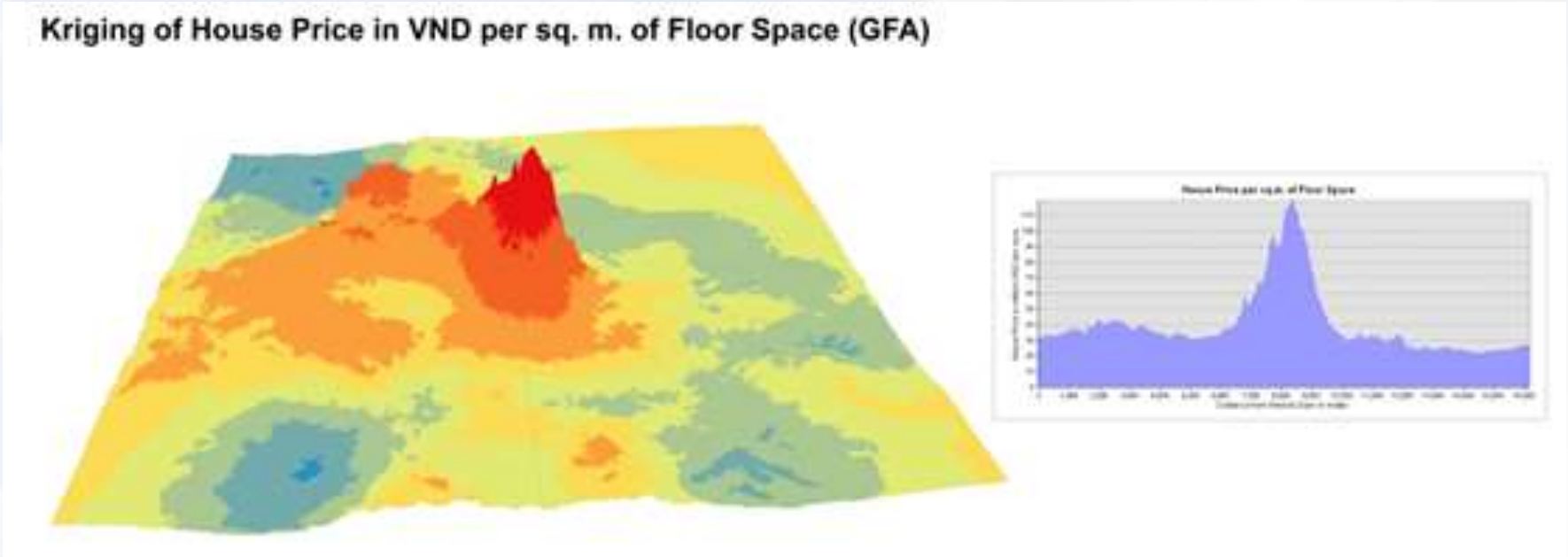
House price modelling

Spatial distribution of house prices in VND per SQ.M of floor space (GFA)

