

應用資料探勘技術於房價預測上之研究

Data Mining on Home Price Prediction

李御璽 顏秀珍
銘傳大學資工系 銘傳大學資工系
leeys@mcu.edu.tw sjyen@mcu.edu.tw

楊順安 陳心懋 吳政瑋
銘傳大學資工所 銘傳大學資工所 銘傳大學資工所
coolancool@hotmail.com general_x10@yahoo.com.tw silvemoonfox@hotmail.com

摘要

近年來，物價的波漲是一波接著一波。房屋對多數人來說，屬於高價值商品，交易金額相對龐大，所以在決定買賣房屋時的考慮也相對複雜。房屋仲介業者在訂定房屋價格時，如果定的不適當，那便會造成損失，所以在訂定房屋價格時，必定要嚴謹。本研究運用資料探勘(Data Mining)中的預測(Prediction)技術，利用過去成交的房屋資料，建立一套預測房屋價格的模型。資料探勘是一項重要的技術，它可以找出隱藏在龐大資料背後的資訊，做為業務行銷、推展的依據。在現今競爭激烈的市場環境中，它是一個相當重要的分析工具。此外，在資料探勘的過程中，資料的前置處理(Data Preprocessing)是相當重要的一環，不但消耗的時間和精力最多，也往往是影響探勘結果最重要的因素。因此，本實驗會先將資料做前置處理再利用資料探勘的技術預測房屋的成交價。本研究所使用的技術有決策樹(Decision Tree)、類神經網路(Neural Networks)、和線性迴歸(Linear Regression)三種模型，而使用的工具為 SQL Server 2005。本研究也會使用自己的方法來挑選欄位。此實驗最後會比較三種模型的預測結果，並且還會比較我們挑選欄位的方法與工具所建議的欄位哪個會具有較好的探勘結果。根據實驗結果顯示決策樹模型會有較接近實際價格的預測值，而且本實驗挑選屬性之方法優於 SQL Server 2005 所建議之欄位。

關鍵字：資料探勘、決策樹、類神經網路、線性迴歸、預測

應用資料探勘技術於房價預測上之研究

Data Mining on Home Price Prediction

李御璽
銘傳大學資工系
leeys@mcu.edu.tw

顏秀珍
銘傳大學資工系
sjyen@mcu.edu.tw

楊順安
銘傳大學資工所
coolancool@hotmail.com

陳心懋
銘傳大學資工所
general_x10@yahoo.com.tw

吳政瑋
銘傳大學資工所
silvemoonfox@hotmail.com

摘要

近年來，物價的波漲是一波接著一波。房屋對多數人來說，屬於高價值商品，交易金額相對龐大，所以在決定買賣房屋時的考慮也相對複雜。房屋仲介業者在訂定房屋價格時，如果定的不適當，那便會造成損失，所以在訂定房屋價格時，必定要嚴謹。本研究運用資料探勘(Data Mining)中的預測(Prediction)技術，利用過去成交的房屋資料，建立一套預測房屋價格的模型。資料探勘是一項重要的技術，它可以找出隱藏在龐大資料背後的資訊，做為業務行銷、推展的依據。在現今競爭激烈的市場環境中，它是一個相當重要的分析工具。此外，在資料探勘的過程中，資料的前置處理(Data Preprocessing)是相當重要的一環，不但消耗的時間和精力最多，也往往是影響探勘結果最重要的因素。因此，本實驗會先將資料做前置處理再利用資料探勘的技術預測房屋的成交價。本研究所使用的技術有決策樹(Decision Tree)、類神經網路(Neural Networks)、和線性迴歸(Linear Regression)三種模型，而使用的工具為SQL Server 2005。本研究也會使用自己的方法來挑選欄位。此實驗最後會比較三種模型的預測結果，並且還會比較我們挑選欄位的方法與工具所建議的欄位哪個會具有較好的探勘結果。根據實驗結果顯示決策樹模型會有較接近實際價格的預測值，而且本實驗挑選屬性之方

