

# 데이터 마이닝을 이용한 아파트 초기계약 예측모형 개발: 위례 신도시 미분양 아파트 단지를 사례로

김지영<sup>1</sup>, 이상경<sup>2\*</sup>

<sup>1</sup>하남도시공사 대리, <sup>2</sup>가천대학교 도시계획학과 교수

## Development of Forecasting Model for the Initial Sale of Apartment Using Data Mining: The Case of Unsold Apartment Complex in Wirye New Town

Ji Young Kim<sup>1</sup>, Sang-Kyeong Lee<sup>2\*</sup>

<sup>1</sup>Deputy Manager, Hanam Urban Innovation Corporation

<sup>2</sup>Professor, Dept. of Urban Planning, Gachon University

요 약 이 연구에서는 미분양 아파트 단지의 세대별 계약 자료에 데이터 마이닝 기법인 의사결정나무, 신경망, 로지스틱 모형을 적용하여 세대별 초기계약을 예측하는 모형을 개발한다. 모형 개발에는 위례신도시 미분양 아파트 단지의 계약 자료가 이용되며, 이 자료는 훈련용 자료와 검증용 자료로 분할되어 분석에 투입된다. 훈련용 자료에서는 신경망, 의사결정나무, 로지스틱 모형 순으로 예측력이 뛰어났지만 검증용 자료에서는 로지스틱 모형이 가장 우수하게 나타났다. 이 같은 결과는 신경망이 훈련용 자료에 최적화된 모형으로 구축되면서 검증용 자료에 대한 적응성이 떨어져 나타난 결과로 판단된다. 의사결정나무와 로지스틱 모형을 병행 적용한 결과, 층수, 향, 세대 위치, 전기 및 발전기실의 소음, 청약자 거주지, 청약 종류가 초기계약에 영향을 주는 것으로 나타났다. 이는 두 가지 모형을 같이 사용하는 것이 초기계약 결정요인 발굴에 더 효과적이라는 것을 의미한다. 이 연구는 데이터 마이닝의 적용 범위를 주택 분양 예측까지 확장함으로써 융복합 분야 발전에 기여하고 있다.

주제어 : 미분양 아파트, 초기계약, 데이터 마이닝, 의사결정나무, 신경망, 로지스틱 모형

**Abstract** This paper aims at applying the data mining such as decision tree, neural network, and logistic regression to an unsold apartment complex in Wirye new town and developing the model forecasting the result of initial sale contract by house unit. Raw data are divided into training data and test data. The order of predictability in training data is neural network, decision tree, and logistic regression. On the contrary, the results of test data show that logistic regression is the best model. This means that logistic regression has more data adaptability than neural network which is developed as the model optimized for training data. Determinants of initial sale are the location of floor, direction, the location of unit, the proximity of electricity and generator room, subscriber's residential region and the type of subscription. This suggests that using two models together is more effective in exploring determinants of initial sales. This paper contributes to the development of convergence field by expanding the scope of data mining.

**Key Words** : Unsold apartment, Initial sale, Data mining, Decision tree, Neural network, Logistic regression

This paper is based on Gachon University master's thesis of Ji Young Kim.

\*Corresponding Author : Sang-Kyeong Lee(skylee@gachon.ac.kr)

Received October 23, 2018

Revised November 29, 2018

Accepted December 20, 2018

Published December 28, 2018

## 1. 서론

아파트 미분양은 사업시행자의 자금조달을 어렵게 하고 입주 예정자들에게도 가격 하락에 대한 두려움을 준다는 점에서 주택시장에서 관심의 대상이 되고 있다. 아파트 청약률이 1대 1이 되지 않는 경우 미분양 발생은 당연하지만 1대 1을 넘더라도 배정된 동·호수에 만족하지 않는 당첨자들이 있을 경우 계약포기와 함께 미분양 발생으로 이어질 수 있다. 초기계약은 최초 분양단계에서 아파트사업 시행자와 당첨자가 맺는 분양계약을 의미하며, 당첨자들이 초기계약을 포기할 경우 해당 세대는 미분양으로 남게 된다. 초기계약의 중요성 때문에 이에 대한 관심은 많지만 기업의 영업활동과 관련된 사항이기 때문에 관련 자료가 공개되지는 않고 있으며, 그 결과 학술차원의 연구도 거의 이루어지지 않고 있다. 다만, 세대별 초기계약이 아닌 아파트 단지별 초기계약률을 분석한 연구는 일부 있는데[1-3], 이 연구들은 거시경제요인과 마케팅 활동, 단지의 입지요인이 아파트 단지의 초기계약률에 미치는 영향을 분석하고 있다. 아파트 세대별 초기계약 결정요인은 거의 연구되지 않고 있다는 점에 대해 연구의 필요성이 제기되고 있다.

