



โครงการ

เรื่อง การทำนายราคาบ้าน

(Predicting House Prices using Machine Learning)

จัดทำโดย

กลุ่ม R008

- | | |
|------------------------|-----------------------|
| 1. สิริวิชญ์ ทิมสุวรรณ | เลขทะเบียน 639682067 |
| 2. ณัฏฐณิชา บุญเหลือ | เลขทะเบียน 6309682315 |
| 3. นาดา ทองคำ | เลขทะเบียน 6309682430 |

เสนอ

ผู้ช่วยศาสตราจารย์ ดร.วนิดา พุทธิวิทยา

รายวิชา คพ.245 หลักวิทยาการข้อมูล

2.3.1 บทนำ ที่มาของโครงการ

ในปัจจุบัน Big data และ Machine learning ได้เข้ามามีบทบาทในชีวิตประจำวันมากขึ้น คณะผู้จัดทำจึงได้เล็งเห็นที่จะนำความรู้ในเรื่อง Big Data และ Machine learning เข้ามาช่วยในการแก้ปัญหาในความสงสัยของเราว่า ทำไมราคาบ้านในแต่ละแห่งนั้นจึงมีราคาที่ไม่เท่ากัน อะไรคือปัจจัยหรือสาเหตุที่ทำให้ราคาบ้านแต่ละแห่งนั้นมีราคาที่แตกต่างกัน

คณะผู้จัดทำจึงได้ทำการศึกษา และวิเคราะห์ข้อมูลทางสถิติจากชุดข้อมูลที่ได้จากการสืบค้น เพื่อหาปัจจัยและสาเหตุที่เป็นตัวกำหนดราคาบ้านนั้น ๆ แล้วนำปัจจัยนั้น ๆ มาทำการสร้างโมเดลในการทำนายราคาบ้าน เพื่อให้เป็นประโยชน์แก่บุคคลที่สนใจที่จะซื้อหรือขายบ้านหลังหนึ่ง ๆ นั้น ที่มีองค์ประกอบของบ้านและสิ่งแวดล้อมที่แตกต่างกันในแต่ละหลังในราคาที่เหมาะสม

วัตถุประสงค์ของโครงการ

1. เพื่อให้นักศึกษาได้ทดลองประยุกต์ใช้ขั้นตอนวิธีการวิเคราะห์ข้อมูลที่ได้ศึกษาจากรายวิชา ในการทำงานกับชุดข้อมูลจริง

2. เพื่อให้นักศึกษาได้ฝึกการทำงานเป็นทีม

3. เพื่อศึกษาและประยุกต์การทำงานของ machine learning ในด้าน regression กับชุดข้อมูลจริง

ประโยชน์ของโครงการ

1. ได้ความรู้เพิ่มเติมจากการค้นคว้าหาข้อมูลที่นอกเหนือจากรายวิชา

2. ได้ฝึกฝนการทำงานเป็นทีม

3. ได้ปฏิบัติกับชุดข้อมูลจริง

4. ได้โมเดลการทำนายราคาบ้าน

2.3.2 ชุดข้อมูล (dataset)

อภิปรายถึง องค์ประกอบ รายละเอียดสิ่งแวดล้อมต่าง ๆ ของบ้านหลังหนึ่ง ๆ

เป็นชุดข้อมูลเรื่อง การทำนายราคาบ้าน (Predict sales house prices)

ที่มาของข้อมูล <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>

ความหมายฟิลด์ข้อมูล

Data fields

Here's a brief version of what you'll find in the data description file.

MSSubClass: Identifies the type of dwelling involved in the sale.

- 20 1-STORY 1946 & NEWER ALL STYLES
- 30 1-STORY 1945 & OLDER
- 40 1-STORY W/FINISHED ATTIC ALL AGES
- 45 1-1/2 STORY - UNFINISHED ALL AGES
- 50 1-1/2 STORY FINISHED ALL AGES
- 60 2-STORY 1946 & NEWER
- 70 2-STORY 1945 & OLDER
- 75 2-1/2 STORY ALL AGES
- 80 SPLIT OR MULTI-LEVEL
- 85 SPLIT FOYER
- 90 DUPLEX - ALL STYLES AND AGES
- 120 1-STORY PUD (Planned Unit Development) - 1946 & NEWER
- 150 1-1/2 STORY PUD - ALL AGES
- 160 2-STORY PUD - 1946 & NEWER
- 180 PUD - MULTILEVEL - INCL SPLIT LEV/FOYER
- 190 2 FAMILY CONVERSION - ALL STYLES AND AGES

MSZoning: Identifies the general zoning classification of the sale.