法優於SQL Server 2005所建議之欄位。

關鍵字：資料探勘、決策樹、類神經網路、線性迴歸、預測

1. 導論

房地產有位置、數量、品質上差異的個別特性，表現在市場上即為各種各樣的房地產商品，如商場(店)、廠房、辦公及住宅等。而在國人傳統「有土斯有財」的觀念下，房屋被視為投資保值的最佳標的。房價高低則被視為衡量個人與社會財富的表徵，且房屋對多數人來說，屬於高價值商品，交易金額相對龐大，所以在決定買賣房屋時的考慮也相對複雜。所以房屋價格不能隨便亂訂定，定的太高反而賣不出去；定的太低有利潤太少或是沒有利潤的問題。如何訂定合理的價格，我們可以利用過去房屋成交的資料來進行資料探勘。

資料探勘(Data Mining)是一項重要的技術，它可以找出隱藏在龐大資料背後的資訊以做為業務行銷、推展的依據，在現今競爭激烈的市場環境是一個相當重要的分析工具[1]。本研究是使用資料探勘中的預測(Prediction)技術[2,3,4,5,6]及近幾年台北縣房屋成交的資料庫，以Microsoft SQL Server 2005為探勘工具，並使用三種模型：決策樹(Decision Tree)、類神經網路(Neural Networks)、和線性迴歸

(Linear Regression)，並比較此三種模型預測的結果，此外，我們還會針對有挑選欄位的方法與工具所提之欄位進行比較。

進行資料探勘時，資料的正確性、完整性皆會影響探勘結果，然而資料探勘的資料往往存在著許多的雜訊，像是缺值、空白字串或與屬性定義不符合的值，另外也存在著許多意義不明之值，在這種情況下探勘得到之結果，可能會發生嚴重的偏差，所以在做資料探勘前，先將資料做前置處理(Data Preprocessing)可以減少探勘過程中所花費的時間和成本，更提高所獲取知識的準確性、利用性和可讀性，故前置處理是非常重要的。

本論文的研究過程如下，首先進行資料前置處理的工作。處理的項目包括檢視原始資料、填補空值等。接著將資料分成 80% 的訓練資料(Training Data)以及 20% 的測試資料(Testing Data)，再將訓練資料分別建立決策樹，類神經網路，線性迴歸等三種模型。最後評估模型的效能。其過程將在第三節詳述。

本論文的章節安排如下：下一節將介紹近年來成將房屋的資料庫，以及資料庫中每個欄位的意義及屬性。第三節介紹研究方法，包括資料前置處理、屬性選擇的計算方法、評估效能的準則。第四節介紹實驗與評估的結果。最後一節為結論與未來研究方向。

2. 資料庫描述

料庫共有 4804 筆資料。每一筆紀錄代表一筆房屋交時的資料。每筆紀錄共有 28 個屬性，包括房屋的 ID 號碼、1 個目標屬性與 26 個條件屬性(8 個類別屬性，18 個數值屬性)。以下將介紹每個欄位的名稱、意義及屬性值。

