A Prediction Model for Forecasting the Trend of Macau Property Price Movements and Understanding the Influential Factors

Simon Fong
Faculty of Science and Technology, University of Macau, Macau SAR
Email: ccfong@umac.mo

Yap Bee Wah
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
Email: yapbeewah@salam.uitm.edu.my

***Abstract*—Property price and transaction activities may contain many factors. Data mining for property research provides a feasible way to analyze the trend and to understand the underlying influential factors. This paper addresses the issues and techniques on experiences in data mining for Macau property market. The original data for property price and factors are obtained as a multi-attribute dataset from the Statistics and Census Service of Macao SAR Government. The challenge is to apply different data mining methods and algorithms which include SVM, Neural Network, C&R Tree, Weka, SPSS, Multilayer Perception Model in order to identify hidden knowledge.**

***Index Terms*— data mining, SVM, Neural Network, C&R Tree, Weka, SPSS, Multilayer Perception Model**

I. Introduction

In this paper, we are presenting an example data mining application in the Macau Property policy domain. A multi-attribute dataset is downloaded from the website of Statistics and Census Service of Macao SAR Government for our data mining project. Although many statistical methods [1, 2, 3, 4] exist in analyzing property markets, using data mining methods could be an alternative; hence the objective of this project is to explore the potential powers of data mining algorithms in analyzing multi-attribute property market data. Intuitively it is known that multiple if not many underlying factors that can more or less influence the movement of the price trend, in terms of average transaction price of Macau's property. The data mining methods, such as feature selection and attribute relevance ranking [5], may help us to identify those factors, study their relative importance, and allow us to optionally eliminate the irrelevant factors when it comes to inducing a prediction model. The insights extracted from the analysis could serve as some reference information for the policy makers of Macau SAR Government to adjust its property policy, so as to have a balance in terms of supply and demand, and initiate policies that may better look after those potential home buyers to afford a house.

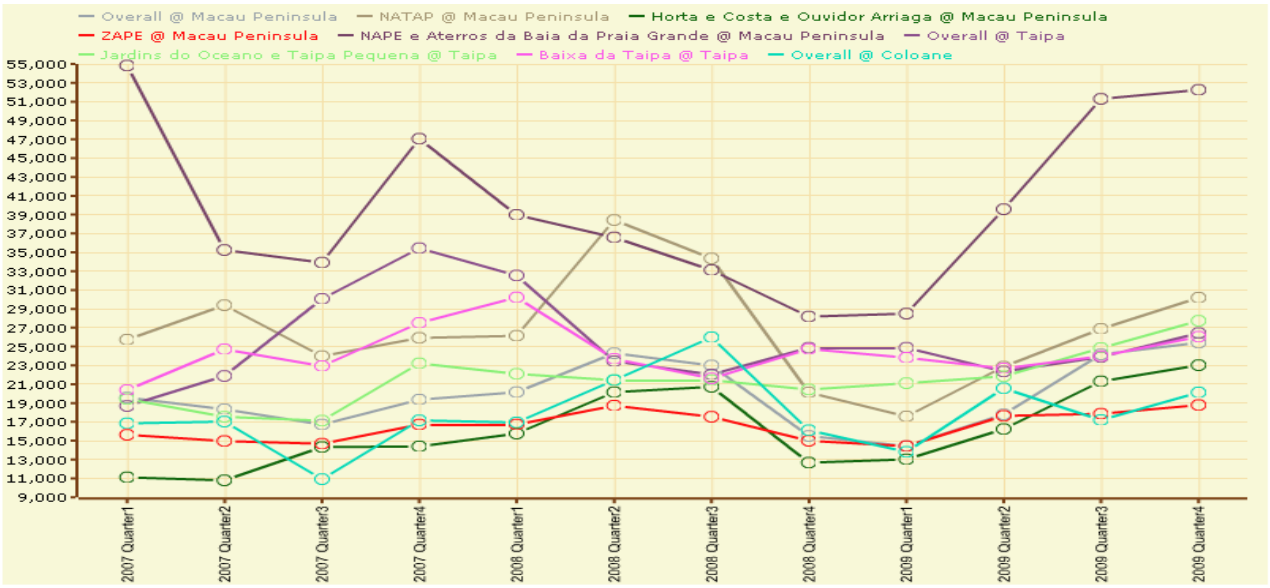

Figure 1. Average Transaction Price by Regions (MOP per sq. meter).

doi:10.4304/jetwi.5.2.122-131

By using data mining methods, a prediction model will be established that can scientifically predict the percentage of up-or-downwards movements in some of the important variables. The data mining methods can further be extended to a decision support system with easy-to-use user interface for market analysts.

In addition to predicting the moving trend, special events could be analyzed too. For instance, we can notice that there is an unusual drop of transaction price between the periods of 2008 quarter 4 and 2009 quarter 1, as shown in Figure 1. It would be intriguing to analyze, hence to find some clues on what the internal factors are for explaining the price drop, purely from data (quantitative) perspective.