- A Agriculture
- C Commercial
- FV Floating Village Residential
- I Industrial
- RH Residential High Density
- RL Residential Low Density
- RP Residential Low Density Park
- RM Residential Medium Density

LotFrontage: Linear feet of street connected to property

LotArea: Lot size in square feet

Street: Type of road access to property

- Grvl Gravel
- Pave Paved

Alley: Type of alley access to property

Grvl Gravel

Pave Paved

NA No alley access

LotShape: General shape of property

Reg Regular

IR1 Slightly irregular

IR2 Moderately Irregular

IR3 Irregular

LandContour: Flatness of the property

Lvl Near Flat/Level

Bnk Banked - Quick and significant rise from street grade to building

HLS Hillside - Significant slope from side to side

Low Depression

Utilities: Type of utilities available

AllPub All public Utilities (E,G,W,& S)

NoSewr Electricity, Gas, and Water (Septic Tank)

NoSeWa Electricity and Gas Only

ELO Electricity only

LotConfig: Lot configuration

Inside Inside lot

Corner Corner lot

CulDSac Cul-de-sac

FR2 Frontage on 2 sides of property

FR3 Frontage on 3 sides of property

LandSlope: Slope of property

Gtl Gentle slope

Mod Moderate Slope

Sev Severe Slope

Neighborhood: Physical locations within Ames city limits

Blmngtn	Bloomington Heights
Blueste	Bluestem
BrDale	Briardale
BrkSide	Brookside
ClearCr	Clear Creek
CollgCr	College Creek
Crawfor	Crawford
Edwards	Edwards
Gilbert	Gilbert
IDOTRR	Iowa DOT and Rail Road
MeadowV	Meadow Village
Mitchel	Mitchell
Names	North Ames
NoRidge	Northridge
NPkVill	Northpark Villa
NridgHt	Northridge Heights
NWAmes	Northwest Ames
OldTown	Old Town
SWISU	South & West of Iowa State University
Sawyer	Sawyer
SawyerW	Sawyer West
Somerst	Somerset
StoneBr	Stone Brook
Timber	Timberland
Veenker	Veenker

Condition1: Proximity to various conditions

Artery	Adjacent to arterial street
Feedr	Adjacent to feeder street
Norm	Normal
RRNn	Within 200' of North-South Railroad
RRAn	Adjacent to North-South Railroad

PosN	Near positive off-site feature--park, greenbelt, etc.
PosA	Adjacent to positive off-site feature
RRNe	Within 200' of East-West Railroad
RRAe	Adjacent to East-West Railroad

Condition2: Proximity to various conditions (if more than one is present)

Artery	Adjacent to arterial street
Feedr	Adjacent to feeder street
Norm	Normal
RRNn	Within 200' of North-South Railroad
RRAn	Adjacent to North-South Railroad
PosN	Near positive off-site feature--park, greenbelt, etc.
PosA	Adjacent to positive off-site feature
RRNe	Within 200' of East-West Railroad
RRAe	Adjacent to East-West Railroad

BldgType: Type of dwelling

1Fam	Single-family Detached
2FmCon	Two-family Conversion; originally built as one-family dwelling
Duplx	Duplex
TwnhSE	Townhouse End Unit
TwnhSI	Townhouse Inside Unit

HouseStyle: Style of dwelling

1Story	One story
1.5Fin	One and one-half story: 2nd level finished
1.5Unf	One and one-half story: 2nd level unfinished
2Story	Two story
2.5Fin	Two and one-half story: 2nd level finished
2.5Unf	Two and one-half story: 2nd level unfinished
SFoyer	Split Foyer
SLvlSplit Level	

OverallQual: Rates the overall material and finish of the house

10 Very Excellent

- 9 Excellent
- 8 Very Good
- 7 Good
- 6 Above Average
- 5 Average
- 4 Below Average
- 3 Fair
- 2 Poor
- 1 Very Poor

OverallCond: Rates the overall condition of the house

- 10 Very Excellent
- 9 Excellent
- 8 Very Good
- 7 Good
- 6 Above Average
- 5 Average
- 4 Below Average
- 3 Fair
- 2 Poor
- 1 Very Poor

YearBuilt: Original construction date

YearRemodAdd: Remodel date (same as construction date if no remodeling or additions)

RoofStyle: Type of roof

- Flat Flat
- Gable Gable
- Gambrel Gabrel (Barn)
- Hip Hip
- Mansard Mansard
- Shed Shed

RoofMatl: Roof material

- ClyTile Clay or Tile

CompShg	Standard (Composite) Shingle
Membran	Membrane
Metal	Metal
Roll Roll	
Tar&Grv	Gravel & Tar
WdShake	Wood Shakes
WdShngl	Wood Shingles