屬性 1：[ID](數值屬性)：此為每一筆 ID 的編號。

屬性 2：[成交價](目標屬性)：房屋成交的價格。

屬性 3：[面積(坪)_(主建物加公設)](數值屬性)：主建物 + 公共建設的坪數。

屬性 4：[面積(坪)_(主建物車位加公設)](數值屬性)：主建物 + 車位 + 公共建設的坪數。

屬性 5：[公設比](數值屬性)：房屋權狀坪數與不屬於房屋室內實際坪數，二者之間的比值。

屬性 6：[路寬](數值屬性)：馬路面寬。

屬性 7：[屋齡(年)](數值屬性)：房屋的年齡。

屬性 8：[樓層](數值屬性)：房屋所在的樓層。

屬性 9：[當月縣市_戶口數](數值屬性)：房屋成交時，當月的縣市戶口數。

屬性 10：[當月縣市_流入人口](數值屬性)：房屋成交時，當月的縣市流入人口數。

屬性 11：[當月縣市_流出口](數值屬性)：房屋成交時，當月的縣市流出口數。

屬性 12：[當月縣市建照數](數值屬性)：房屋成交時，當月的縣市核發的建照數。

屬性 13：[當月縣市使用執照數](數值屬性)：房屋成交時，當月的縣市使用執照數。

屬性 14：[房價指數_月成長率](數值屬性)：房屋成交時，當月的房價指數成長率(月)。

屬性 15：[建築生產指數_月成長率](數值屬性)：房屋成交時，當月的建築生產指數成長率(月)。

屬性 16：[購屋貸款利率](數值屬性)：房屋成交時，購屋貸款利率。

屬性 17：[台股指數_月成長率](數值屬性)：房屋成交時，當月的台股指數成長率(月)。

屬性 18：[所得成長率](數值屬性)：房屋成交時，當月的所得成長率(月)。

表二、各數值屬性欄位統計資訊

屬性	最小值	50%	最大值	平均值	標準差
成交價(萬元)	190	445	1100	479.763	195.8
面積(坪)(主建物加公設)	9.39	32.119	134.34	32.486	9.72
面積(坪)(主建物車位加公設)	10.58	33.49	134.34	35.318	11.971
公設比	0	0.16	0.589	0.142	0.103
路寬	2	6	50	8.139	5.325
屋齡(年)	0.02	6.15	38.349	9.315	8.924
樓層	1	5	36	6.817	5.361
當月縣市_戶口數	132237	1186498	1222079	1118465.048	270266.507
當月縣市_流入人口	1686	19471	25351	18902.051	5250.23
當月縣市_流出人口	1635	17838	24900	17879.994	5035.175
當月縣市建照數	3	48	86	47.052	16.628
當月縣市使用執照數	3	43	75	42.081	13.509
房價指數_月成長率	-2.217	-0.219	3.139	-0.025	1.488
建築生產指數_月成長率	-61.94	4.477	96.651	3.319	29.547
購屋貸款利率	2.48	3.686	4.962	3.704	0.81
台股指數_月成長率	-12.028	-2.509	12.638	-0.227	7.107
所得成長率	-0.54	4.25	7.63	3.992	2.124
土地(坪)	0.11	6.07	1323.66	7.747	26.792
公告現值	1	173480.98	581947.13	173941.514	84727.643

屬性 19：[土地(坪)](數值屬性)：房屋的土地坪數。

屬性 20：[公告現值](數值屬性)：房屋成交時，當月的公告現值。

屬性 21：[車位](類別屬性)：

有：有附車位。

無：沒附車位。

屬性 22：[風險](類別屬性)：這是由房屋仲介業者自行的判斷。

低：房屋屬於低風險。

中：房屋屬於中風險。

高：房屋屬於高風險。

屬性 23：[樓層屬性](類別屬性)：房屋的屬性。

頂樓：房屋位於頂樓。

透天：房屋屬於整棟。

一樓：房屋位於一樓。

其他：房屋未於其他樓層。

屬性 24：[房屋型態](類別屬性)：房

屋屬於種型態：別墅、公寓、住辦、

華廈、辦公室、店面、透天、國宅、套房、工宅。

屬性 25：[建材](類別屬性)：房屋所使用的建材：鋼骨鋼筋混凝土、鋼骨結構、加強磚造、鋼筋混凝土。

屬性 26：[用途](類別屬性)：房屋的用途：商業、店面、住家、工業、國宅、其他、住商。

屬性 27：[分區](類別屬性)：房屋屬於哪一區：住宅區、工業區、商業區、住商。

屬性 28：[鄉鎮區](類別屬性)：將台北縣劃分為 25 區而此房屋落在哪一區。

表一為屬性空值分析表，可以檢視資料庫的資料品質。可以看出只有鄉鎮區此屬性有 604 筆記錄有空值，在這裡我們是將它填入[未知]。

表三、數值屬性欄位比較

屬性	最小值	50%	最大值	平均值	標準差
面積(坪)_(主建物加公設)	9.39	32.119	134.34	32.486	9.72
	2.239	3.469	4.9	3.436	0.303
面積(坪)_(主建物車位加公設)	10.58	33.49	134.34	35.318	11.971
	2.358	3.511	4.9	3.507	0.343
路寬	2	6	50	8.139	5.325
	0.693	1.791	3.912	1.948	0.517
當月縣市_戶口數	132237	1186498	1222079	1118465.048	270266.507
	11.792	13.986	14.016	13.838	0.556
當月縣市_流入人口	1686	19471	25351	18902.051	5250.23
	7.43	9.876	10.14	9.753	0.555
當月縣市_流出人口	1635	17838	24900	17879.994	5035.175
	7.399	9.789	10.122	9.7	0.542
土地(坪)	0.11	6.07	1323.66	7.747	26.792
	-2.207	1.803	7.188	1.732	0.679
公告現值	1	173480.98	581947.13	173941.514	84727.643
	0	12.063	13.274	11.895	0.74