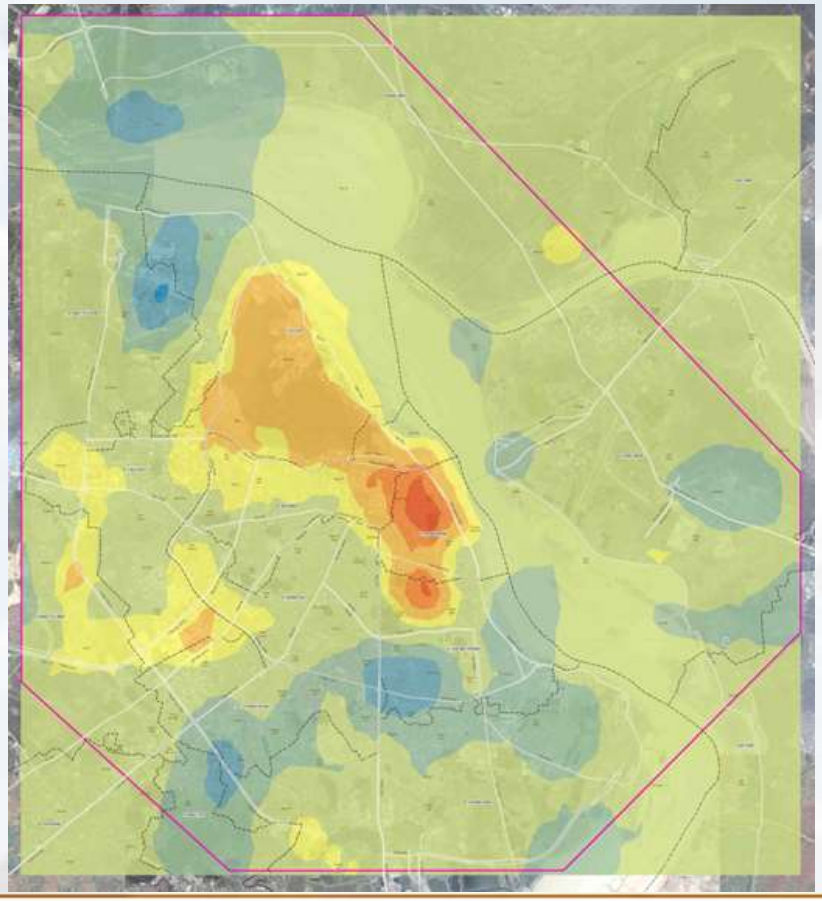




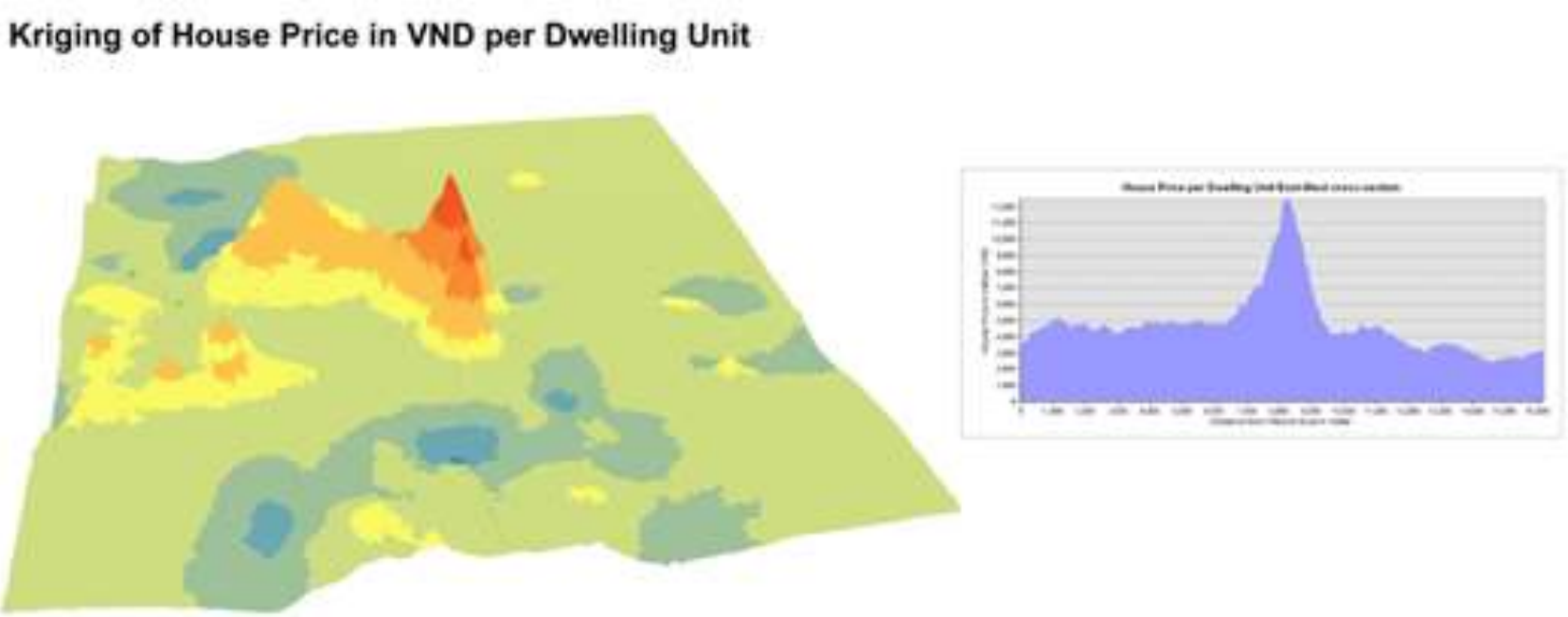
Spatial distribution of house prices in VND per SQ.M of floor space (GFA)