Exterior1st: Exterior covering on house

AsbShng	Asbestos Shingles
AsphShn	Asphalt Shingles
BrkComm	Brick Common
BrkFace	Brick Face
CBlock	Cinder Block
CemntBd	Cement Board
HdBoard	Hard Board
ImStucc	Imitation Stucco
MetalSd	Metal Siding
Other	Other
Plywood	Plywood
PreCast	PreCast
Stone	Stone
Stucco	Stucco
VinylSd	Vinyl Siding
Wd Sdng	Wood Siding
WdShing	Wood Shingles

Exterior2nd: Exterior covering on house (if more than one material)

AsbShng	Asbestos Shingles
AsphShn	Asphalt Shingles
BrkComm	Brick Common
BrkFace	Brick Face
CBlock	Cinder Block

CemntBd	Cement Board
HdBoard	Hard Board
ImStucc	Imitation Stucco
MetalSd	Metal Siding
Other	Other
Plywood	Plywood
PreCast	PreCast
Stone	Stone
Stucco	Stucco
VinylSd	Vinyl Siding
Wd Sdng	Wood Siding
WdShing	Wood Shingles

MasVnrType: Masonry veneer type

BrkCmn	Brick Common
BrkFace	Brick Face
CBlock	Cinder Block
None	None
Stone	Stone

MasVnrArea: Masonry veneer area in square feet

ExterQual: Evaluates the quality of the material on the exterior

Ex	Excellent
Gd	Good
TA	Average/Typical
Fa	Fair
Po	Poor

ExterCond: Evaluates the present condition of the material on the exterior

Ex	Excellent
Gd	Good
TA	Average/Typical
Fa	Fair
Po	Poor

Foundation: Type of foundation

BrkTil	Brick & Tile
CBlock	Cinder Block
PConc	Poured Contrete
Slab	Slab
Stone	Stone
Wood	Wood

BsmtQual: Evaluates the height of the basement

Ex	Excellent (100+ inches)
Gd	Good (90-99 inches)
TA	Typical (80-89 inches)
Fa	Fair (70-79 inches)
Po	Poor (<70 inches)
NA	No Basement

BsmtCond: Evaluates the general condition of the basement

Ex	Excellent
Gd	Good
TA	Typical - slight dampness allowed
Fa	Fair - dampness or some cracking or settling
Po	Poor - Severe cracking, settling, or wetness
NA	No Basement

BsmtExposure: Refers to walkout or garden level walls

Gd	Good Exposure
Av	Average Exposure (split levels or foyers typically score average or above)
Mn	Mimimum Exposure
No	No Exposure
NA	No Basement

BsmtFinType1: Rating of basement finished area

GLQ	Good Living Quarters
ALQ	Average Living Quarters

BLQ Below Average Living Quarters

Rec Average Rec Room

LwQ Low Quality

Unf Unfinished

NA No Basement

BsmtFinSF1: Type 1 finished square feet

BsmtFinType2: Rating of basement finished area (if multiple types)

GLQ Good Living Quarters

ALQ Average Living Quarters

BLQ Below Average Living Quarters

Rec Average Rec Room

LwQ Low Quality

Unf Unfinished

NA No Basement

BsmtFinSF2: Type 2 finished square feet

BsmtUnfSF: Unfinished square feet of basement area

TotalBsmtSF: Total square feet of basement area

Heating: Type of heating

Floor Floor Furnace

GasA Gas forced warm air furnace

GasW Gas hot water or steam heat

Grav Gravity furnace

OthW Hot water or steam heat other than gas

WallWall furnace

HeatingQC: Heating quality and condition

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

CentralAir: Central air conditioning

N No

Y Yes

Electrical: Electrical system

SBkr Standard Circuit Breakers & Romex

FuseA Fuse Box over 60 AMP and all Romex wiring (Average)

FuseF 60 AMP Fuse Box and mostly Romex wiring (Fair)

FuseP 60 AMP Fuse Box and mostly knob & tube wiring (poor)

Mix Mixed

1stFlrSF: First Floor square feet

2ndFlrSF: Second floor square feet

LowQualFinSF: Low quality finished square feet (all floors)

GrLivArea: Above grade (ground) living area square feet

BsmtFullBath: Basement full bathrooms

BsmtHalfBath: Basement half bathrooms

FullBath: Full bathrooms above grade

HalfBath: Half baths above grade

Bedroom: Bedrooms above grade (does NOT include basement bedrooms)

Kitchen: Kitchens above grade

KitchenQual: Kitchen quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

Functional: Home functionality (Assume typical unless deductions are warranted)