II. Data Preperation

A. Datasets

The original raw data are archived at different sections of the Time Series Database[1] in the Statistics and Census Service of Macau SAR Government, free for the public to view and download. In general the time series data are recorded in quarters of a year, ranging from 2007 to 2009. Along the same axis of time and time intervals, other data such as the total number of building units, data from different industries like gaming, construction, tourism, building units transacted, as well as demographical data are acquired from the database. The data from almost all the available domains are extracted, and they are concatenated together as large multi-variate dataset, leaving the market price data as the predictable column. In total there are 121 attributes; the last column which is the predictable column is precisely defined as "average transaction price of building units transacted, calculated as percentage of change when compared to the previous quarter".

B. Pre-processing

The tasks in the process of data pre-processing include: combining data from multiple disparate sources, normalizing the numerical attributes vales into z-scores and some simple transformations such as computing the moving averages etc. The raw dataset is in a standard CSV format. It can be loaded into Weka [6] directly for instant analysis. This project mainly uses Weka and SPSS Predictive Analytics SoftWare [7] (or just SPSS) to perform the data mining research. Important attributes are screened by using feature selection techniques, and the filtered data that contain only the significant attributes will be passed through SPSS and Weka to build up a predictive model by using different kind of data mining techniques. As shown in Figure 2, the workflow of the analysis involves integrating data from separate sources into a large training set and subsequently feeding them to the two data mining software programs. The original dataset contains around the data ranging from 2007 to 2009 in 15 Excel files, each row of them represents data of a quarter of a year. The downloaded data are already in hierarchically structured data. The original data contains 12 rows with 11 catalogs. Combining those data to one multi-variable catalog produces a total of 132 data instances for training the classification model.

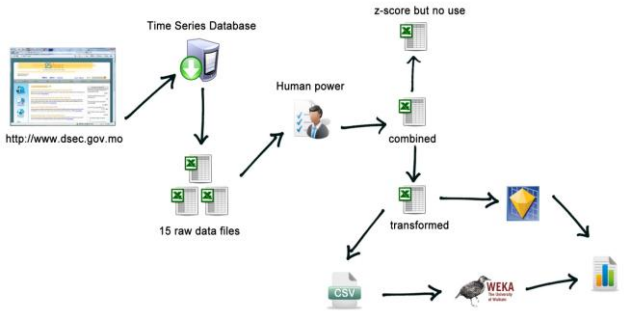

Figure 2. Workflow of the analysis.

Without being exhaustive, the full listing of the 121 attributes is presented in Appendix A.

C. Feature Selection

Some of the attributes may not be so relevant to the predictive power of the prediction model. Also having too many attributes and relatively too few the training instances may result in 'under-fitting' that means the prediction model is not generalized enough and prediction accuracy drops. Hence it is necessary to select only the significant attributes and eliminate those that are not so significant. Various feature selection methods and tools exist in both SPSS and Weka software programs [8]. In this project, two methods for evaluating the importance of the variables are used in SPSS. They are SVM, and Neural Net. The tool called 'Auto Numeric' produce the following result: Using the Auto Numeric node, the researcher (user) may compare a number of models that predict the property values based on the building type, neighborhoods or proximity, area size, and other known factors. According to the C&R Tree [9], the Number of Fields used only 23 original data instances that are described by 121 fields. In the following stream in SPSS, we only select top 23 important attributes based on the consideration of the balance of each method. A screenshot of the SPSS feature selection process is shown in Figure 3. The two resultant C&R trees by SVM and Neural Net are shown in Figures 4 and 5 respectively. In SVM and Neural Net, the time period and regions were selected by the program results to be eliminated.

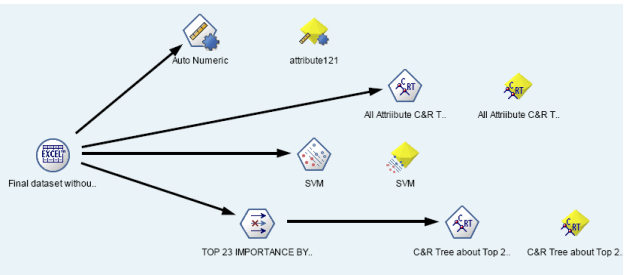

Figure 3. Stream of 'Select 23 top important attributes' by SPSS.

[1] http://www.dsec.gov.mo/TimeSeriesDatabase.aspx

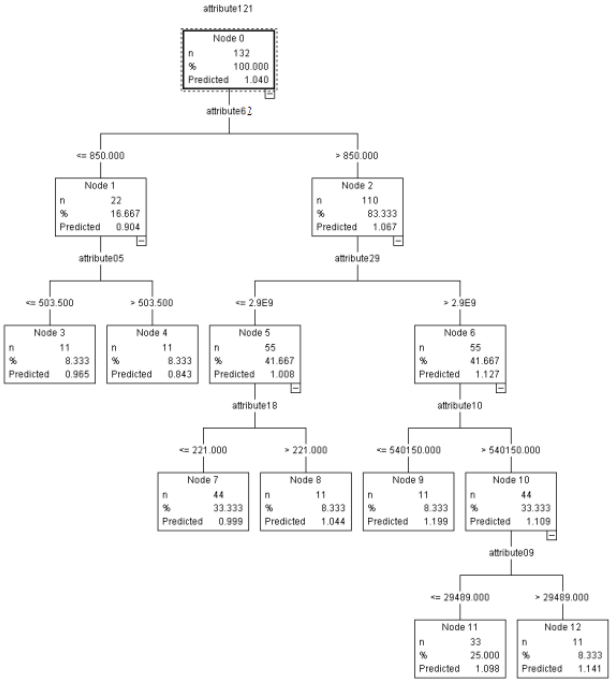

Figure 4. C&R tree for feature selection by SVM.