Spatial distribution of house prices in VND per dwelling unit

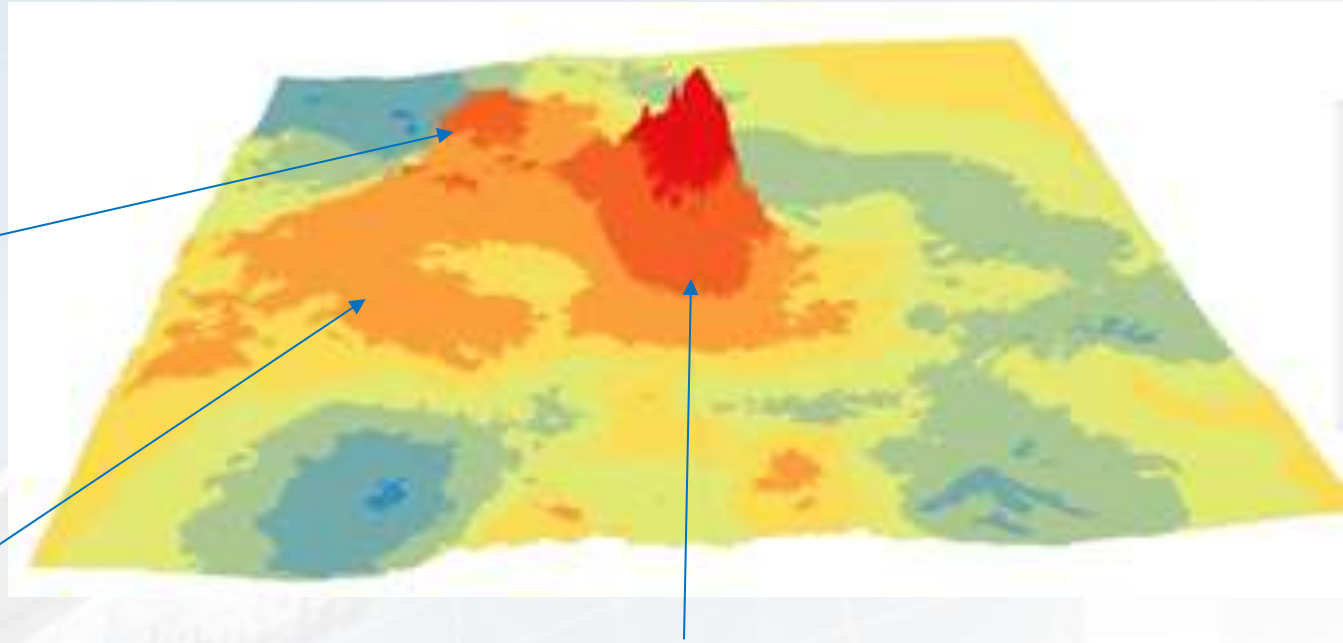


Spatial distribution of house prices in VND per dwelling unit



2) the West Lake area (Hồ Tây): The West Lake area (Hồ Tây) is emerging as the most important food, fashion and creative quarter of Hanoi City

3) the South West area (around Trung Hoa Nhan Chinh New Urban Area): The area was deliberately targeting the yuppies group



1) the Restored Sword Lake area (Hồ Gươm): The role of the Restored Sword area (Hồ Gươm) as the most important city centre, is a historical fact, widely recognised

Step 1. Division of standardized dataset (Training, Validation, Test)

Step 2. Selection of regresion techniques (Neural Network, SVM, Ensemble).