表一、屬性空值分析表

屬 性	空值個數	空值所佔的比例
ID	0	0.00%
成交價	0	0.00%
面積(坪)_(主建物加公設)	0	0.00%
面積(坪)_(主建物車位加公設)	0	0.00%
公設比	0	0.00%
路寬	0	0.00%
屋齡(年)	0	0.00%
樓層	0	0.00%
當月縣市_戶口數	0	0.00%
當月縣市_流入人口	0	0.00%
當月縣市_流出人口	0	0.00%
當月縣市建照數	0	0.00%
當月縣市使用執照數	0	0.00%
房價指數_月成長率	0	0.00%
建築生產指數_月成長率	0	0.00%
購屋貸款利率	0	0.00%
台股指數_月成長率	0	0.00%
所得成長率	0	0.00%
土地(坪)	0	0.00%
公告現值	0	0.00%
車位	0	0.00%
風險	0	0.00%
樓層屬性	0	0.00%
房屋型態	0	0.00%
建材	0	0.00%
用途	0	0.00%
分區	0	0.00%
鄉鎮區	604	12.57%

3.1 資料處理

我們將資料庫切割成 80% 的訓練資料 (Training Data) 以及 20% 的測試資料 (Testing Data)，而訓練資料有 3817 筆、測試資料有 987 筆。資料的品質對於探勘的結果有著極重大的影響，如果進件資料當中存在著許多“雜訊”，則探勘的結果將會變的沒有參考價值；因此，透過數值資料分佈，以瞭解數值資料的分佈及品質，並進行後續的處理。表二為數值屬性欄位的統計資訊。

由表二可以得知，面積(坪)_(主建物加公設)、面積(坪)_(主建物車位加公設)、路寬、當月縣市_戶口數、當月縣市_流入人口、當月縣市_流出人口、土地(坪)、公告現值等六個屬性，它們的最大值或最小值超出平均值加減三個標準差，為了使資料更接近常態分布所以我們將這六個欄位的數值做取 LOG 的動作，並將取過 LOG 的資料作為一個新的訓練資料，屆時與原訓練資料做出的結果做比較。而表三則是原始的統計與取 LOG 之後(每個屬性第二列的值)的統計比較，可以看出取 LOG 後的資料較接近於常態分布。

3. 研究方法

表四、數值屬性重要程度

欄位名稱	importance	欄位名稱	importance
面積(坪)_(主建物車位加公設)	0.733	當月縣市使用執照數	0.167
面積(坪)_(主建物加公設)	0.688	當月縣市建照數	0.161
當月縣市_戶口數	0.235	所得成長率	0.048
路寬	0.228	土地(坪)	0.045
樓層	0.225	購屋貸款利率	0.021
公告現值	0.219	房價指數_月成長率	0.001
公設比	0.218	建築生產指數_月成長率	-0.019
當月縣市_流入人口	0.2	台股指數_月成長率	-0.035
當月縣市_流出人口	0.192	屋齡(年)	-0.393

3.2 重要屬性程度分析

每個條件屬性對於目標屬性會有不一樣的重要程度，所以本研究會對所有屬性作分析，選擇重要且具有區分能力的屬性作為分類模型的輸入。並非所有的屬性對於分類模型皆有幫助，若將不重要的屬性輸入至分類模型中，不僅會拉長模型的學習時間，也會降低分類系統的正確率。挑選重要的屬性進入分類模型，除了可以避免不重要的屬性加入分類模型，導致降低模型的分類結果。也可避免系統處理不重要的屬性而浪費學習時間，可加速探勘的速度。傳統上挑選重要屬性的方式，是利用人的直覺與經驗來篩選，往往缺乏科學數據支持。於本研究中，挑選重要屬性的計算方式，包括數值型屬性(Numerical Attribute)與類別型屬性(Categorical Attribute)的挑選。以下將詳述屬性重要程度之計算方式。首先介紹挑選數值型重要屬性的計算方法。

3.2.1 數值屬性重要性分析

對於數值屬性，我們採用皮爾森相關係數(Pearson Correlation Coefficient)進行評估。其公式如下：

$$\frac{\sum_{i=1}^n (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\left(\sum_{i=1}^n (A_i - \bar{A})^2\right) \left(\sum_{i=1}^n (B_i - \bar{B})^2\right)}} \quad (1)$$