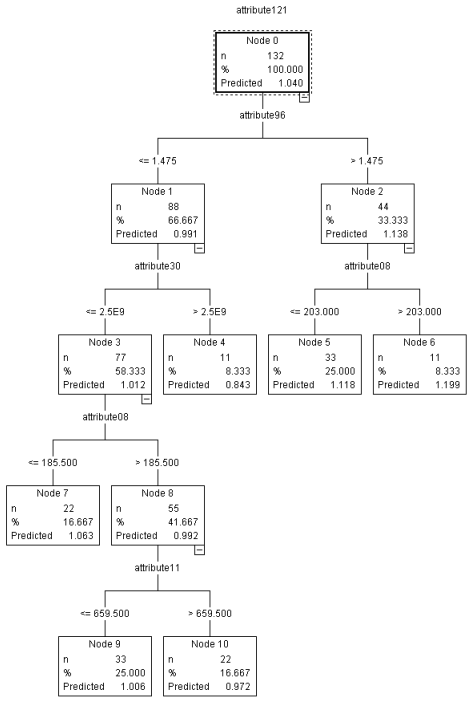

Figure 5. C&R tree for feature selection by Neural Net.

When the average transaction price of building units transacted (percentage change on previous quarter) is less than 1, it means transaction price of building is dropping.

In weka, there are two algorithms for selecting attributes:

a). Select attributes using *CfsSubsetEval* and *GreedyStepwise* search.

b). Select attributes using *CfsSubsetEval* and *ScatterSearchV1* search.

Figure 6 shows a screenshot of Weka where the attributes are ready to be processed by the feature selection functions.

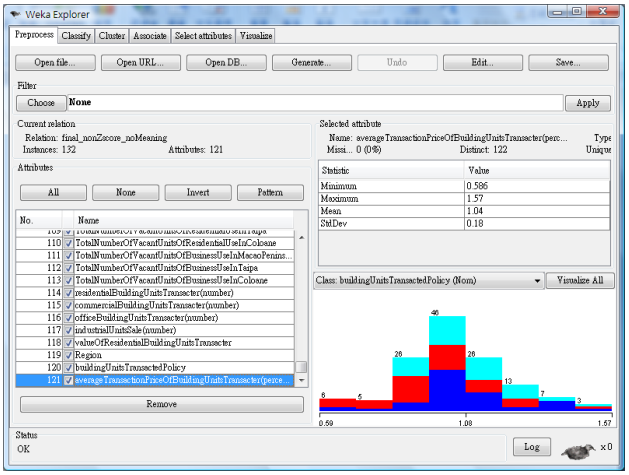

Figure 6. Screenshot of Weka Explorer.

The run information from the two feature selection functions in Weka platform is shown in Appendix B and Appendix C respectively. As the log information shows in Appendix B, 25 significant attributes using *CfsSubsetEval* and *GreedyStepwise* search are obtained.

With the *CfsSubsetEval* and *ScatterSearch* search functions installed in Weka, 11 selected attributes are being created, as shown in the information log in Appendix C. With Weka feature selection analysis, two sets of data are being created with two different amounts of attributes. In total, after the feature selection processes, insignificant features are removed; two sets of data each are made available for SPSS for Weka respectively, for predictive modeling.

III. Prediction Models Construction

A. SPSS

Two machine learning algorithms are selected to build a predictive model in SPSS, they are Neural Network and SVM. The choice is justified by the powerful and non-linear predictive models which are tested to be some of the most suitable for generalizing a predictive model with highly non-linear relationships among the attributes (features) and few available training instances. The relevant processes for training the models with given training data and for testing the models are shown respectively in Figure 7 and Figure 8.

To test the model, we just randomly choose half the amount of the training records for cross-validation. The original class value is 1.17058413. Compare with the predictive value in Neural Network model and SVM model, Neural Network = 1.086 and SVM = 1.119. It shows that SVM is much better in accuracy than Neural Network. And test the record of 100. The original class value is 0.64148287. Neural Network = 1.060 and SVM = 0.989. It shows also that SVM better than Neural Network in the other dataset. Recall the C&R tree with SVM model: attribute05, the number of units new build

which is the most significant attribute. From the analysis by SPSS it is observed that the attributes 05, 09, 10, 18, 29, 62 are the most essential factors directly to property transaction price movement. The meanings of the significant attributes, in an order of importance, are summarized as follow - 09: gross floor area of new building; 10: Population; 18: new services of financial, leasing and provided to company; 29: Government final consumption expenditure; 62: Construction Population.

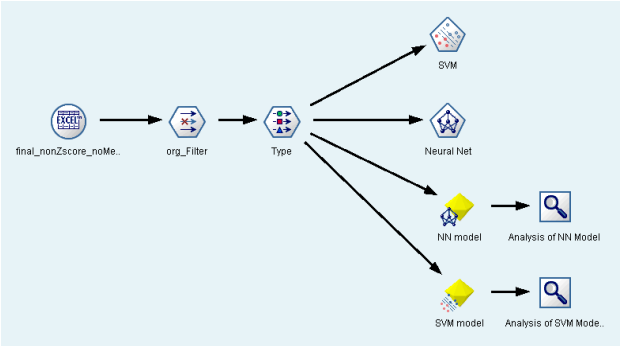

Figure 7. Model training in SPSS.