Typ Typical Functionality

Min1 Minor Deductions 1

Min2 Minor Deductions 2

Mod Moderate Deductions

Maj1 Major Deductions 1

Maj2 Major Deductions 2

Sev Severely Damaged

Sal Salvage only

Fireplaces: Number of fireplaces

FireplaceQu: Fireplace quality

Ex Excellent - Exceptional Masonry Fireplace

Gd Good - Masonry Fireplace in main level

TA Average - Prefabricated Fireplace in main living area or Masonry Fireplace in basement

Fa Fair - Prefabricated Fireplace in basement

Po Poor - Ben Franklin Stove

NA No Fireplace

GarageType: Garage location

2Types More than one type of garage

Attchd Attached to home

Basment Basement Garage

BuiltIn Built-In (Garage part of house - typically has room above garage)

CarPort Car Port

Detchd Detached from home

NA No Garage

GarageYrBlt: Year garage was built

GarageFinish: Interior finish of the garage

Fin Finished

RFn Rough Finished

Unf Unfinished

NA No Garage

GarageCars: Size of garage in car capacity

GarageArea: Size of garage in square feet

GarageQual: Garage quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

NA No Garage

GarageCond: Garage condition

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

NA No Garage

PavedDrive: Paved driveway

Y Paved

P Partial Pavement

N Dirt/Gravel

WoodDeckSF: Wood deck area in square feet

OpenPorchSF: Open porch area in square feet

EnclosedPorch: Enclosed porch area in square feet

3SsnPorch: Three season porch area in square feet

ScreenPorch: Screen porch area in square feet

PoolArea: Pool area in square feet

PoolQC: Pool quality

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

NA No Pool

Fence: Fence quality

GdPrv Good Privacy

MnPrv Minimum Privacy

GdWo Good Wood

MnWw Minimum Wood/Wire

NA No Fence

MiscFeature: Miscellaneous feature not covered in other categories

Elev Elevator

Gar2 2nd Garage (if not described in garage section)

Othr Other

Shed Shed (over 100 SF)

TenC Tennis Court

NA None

MiscVal: \$Value of miscellaneous feature

MoSold: Month Sold (MM)

YrSold: Year Sold (YYYY)

SaleType: Type of sale

WD Warranty Deed - Conventional

CWD Warranty Deed - Cash

VWD Warranty Deed - VA Loan

New Home just constructed and sold

COD Court Officer Deed/Estate

Con Contract 15% Down payment regular terms

ConLw Contract Low Down payment and low interest

ConLI Contract Low Interest

ConLD Contract Low Down

Oth Other

SaleCondition: Condition of sale

Normal Normal Sale

Abnorml Abnormal Sale - trade, foreclosure, short sale

AdjLand Adjoining Land Purchase

Alloca Allocation - two linked properties with separate deeds, typically condo with a garage un

Family Sale between family members

Partial Home was not completed when last assessed (associated with New Homes)

2.3.3 การวิเคราะห์ข้อมูล (Methodology)

link URL ไปยัง notebook [Click here.](#)

2.3.4 สรุปผลการดำเนินงาน (Conclusion)

จากการใช้ขั้นตอนและวิเคราะห์ข้อมูลที่ได้จากการศึกษาในรายวิชานี้ ได้พบว่าชุดข้อมูลที่ได้มายังไม่ใช่ชุดข้อมูลที่สามารถนำมาใช้ได้เลย จำเป็นต้องมีการทำความสะอาดข้อมูลก่อนตามขั้นตอนที่เรียนมาในรายวิชานี้ แล้วจึงสามารถนำมาวิเคราะห์ข้อมูลทางสถิติเพื่อหาตัวแปรที่สามารถอธิบายถึงราคาของบ้านได้

ในส่วนของการทำโมเดลทำนายราคาบ้าน เราได้ทดลองใช้โมเดล RandomForrestRegressor() โดยการสุ่มหาพารามิเตอร์ที่ดีที่สุดในการทำนายโมเดลต่างๆ เนื่องจากการทดสอบตัวแบบ Regression เราเลือกใช้โมเดลที่มีค่า Root Mean Square Error น้อยที่สุด และ ถ้าหากว่าโมเดลนั้นมีค่า R-Square ตั้งแต่ 97% ขึ้นไป เราจะเลือกใช้โมเดลนั้นในการทำนายราคาบ้านจาก test data set ที่เป็นโจทย์ หลังจากนั้นเราได้ทำการหา features ที่สำคัญจากการที่โมเดลตัวนั้นได้เรียนรู้ไป เพื่อจะได้ทราบว่า features ใหนสำคัญต่อการทำนายราคาบ้านมากแค่ไหน