其中 A 是數值屬性的欄位；B 則是目標屬性(成

交價)。利用此公式我們可以計算出所有數值屬性的重要程度，其結果按照重要性高低列在表四。而表五則是比較取 LOG 前後重要程度的變化。

表五、重要程度變化

欄位名稱	importance	
	Normal	Log
面積(坪)_(主建物加公設)	0.688	0.674
面積(坪)_(主建物車位加公設)	0.733	0.715
路寬	0.228	0.194
當月縣市_戶口數	0.235	0.178
當月縣市_流入人口	0.2	0.2
當月縣市_流出人口	0.192	0.206
土地(坪)	0.045	0.083
公告現值	0.219	0.186

3.2.2 類別屬性重要性分析

而類別屬性的計算方式，本研究提出一種計算方式，首先，先計算成交價在此類別欄位所有屬性的平均值與標準差，接著再利用公式(2)算出重要程度。

$$\text{類別屬性重要程度} = \frac{|\max(\text{平均成交價}) - \min(\text{平均成交價})|}{\text{avg}(\text{成交價標準差})} \quad (2)$$

表六、風險欄位所有屬性值的平均值及標準差

	高	中	低
標準差	137.3968	189.2987	197.6039
平均值	362.2495	461.0101	489.0106

我們以風險的欄位為例，表六是如此欄位所有屬

性值的平均值 (Mean) 及標準差 (Standard Deviation)。依據公式我們可以算出

$$importance = \frac{|489.0106 - 362.2495|}{(137.3968 + 189.2987 + 197.6039)/3} = 0.725$$

依此類推我們算出所有類別欄位的重要程度，其結果如表七。

表七、所有欄位的重要程度

欄位名稱	importance
房屋型態	2.987
鄉鎮區	2.374
用途	2.266
建材	1.396
分區	1.132
樓層屬性	1.079
車位	1.063
風險	0.725

4. 實驗與效能評估

4.1 建立模型

本研究會挑選不同門檻值的重要程度，並分別建立決策樹、類神經網路、線性迴歸等三種模型進行研究，而且我們還會使用工具所建議之欄位結果做比較。圖一是工具對原資料所建議之欄位，圖二則是工具對取 LOG 後的資料所建議之欄位。

4.2 效能評估

將測試資料餵入訓練資料所建模型，可以得到預測的成交價。本研究對於模型評估方法是計算每筆預測值與實際值差多少的平均，公式為：

$$AVG(abs(預測值 - 實際值)) \quad (3)$$

依據公式(3)我們可以評估所有模型的結果。本研究對於不同門檻值的重要程度進行訓練，並找出在何門檻值可以得到最佳結果，接著也對測試資料進行不同門檻值的欄位餵入模型。表

八為決策樹模型的結果，從表中我們可以得知在原資料門檻值設定為 0.2 會得到較好的結果，而在測試資料也是在門檻值設定為 0.2 時有較準確的結果；在取 LOG 資料門檻值設定為 0.05 會得到較好的結果，而在測試資料也是在門檻值設定為 0.05 時有較準確的結果。

表九為線性迴歸模型的結果，從表中我們可以得知在原資料門檻值設定為 0.05 會得到較好的結果，而在測試資料也是在門檻值設定為 0.05 時有較準確的結果；在取 LOG 資料門檻值設定為 0.05 會得到較好的結果，而在測試資料也是在門檻值設定為 0.05 時有較準確的結果。

表十為類神經網路模型的結果，從表中我們可以得知在原資料門檻值設定為 0.05 會得到較好的結果，而在測試資料也是在門檻值設定為 0.05 時有較準確的結果；在取 LOG 資料門檻值設定為 0.05 會得到較好的結果，而在測試資料也是在門檻值設定為 0.05 時有較準確的結果。

5. 結論與未來工作

在決策樹、線性迴歸、類神經網路這三種模型中，決策樹所預測的房價和實際房價平均相差最少，而在原資料中門檻值設 0.2 有準確的結果，在測試資料進行預測時也是得到在門檻值 0.2 時有較好的效能；取 LOG 的資料在門檻值 0.05 時有最好的結果，可是在測試資料所得的結果卻比原資料的結果差，而且在所有模型中取 LOG 的結果皆遜於原資料的結果，從這可以得知在此資料取 LOG 使資料較為接近常態分部並沒有較好的效能。本研究提出的屬性重要性分析的方法，根據不同門檻值所挑選出來的屬性，其實驗結果皆優於 SQL Server 2005 所建議之屬性的結果。在本實驗中設定挑選數值屬性與類別屬性的門檻值一樣，然而數值屬性及類別屬性計算重要程度的方式不同，未來工作可以嘗試在挑選屬性時，數