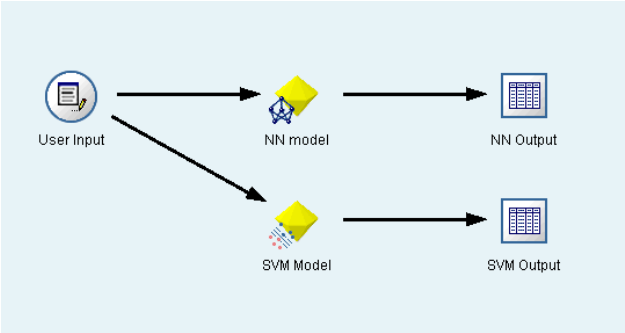

Figure 8. Model testing in SPSS.

The performance comparison of the two predictive models as in SPSS is summarized in Table I, in terms of prediction errors in percentage.

TABLE I.
PERFORMANCE COMPARISON BETWEEN SVM AND NN IN SPSS

Performance	SVM	NN
Minimum Error	-0.411	-0.419
Maximum Error	0.4	0.454
Mean Error	-0.005	0.028
Mean Absolute Error	0.117	0.123
Standard Deviation	0.157	0.158
Linear Correlation	0.492	0.474
Occurrences	132	132

B. *Weka*

In Weka, two of the most effective classification algorithms are adopted; the choice is based on some preliminary testing of most of the algorithms available in Weka over the given dataset. The selected algorithms are Multilayer Perception and Linear Regression, they are known to be able to generate a solution from a highly non-linear data environment such as the one being used here. The weights between the hidden layers and the coefficients of the linear regression model scale up to a large hyperspace. A comparison of performance as well as the complexity of the coefficients values between Multilayer Perception (MP) and Linear Regression (LR) is presented in Table II.

TABLE II.
PERFORMANCE COMPARISON BETWEEN MP AND LR IN WEKA

Performance	MP	LR
Correlation coefficient	0.9978	0.4952
Class complexity at order 0	853.0085 bits	853.0085 bits
Average class complexity at order 0	6.4622 bits/instance	6.4622 bits/instance
Class complexity per scheme	1417.8986 bits	1900.3814 bits
Average class complexity per scheme	10.7417 bits/instance	14.3968 bits/instance
Complexity improvement	-564.8901 bits	-1047.373 bits
Average complexity improvement	-4.2795 bits/instance	-7.9346 bits/instance
Mean absolute error	0.0141	0.1151
Root mean square error	0.0175	0.1557
Relative absolute error	10.3586 %	84.6007 %
Root relative squared error	9.7364 %	86.8766 %
Total instances	132	132

Comparing over MP and LR, it can easily be seen that MP can observe superior performance in terms of lower error rate and simpler model complexity than LR. While SPSS is able to achieve the best model so far with only 6 selected features, analysts may opt to investigate a larger pool of features and their impact on the model accuracy. Therefore, in Weka, the MP as the selected candidate would be subject to a filtered dataset that consist of 25 selected features, as obtained from the process in Section II C.

In this case the model is built again in Weka, and the experiment proceeds to compare the net results from the two feature selection algorithms. The performance results are then produced solely by using MP this time, over the datasets extracted by using two feature selection algorithms, and the results are listed in Table III.

TABLE III.
PERFORMANCE COMPARISON BETWEEN TWO FEATURE SELECTION ALGORITHMS BY MP IN WEKA

Performance	MP with *CfsSubsetEval* and *GreedyStepwise*	MP with *CfsSubsetEval* and *ScatterSearch*
Correlation coefficient	**0.9978**	0.9737
Class complexity at order 0	853.0085 bits	853.0085 bits
Average class complexity at order 0	6.4622 bits/instance	6.4622 bits/instance
Class complexity per scheme	1417.8986 bits	1653.6309 bits
Average class complexity per scheme	10.7417 bits/instance	12.5275 bits/instance
Complexity improvement	-564.8901 bits	-800.6224 bits
Average complexity improvement	-4.2795 bits/instance	-6.0653 bits/instance
Mean absolute error	0.0141	0.0318
Root mean square error	0.0175	0.0419
Relative absolute error	10.3586 %	23.3988 %
Root relative squared error	9.7364 %	23.3719 %
Total instances	132	132

Observing from Table III, it can be confirmed that the Multilayer Perception Model of reduced data by using *CfsSubsetEval* and *GreedyStepwise* feature selection functions give a better result than its counter-part. This is evident by its larger correlation coefficient value, 0.9978. In fact, the coefficient value is even greater than those obtained from SPSS model, 0.492 and 0.474. From the empirical evidences shown so far, it is sensible to adopt the Multilayer Perception model in Weka together with the feature selection function combo, *CfsSubsetEval* and *GreedyStepwise*, for doing the analysis. As a result, the essential attributes being selected from the selected algorithms are: 3, 4, 6, 8, 13, 15, 18, 21, 22, 23, 28, 31, 65, 66, 70, 76, 81, 86, 90, 96, 98, 115, 117, 118, and 119.