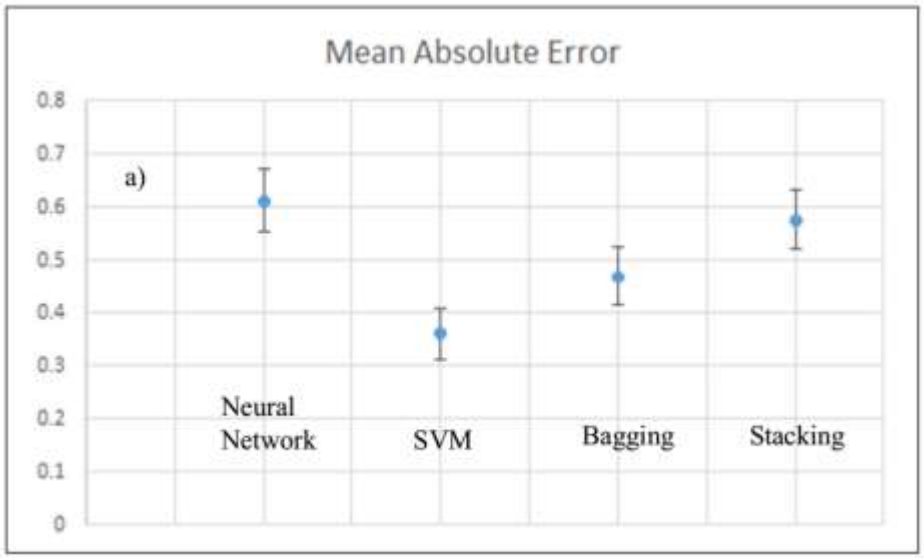
Step 3. Parameter tuning

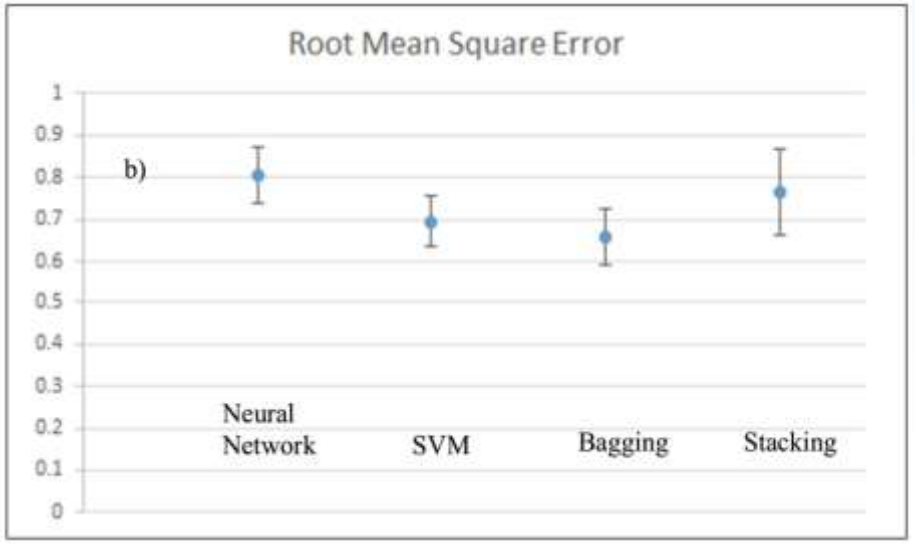
Step 4. Model training with 10-fold-validation

Step 5. Model evaluation by test set

Step 6. Measurement of errors from test set

Fold ID	Neural Network		SVM Regression		Bagging		Stacking	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.689968	0.895853	0.388417	0.764889	0.525214	0.753934	0.60042	0.904827
2	0.644308	0.849578	0.400296	0.705495	0.478905	0.667503	0.534253	0.656751
3	0.564123	0.802232	0.400423	0.722713	0.514395	0.693755	0.677428	0.739634
4	0.694555	0.909816	0.387871	0.764044	0.458534	0.6261	0.601731	0.875607
5	0.552275	0.736359	0.275716	0.630406	0.340802	0.531472	0.529669	0.759271
6	0.55723	0.795158	0.375069	0.605912	0.471455	0.646467	0.581468	0.634263
7	0.67179	0.843183	0.375215	0.739738	0.429268	0.616813	0.509787	0.612474
8	0.59837	0.767273	0.304135	0.634661	0.476881	0.647126	0.500272	0.798604
9	0.63935	0.74523	0.292092	0.623789	0.45418	0.622053	0.65062	0.75597
10	0.521899	0.69788	0.418721	0.752136	0.560184	0.778386	0.59012	0.907561
Average	0.6133868	0.8042562	0.3617955	0.6943783	0.4709818	0.6583609	0.5775768	0.7644962
Standard deviation	0.059501917	0.066307508	0.048531208	0.060508434	0.056572733	0.067327781	0.055970421	0.102989229


Mean Absolute Error
a)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Neural Network
SVM
Bagging
Stacking


Root Mean Square Error
b)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Neural Network
SVM
Bagging
Stacking

SQTO theory can fully be used to explain the structure of urban residents.

In addition to the systematic prediction for the presence of different poles in Hanoi, the estimation of hidden (implicit) value of tangible and intangible attributes based on data mining algorithm has proved to be very useful in identifying the driving forces of house price formation across the market segments.

The result also desbribes a strong movement of city structure from uni-polar to multiple-polar structure and it can be considered as preliminary foundation for a urban decision support system.

Thanks for your attention