Column Name	Score	Input
當月縣市_流出口	0.293	x
當月縣市_戶口數	0.293	x
當月縣市_流入人口	0.293	x
台股指數_月成長率	0.253	x
購屋貸款利率	0.253	x
建築生產指數_月成長率	0.253	x
所得成長率	0.248	x
當月縣市建照數	0.247	x
當月縣市使用執照數	0.245	x
樓層	0.219	x
鄉鎮區	0.183	x
路寬	0.162	x
房價指數_月成長率	0.131	x
房屋型態	0.062	x
車位	0.046	
樓層屬性	0.041	
風險	0.029	
建材	0.029	
分區	0.029	
用途	0.025	
公告現值		
土地(坪)		
屋齡(年)		

圖一、工具對原資料所建議之欄位

Column Name	Score	Input
當月縣市_流出口	0.290	x
當月縣市_流入人口	0.290	x
當月縣市_戶口數	0.290	x
台股指數_月成長率	0.253	x
購屋貸款利率	0.253	x
建築生產指數_月成長率	0.253	x
所得成長率	0.247	x
當月縣市建照數	0.245	x
當月縣市使用執照數	0.243	x
樓層	0.221	x
鄉鎮區	0.185	x
路寬	0.162	x
房價指數_月成長率	0.131	x
房屋型態	0.066	x
車位	0.047	
樓層屬性	0.042	
建材	0.032	
分區	0.031	
風險	0.031	
用途	0.027	
公告現值		
土地(坪)		
屋齡(年)		

圖二、工具對取 LOG 後的資料所建議之欄位

表八、決策樹模型的結果

決策樹

訓練

門檻	0.05	0.1	0.15	0.2	0.25	SQL
Log	46	49	49	54	62	129
Normal	52	52	52	50	58	129

測試

門檻	0.05	0.1	0.15	0.2	0.25	SQL
Log	64	64	64	64	64	135
Normal	66	66	66	59	63	135

表九、線性迴歸模型的結果

線性迴歸

訓練

門檻	0.05	0.1	0.15	0.2	0.25	SQL
Log	84	84	84	88	104	142
Normal	79	79	79	79	99	143

測試

門檻	0.05	0.1	0.15	0.2	0.25	SQL
Log	86	86	86	89	105	145
Normal	82	82	82	82	100	146

表十、類神經網路模型的結果

類神經

訓練

門檻	0.05	0.1	0.15	0.2	0.25	SQL
Log	66	66	66	66	69	121
Normal	62	62	62	65	67	121

測試

門檻	0.05	0.1	0.15	0.2	0.25	SQL
Log	68	67	67	67	70	125
Normal	65	65	65	66	68	125

值屬性與類別屬性的門檻值可設為不一樣，以驗證門檻值的不同對系統效能之影響為何。

參考文獻

1. Han, J. and Kamber, M. (2000), Data Mining - Concepts and Techniques, Morgan Kaufmann Publishers, 2000.
2. Lee, Y. S. and Yen, S. J. (2002), "Neural-Based Approaches for Improving the Accuracy of Decision Trees," Lecture Notes in Computer Science (LNCS): Data Warehousing and Knowledge Discovery, 2454, September 2002, pp. 114-123.
3. Lee, Y. S. and Yen, S. J. (2004), "Classification Based on Attribute Dependency," Lecture Notes in Computer Science (LNCS): Data Warehousing and Knowledge Discovery, 3181, September 2004, pp. 259-268.
4. Lee, Y. S., et al. (2004), "A Data Mining Approach to Constructing Credit Risk Scoring Model," Proceedings of 10th Conference on Information Management and Practice, 2004, pp. 1799-1813.
5. Quinlan, J. R. (1996), "Improved Use of Continuous Attributes in C4.5," Journal of Artificial Intelligence Approach, 4, 1996, pp. 77-90.
6. Wang, K., Zhou, S. Q. and He, Y. (2000), "Growing Decision Trees on Support-Less Association Rules," Proceedings of International Conference on Knowledge Discovery and Data Mining, 2000, pp. 265-269.