A summary of the essential attributes are listed below in Table IV.

TABLE IV.
THE 25 ATTRIBUTES ADVOCATED BY THE BEST PERFORMING ALGORITHMS IN WEKA

3.Gaming tax revenue	4.Number of completed buildings	6.Building area	8.Number of new units	13.Persons authorized to reside in Macao (Inflow)
15. 25-34 years old (compared with 2004)	18.New services of financial, leasing and provided to company	21.Financial activities, renting and services to companies	22.Number of dissolved companies	23.Captial of dissolved companies
28.Final consumption	31.Net purchases of goods and services of government final consumption expenditure	65.Transport-Storage & Communications population	66.Finance population	70.Health and social welfare population
76.Technicians and associate professionals population	81.Plant and machine operator, drivers and assemblers population	86.Unemployment rate(%) female	90.Retail sales of motor vehicles	96.No. of hotel guest in hotel sector
98.Income of inward direct investment	115.Commercial building units transacted (number)	117.Industrial units sale (number)	118.Value of residential building units transacted	119.Region

IV. CONCLUSION

This paper contributes to a method that uses SPSS Modeler and Weka data mining software to explore the underlying relations of Macau property market. The dataset is multi-variate consists of 121 attributes and 132 instances dataset, obtained from Statistics and Census Service of Macao SAR Government, under the section of Macau Property. With the C&R tree for selection by SVM, Neutral Net of SPSS and the feature selection functions such as *CfsSubsetEval* and *GreedyStepwise* search, *CfsSubsetEval* and *ScatterSearch* functions of Weka, the less significant attributes that are affecting property price were eliminated. Predictive models according to the filtered data with SPSS, Weka are built. The algorithms deployed are Multilayer Perception Neural Network model. From the results of correlation coefficients and accuracies of the models, Multilayer Perception seems to be the most effective; subsequently 25 most relevant attributes are harvested that are computed to be the most influential factors to the property market price for that given period of time.

ACKNOWLEDGMENT

The authors wish to thank the following MSc students of the University of Macau who have contributed both technically and intellectually to this project: Mr. Ng Ka Kit, Mr. Chan Io Weng and Ms. Liu Liyang.

REFERENCES

[1] C. Leishman, *Real Estate Market Research and Analysis*. Palgrave Macmillan, April 2003.
[2] J. M. Clapp, *Handbook for real estate market analysis*. Prentice-Hall, 1987.
[3] J. M. Clapp and S. D. Messner, *Real Estate Market Analysis: Methods and Applications*, ABC-CLIO, 1998.
[4] N. G. Carn, *Real estate market analysis: techniques and applications*, Prentice-Hall, 1988.
[5] S. Fong, Y. Zhuang, M. Yu and I. Ma, "Quantitative analysis of trust factors on social network using data mining approach," *IEEE 2012 International Conference on Future Generation Communication Technology (FGCT)*, pp. 70-75, December 2012.
[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, *The Weka Data Mining Software: An Update*. SIGKDD Explorations, Volume 11, Issue 1, 2009.
[7] SPSS Inc. Released 2009. *PASW Statistics for Windows*, Version 18.0. Chicago: SPSS Inc.
[8] S. Fong and A. Cerone, "Attribute Overlap Minimization and Outlier Elimination as Dimensionality Reduction Techniques for Text Classification Algorithms," *Journal of Emerging Technologies in Web Intelligence (JETWI)*, Academy Publisher, ISSN 1798-0461, Volume 4, Issue 3, Oulu, Finland, pp. 259-263, August 2012.
[9] L. Rokach and O. Maimon, *Data mining with decision trees: theory and applications*. World Scientific Pub Co Inc., 2008.

APPENDIX A FULL LISTING OF ATTRIBUTES

Name	Meaning of Attribute	Type
attribute01	Time	Nominal
attribute02	gross gaming revenue total	Numerical
attribute03	gaming tax revenue	Numerical
attribute04	the number of completed buildings	Numerical
attribute05	the number of units on new build	Numerical
attribute06	building area *	Numerical
attribute07	the number of new buildings	Numerical
attribute08	the number of new units	Numerical
attribute09	gross floor area of new building	Numerical
attribute10	Population	Numerical
attribute11	number of married	Numerical
attribute12	newborn baby	Numerical
attribute13	Persons authorized to reside in Macao (Inflow)	Numerical
attribute14	Non-resident workers (End balance)	Numerical
attribute15	25-34 years old (compared with 2004)	Numerical
attribute16	newly construction company	Numerical
attribute17	the newly Real Estate Activities Business Company	Numerical
attribute18	new services of financial, leasing and provided to company	Numerical
attribute19	real estate activities business capital	Numerical
attribute20	Building services capital	Numerical
attribute21	Financial activities, renting and services to companies	Numerical
attribute22	number of dissolved companies	Numerical
attribute23	capital of dissolved companies	Numerical
attribute24	Per capita GDP	Numerical
attribute25	Gross Domestic Product (GDP)	Numerical
attribute26	the total number of private consumption expenditure	Numerical
attribute27	Household final consumption expenditure	Numerical
attribute28	Final consumption expenditure of NPISHs	Numerical
attribute29	Government final consumption expenditure	Numerical
attribute30	Compensation of employees of government final consumption expenditure	Numerical
attribute31	Net purchases of goods and services of government final consumption expenditure	Numerical
attribute32	the total number of gross fixed capital formation	Numerical
attribute33	Construction	Numerical
attribute34	Machinery and equipment	Numerical
attribute35	Changes in inventories	Numerical
attribute36	the total number of export	Numerical
attribute37	Exports of goods	Numerical
attribute38	Exports of services	Numerical
attribute39	total number of import	Numerical
attribute40	Imports of goods	Numerical
attribute41	Imports of services	Numerical
attribute42	total number of median of monthly employment earnings	Numerical
attribute43	the median monthly employment earnings in manufacturing	Numerical
attribute44	the median monthly employment earnings in Electricity, gas and water supply	Numerical
attribute45	the median monthly employment earnings in construction	Numerical
attribute46	the median monthly employment earnings in Wholesale and retail trade	Numerical
attribute47	the median monthly employment earnings in Hotels and restaurants	Numerical
attribute48	the median monthly employment	Numerical