아파트 분양 당첨자의 초기계약 결정은 다양한 요인에 의해 영향을 받기 때문에 계약 결과를 사전에 예측한다는 것은 쉽지 않은 일이 된다. 이는 연구자가 세대별로 초기계약 결과를 예측하는 모형을 개발할 경우에도 마주치는 문제가 된다. 초기계약 여부를 이진변수로 표현하고 이를 종속변수로 설정할 경우 로지스틱 회귀모형과 같은 통계학적 모형의 적용을 우선 고려할 수 있다. 하지만 이 경우 회귀분석의 원리를 따라 독립변수들의 선형성에 입각하여 모형을 설정해야하는 관계로 비선형적 관계를 보이는 변수들을 모형에서 배제시키는 결과를 초래할 수도 있다. 이를 극복하기 위해 다양한 예측모형의 적용을 고려할 수 있는데 이 경우 자료 탐색, 변수 선택, 모형 설정, 모형 평가로 이어지는 일련의 작업이 모든 모형들에 동일하게 적용되어야 하는 과제를 안게 된다. 이 같은 작업은 반자동 또는 자동화된 시스템을 필요로 하게 되는데 데이터 마이닝(data mining)은 이를 충족시킬 수 있는 효과적인 수단이 될 수 있다.

데이터 마이닝은 대용량의 데이터로부터 가치 있는 정보를 추출하는 것으로 자동화된 기법을 활용하여 유의미한 경향과 규칙을 발견하기 위해 탐색하고 분석하는

과정을 지칭한다[4]. 데이터 마이닝은 1990년대 컴퓨터 성능 향상과 네트워크 기술이 발전하면서 기업, 학교, 정부기관 등에서 대용량의 데이터를 보관할 수 있게 되었고 데이터가 가지고 있는 유용한 정보를 활용하는 것에 대한 관심이 높아지면서 처음 대두되었다[5]. 이후 보험사와 이동통신사, 신용카드사와 같이 고객에 대한 인구통계학적 데이터와 거래기록을 대규모로 확보하고 있는 기업들을 중심으로 사용범위가 확대되었으며[5], 최근 들어서는 빅데이터를 보유하고 있는 인터넷 기업들을 중심으로 활용범위를 넓혀가고 있다. 데이터 마이닝을 이용한 학술적 연구는 통계학, 관공학, 경영학 등 다양한 분야에서 활발하게 이루어지고 있으며[6-8], 부동산 분야에서도 가계소득 분석[9], 리츠 투자특성 연구[10], 주택가격 예측모형 개발[11-13], 오피스 가격 및 투자분석[14-16] 등 다양한 연구가 시도되고 있다. 하지만 아직까지 아파트 미분양을 대상으로 데이터 마이닝을 적용한 사례는 없는 것으로 확인되고 있다.

이 같은 인식하에, 본 연구에서는 미분양 아파트 사례 단지의 세대별 계약 자료에 데이터 마이닝을 적용하여 세대별 초기계약 예측모형을 개발하고자 한다. 사례단지는 하남도시공사가 위례신도시에 분양한 A아파트단지로 분양 당시 당첨자들의 계약포기로 미분양 세대가 발생한 경우에 해당한다. 초기계약 예측모형 개발은 데이터 마이닝 기법 중 의사결정나무와 신경망, 로지스틱 모형을 이용한다. 세대별 초기계약 결과가 이진형의 명목척도 형태로 종속변수에 투입되는 관계로 로지스틱 모형 중에서 이항 로지스틱 모형을 분석에 이용한다. 본 연구는 기존의 거시적 관점의 연구와 내용과 방법론 측면에서 차별화된 연구를 지향한다. 내용적으로 미분양 아파트의 초기계약 결정요인을 실증적으로 분석하게 되며 데이터 마이닝 기법을 아파트 미분양 분야에 처음으로 도입하게 된다.

## 2. 선행연구 고찰

### 2.1 아파트 초기계약 관련 연구 고찰

분양아파트 초기계약에 관한 실증적 연구는 자료 취득의 어려움 등으로 활성화되지 않고 있으며 학술적 논문도 소수에 그치고 있다. 분양아파트 단지들의 초기계약률 자료를 이용하여 초계약률에 영향을 미치는 요인

들을 분석한 연구들과[1-3] 부산의 분양 아파트 단지를 사례로 미분양 세대의 시장잔류시간에 영향을 주는 요인을 분석한 연구가[17] 그나마 본 연구에 시사점을 제공할 수 있는 연구라고 할 수 있다.