	earnings in transport and communications and storage industry	
attribute49	the median monthly employment earnings in financial industry	Numerical
attribute50	the median monthly employment earnings in Real Estate and Business Activities	Numerical
attribute51	the median monthly employment earnings in Public administration and social security	Numerical
attribute52	the median monthly employment earnings in Education	Numerical
attribute53	the median monthly employment earnings in health and social	Numerical
attribute54	the median monthly employment earnings in Recreational, cultural, gaming and other services	Numerical
attribute55	the median monthly employment earnings in Domestic workers domestic	Numerical
attribute56	local household savings deposits	Numerical
attribute57	residents deposits	Numerical
attribute58	non-resident deposits	Numerical
attribute59	Private loans / residential purpose loans	Numerical
attribute60	manufacturing population of	Numerical
attribute61	water and electricity and gas and water supply of population	Numerical
attribute62	Construction Population	Numerical
attribute63	Wholesale and retail trade population	Numerical
attribute64	Hotels and restaurants Population	Numerical
attribute65	Transport-Storage & Communications Population	Numerical
attribute66	finance Population	Numerical
attribute67	Real Estate and Business Activities	Numerical
	Population	
attribute68	public administration and social security services Population	Numerical
attribute69	Education Population	Numerical
attribute70	Health and Social Welfare Population	Numerical
attribute71	Recreational, cultural, gaming and other services Population	Numerical
attribute72	Of which: Gaming industry	Numerical
attribute73	domestic worker population	Numerical
attribute74	Legislators, senior officials, directors and managers of companies Population	Numerical
attribute75	Professionals	Numerical
attribute76	technicians and associate professionals	Numerical
attribute77	clerks	Numerical
attribute78	service,sales and similar staff	Numerical
attribute79	Skilled agricultural and fishery workers	Numerical
attribute80	Craftsmen and similar workers	Numerical
attribute81	Plant and machine operator, drivers and assemblers	Numerical
attribute82	unskilled workers	Numerical
attribute83	Labor Force Participation Rate (%) male	Numerical
attribute84	Labor Force Participation Rate (%) Female	Numerical
attribute85	unemployment rate (%)-Men	Numerical
attribute86	Unemployment rate (%) Female	Numerical
attribute87	underemployment rate (%)	Numerical
attribute88	Retail sales of goods in department store retail sales (Million MOP)	Numerical
attribute89	Retail sales of goods in department store retail sales	Numerical
attribute90	Retail sales of motor vehicles	Numerical
attribute91	Retail sales of goods in supermarkets	Numerical

attribute92	retail sales of watches, clocks and jewellery	Numerical
attribute93	Retail sales in adults' clothing	Numerical
attribute94	Retail sales in automotive fuels	Numerical
attribute95	Hotel occupancy rate by classification of establishments in hotel sector	Numerical
attribute96	No. of hotel guest in hotel sector	Numerical
attribute97	Flows of inward direct investment	Numerical
attribute98	Income of inward direct investmen	Numerical
attribute99	Stock of outward direct investment	Numerical
attribute100	Flows of outward direct investment	Numerical
attribute101	Income of outward direct investmen	Numerical
attribute102	Total number of building units of residential use in Macau Peninsula	Numerical
attribute103	Total number of building units of residential use in Taipa	Numerical
attribute104	Total number of building units of residential use in Coloane	Numerical
attribute105	Total number of building units of Business use in Macau Peninsula	Numerical
attribute106	Total number of building units of Business use in Taipa	Numerical
attribute107	Total number of building units of Business use in Coloane	Numerical
attribute108	Total number of vacant units of Residential use in Macau Peninsula	Numerical
attribute109	Total number of vacant units of Residential use in Taipa	Numerical
attribute110	Total number of vacant units of Residential use in Coloane	Numerical
attribute111	Total number of vacant units of Business use in Macao Peninsula	Numerical
attribute112	Total number of vacant units of Business use in Taipa	Numerical
attribute113	Total number of vacant units of Business use in Coloane	Numerical
attribute114	residential building units transacted (number)	Numerical
attribute115	commercial building units transacter (number)	Numerical
attribute116	office building units transacted (number)	Numerical
attribute117	industrial units sale (number)	Numerical
attribute118	value of residential building units transacted	Numerical
attribute119	Region	Nominal
attribute120	building units transacted policy **a** - Including the number of residential building units sold in 2008 at or below the value of MOP3 million and thus exempted from the payment of stamp duty, in accordance with Article 14 of Law no.7/2007. **b** - Including the number of residential building units sold in 2009 at or below the value of MOP3 million and thus exempted from the payment of stamp duty, in accordance with Article 14 of Law no.15/2008 .	Nominal
attribute121	average transaction price of building units transacted (percentage change on previous quarter)	Numerical

APPENDIX B RUN INFORMATION BY CFSSUBSETEVAL AND GREEDYSTEPWISE IN WEKA

```
=== Run information ===
Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:       weka.attributeSelection.GreedyStepwise -T -
1.7976931348623157E308 -N -1
Relation:     final_nonZscore_noMeaning
Instances:    132
Attributes:   121
              [list of attributes omitted]
Evaluation mode:    evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
              Greedy Stepwise (forwards).
              Start set: no attributes
              Merit of best subset found:    0.344
Attribute Subset Evaluator (supervised, Class (numeric): 121
averageTransactionPriceOfBuildingUnitsTransacter(percentage
ChangeOnPreviousQuarter)):
              CFS Subset Evaluator
              Including locally predictive attributes
Selected attributes:
3,4,6,8,13,15,18,21,22,23,28,31,65,66,70,76,81,86,90,96,98,115
,117,118,119 : 25
        gamingTaxRevenue
        theNumberOfCompletedBuildings
        buildingArea
        theNumberOfNewUnits
        PersonsAuthorizedToResideInMacao(Inflow)
        25-34YearsOld(comparedWith2004)
        newServicesOfFinancial_leasingAndProvidedToCompany
        FinancialActivities_RentingAndServicesToCompanies
        numberOfDissolvedCompanies
        capitalOfDissolvedCompanies
        FinalConsumptionExpenditureOfNPISHs
NetPurchasesOfGoodsAndServicesOfGovernmentFinalConsum
ptionExpenditure
        Transport-StorageAndCommunicationsPopulation
        financePopulation
        HealthAndSocialWelfarePopulation
        techniciansAndAssociateProfessionals
        PlantAndMachineOperator_DriversAndAssemblers
        UnemploymentRateFemale
        RetailSalesOfMotorVehicles
        No_OfHotelGuestInHotelSector
        IncomeOfInwardDirectInvestmen
        commercialBuildingUnitsTransacter(number)
        industrialUnitsSale(number)
        valueOfResidentialBuildingUnitsTransacter
        Region
```

APPENDIX C RUN INFORMATION BY CFSSUBSETEVAL AND SCATTERSEARCH IN WEKA

```
=== Run information ===
Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:       weka.attributeSelection.ScatterSearchV1 -T 0.0 -Z -1 -R 0 -
S 1 -D
Relation:     final_nonZscore_noMeaning
Instances:    132
Attributes:   121
              [list of attributes omitted]
Evaluation mode:    evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
              Scatter Search
              Init Population: 60
              Kind of Combination: Greedy Combination
              Random number seed: 1
              Debug: true
              Treshold:    0
              Total number of subsets evaluated: 289888
              Merit of best subset found:    0.397
Population: 60
merit       subset
 0.30073    [30, 66, 84, 98, 104]
 0.29014    [81, 86]
 0.28892    [76]
 0.28537    [15]
 0.28309    [21, 113]
 0.31566    [1, 90, 99]
 0.27164    [107]
 0.2675     [55, 90]
 0.2574     [109]
 0.25722    [115]
 0.25309    [113]
 0.24766    [117]
 0.28309    [10, 21]
 0.24687    [108]
 0.24396    [1, 22]
 0.23938    [65]
 0.23059    [86]
 0.22001    [101]
 0.2675     [90, 102]
 0.19381    [96]
 0.19355    [13]
 0.19225    [6, 55, 104]
 0.19678    [60, 102]
 0.19174    [120]
 0.19089    [104]
 0.18403    [70]
 0.17739    [55]
 0.17678    [28]
 0.17599    [18]
 0.20186    [1, 4, 19, 36, 45, 60, 84, 101, 107, 118]
 0.16998    [56, 85]
 0.2546     [65, 79]
 0.15662    [85]
 0.15654    [84, 105]
 0.27321    [14, 60, 99]
 0.19326    [15, 19, 28, 30, 45, 56, 81, 90, 94, 118]
 0.14488    [66]
 0.14404    [8]
 0.14173    [84]
 0.21065    [2, 8, 10, 20, 22, 24, 28, 31, 35, 43, 56, 57, 66, 70, 76, 84,
90, 98, 107, 108, 115, 117]
 0.13777    [23, 70, 81, 109]
 0.17265    [14, 24, 28, 35, 55, 56, 102, 117]
 0.22526    [35, 45, 55, 101]
 0.25356    [6, 21, 33, 35, 36, 45, 55, 60, 84, 85, 98, 108]
 0.19893    [60, 114]
 0.26603    [5, 65]
 0.12721    [6, 10, 18, 19, 30, 55, 61, 76, 81, 105, 108]
 0.17718    [2, 3, 5, 8, 18, 25, 55, 66, 74, 81, 82, 94, 102, 113, 118]
 0.22402    [4, 10, 15, 28, 29, 81, 99]
 0.14305    [20, 21, 22, 38, 57, 85, 86]
 0.28309    [1, 5, 15, 21, 61, 120]
 0.13979    [1, 3, 4, 18, 19, 30, 38, 45, 57, 65, 94, 114, 115, 117]
 0.15749    [1, 18, 28, 36, 55, 57, 60, 65, 81, 86, 105, 107, 108, 109]
 0.17294    [1, 13, 25, 30, 31, 33, 56, 60, 85, 90, 96, 104, 120]
 0.27321    [4, 64, 99, 104]
 0.17633    [5, 13, 14, 43, 86, 96, 108, 115]
 0.26654    [1, 13, 22, 65, 94, 99, 102, 115]
 0.11735    [43, 61]
 0.17156    [8, 31, 45, 56, 64, 102, 113]
 0.1447     [2, 10, 28, 33, 55, 82, 90]

ReferenceSet:
----------Most Significants Solutions----------
```

0.31566 [1, 90, 99]
0.30073 [30, 66, 84, 98, 104]
0.29014 [81, 86]
0.28892 [76]
0.28537 [15]
0.28309 [10, 21]
0.28309 [1, 5, 15, 21, 61, 120]
0.28309 [21, 113]
0.27321 [4, 64, 99, 104]
0.27321 [14, 60, 99]
0.27164 [107]
0.2675 [90, 102]
0.2675 [55, 90]
0.26654 [1, 13, 22, 65, 94, 99, 102, 115]
0.26603 [5, 65]
----------Most Diverses Solutions----------
0.21065 [2, 8, 10, 20, 22, 24, 28, 31, 35, 43, 56, 57, 66, 70, 76, 84, 90, 98, 107, 108, 115, 117]
0.15662 [85]
0.22001 [101]
0.17599 [18]
0.19381 [96]
0.13777 [23, 70, 81, 109]
0.19893 [60, 114]
0.15654 [84, 105]
0.17678 [28]
0.14488 [66]
0.14173 [84]
0.17739 [55]
0.19678 [60, 102]
0.19355 [13]
0.19326 [15, 19, 28, 30, 45, 56, 81, 90, 94, 118]

Last Reference Set Updated:
merit subset
0.39689 [18, 22, 55, 65, 81, 86, 105, 108, 115, 117]
0.39688 [18, 22, 55, 65, 81, 105, 108, 115, 117]
0.39682 [22, 55, 65, 81, 86, 105, 108, 115, 117]
0.39682 [22, 55, 65, 81, 105, 108, 115, 117]
0.39604 [18, 22, 65, 81, 86, 105, 108, 115, 117]
0.39603 [18, 22, 65, 81, 105, 108, 115, 117]
0.39586 [22, 65, 81, 86, 105, 108, 115, 117]
0.39585 [22, 65, 81, 105, 108, 115, 117]
0.21065 [2, 8, 10, 20, 22, 24, 28, 31, 35, 43, 56, 57, 66, 70, 76, 84, 90, 98, 107, 108, 115, 117]
0.28309 [1, 5, 15, 21, 61, 120]
0.3653 [4, 21, 64, 96, 99]
0.19678 [60, 102]
0.15662 [85]
0.22001 [101]
0.19893 [60, 114]
0.19381 [96]
0.28892 [76]
0.27321 [14, 60, 99]
0.3916 [22, 55, 65, 81, 86, 105, 115]
0.39144 [22, 55, 65, 81, 86, 109, 115]
0.38848 [22, 65, 81, 86, 109, 115]
0.38848 [22, 65, 81, 109, 115]
0.3653 [1, 4, 21, 61, 64, 99]
0.3653 [4, 21, 64, 99]
0.32131 [22, 76, 84, 108, 115, 117]
0.31892 [4, 8, 21, 22, 28, 31, 64, 66, 70, 76, 81, 84, 90, 98, 115, 117]
0.31892 [1, 8, 21, 22, 28, 31, 66, 70, 76, 81, 84, 90, 98, 115, 117]
0.3126 [8, 22, 28, 31, 66, 70, 76, 81, 84, 90, 98, 114, 115, 117]
0.3126 [8, 22, 28, 31, 66, 70, 76, 81, 84, 90, 96, 98, 115, 117]
0.30057 [15, 21, 60, 76, 81, 85, 98, 114]
Attribute Subset Evaluator (supervised, Class (numeric): 121 averageTransactionPriceOfBuildingUnitsTransacter(percentageChange OnPreviousQuarter)):
CFS Subset Evaluator
Including locally predictive attributes
Selected attributes: 18,22,55,65,81,86,105,108,115,117,119 : 11
newServicesOfFinancial_leasingAndProvidedToCompany
numberOfDissolvedCompanies

theMedianMonthlyEmploymentEarningsInDomesticWorkersDomestic
Transport-StorageAndCommunicationsPopulation
PlantAndMachineOperator_DriversAndAssemblers
UnemploymentRateFemale
TotalNumberOfBuildingUnitsOfBusinessUseInMacauPeninsula
TotalNumberOfVacantUnitsOfResidentialUseInMacauPeninsula
commercialBuildingUnitsTransacter(number)
industrialUnitsSale(number)
Region