선행연구를 보면, 단지별 초기계약률은 단지특성변수 중 도급순위, 총세대수, 구조, 난방방식, 분양가와 기존 아파트 가격과의 차이, 할인점과의 거리, 공원과의 거리에 영향을 받으며 거시환경변수 중에서는 주택담보대출 금리, 주택부문 소비자전망지수의 영향을 받는 것으로 나타나고 있다[1]. 초기계약률은 단지특성변수와 거시경제변수 외에도 전단광고, 신문광고 등의 마케팅 변수에 영향을 받는 것으로 나타나고 있다[2]. 한편, 아파트 청약률과 초기계약률간의 차이는 단지규모, 대중교통 이용 가능성, 전세 증가율, 미분양 세대수, 지역별 경제활동 인구수, 매수우위지수 등에 영향을 받는 것으로 나타나고 있다[3]. 선행 연구결과를 종합해 보면, 아파트 단지의 초기계약률은 입지 요인과 거시경제 요인, 마케팅 요인, 지역 주택시장 요인 등 다양한 요인들에 의해 영향을 받는다. 이들 연구들은 단지별 초기계약률을 분석대상으로 했기 때문에 세대별 초기계약 여부를 분석대상으로 하는 본 연구와는 분석단위에서 그 틀을 달리한다. 이 점에서 부산시 미분양 아파트 단지를 사례로 미분양 세대의 시장잔류시간에 영향을 주는 요인을 분석한 [17]이 본 연구에 더 많은 시사점을 줄 수 있다. [17]에서는 분석방법으로 생존분석(survival analysis)을 이용하였으며 분양시장의 선호특성과 전매시장의 선호특성을 비교하였다. 분양시장 분석에서는 종속변수로 분양계약까지 걸린 시간을, 전매시장 분석에서는 분양이후 전매까지 걸린 시간을 이용하였다. 분석 결과, 공급자에 의해 가격이 결정되는 분양시장에서는 아파트의 구조특성인 평형, 층수, 조망의 영향이 강하지만 전매시장에서는 이 같은 특성들이 가격에 반영되면서 이들 변수의 영향력이 상대적으로 줄어드는 것으로 나타났다. 이 연구에서는 이들 변수 외에도 세대배치, 계약자 연령, 성별 등이 독립변수로 사용되었는데, 본 연구에서도 이를 참조하여 이들 변수들을 예측모형에 투입하는 독립변수를 설정하고자 한다.

## 2.2 데이터 마이닝 관련 연구 고찰

본 연구가 주택을 대상으로 하는 관계로 주택 분야에 데이터 마이닝을 적용한 경우를 중심으로 선행연구를 살펴보고자 한다. 먼저, 데이터 마이닝 기법을 이용하여 임

차인의 임대계약 선택을 분석한 연구[18]에서는 주택임대 계약형태를 종속변수로 하고 가구주 직업, 주택 형태, 1인당 주거면적, 가구주연령, 1인 가구, 금융자산, 부동산 자산, 월 소득, 가구주 교육수준을 독립변수로 하였다. 로지스틱 모형과 의사결정나무모형, 신경망 모형을 이용하여 분석을 수행한 결과 세 가지 모형의 예측력이 비슷한 수준으로 나타났다. 이 연구에서는 임대 시장의 복잡성을 감안하여 특정 모형을 선택하기보다 분석 모형의 장점들을 적절히 활용하는 것이 연구의 질을 향상시킬 수 있다고 주장하였다. 데이터 마이닝을 적용하여 아파트 가격을 예측한 연구[11]에서는 아파트 가격을 종속변수로 하고 독립변수로 아파트 평형, 방수, 욕실 수, 아파트 수명, 가구당 주차 수, 역까지의 거리, 공원까지의 거리, 아파트 브랜드 점수 등을 변수로 설정하였다. 회귀분석이 가장 예측력이 좋은 모형으로 나타났지만 의사결정나무모형을 통해 강남권 아파트의 재건축 기대심리를 유의한 변수로 찾아냈다는 점을 고려하여 세 모형의 장점들을 살리는 것이 연구의 질을 향상시킬 수 있다고 주장하였다. 이와 유사한 결과는 데이터 마이닝을 아파트 실거래가격에 적용한 연구[12]에서도 나타나고 있는데, 이 연구에서는 독립변수로 아파트면적, 아파트 층수, 건축연도, 시청 직선거리, 2차선 직선거리, 4차선 직선거리, 지하철역 직선거리, 공원 직선거리, 인구밀도로 지역은 서울시 강남권, 강서권, 도심권, 강북권을 설정하였으며, 의사결정나무분석, 다중회귀분석, 신경망분석을 이용해 실증분석을 시도하였다. 분석결과, 신경망 분석이 가장 우수한 것으로 나타났지만 세 모형의 장점을 살릴 경우 연구 내용이 풍부해진다는 점을 강조하였다. 이들 연구들은 회귀분석과 같은 특정 모형이 우수 모형이 선택될 수 있지만 연구의 질 향상 등을 고려하여 세 가지 모형 모두를 병행 사용하는 것이 필요하다는 주장을 펴고 있다. 이는 세 가지 모형을 보완적으로 사용할 경우 유의한 변수를 더 많이 찾아낼 수 있다는 사실에 근거하는 것으로 본 연구에서도 세 가지 모형의 병행 사용을 통해 유의한 변수들을 더 많이 찾아내고자 한다.

## 3. 연구방법

### 3.1 의사결정나무 분석

의사결정나무는 자료의 변수들을 분석하여 이들 사이

의 패턴과 분류별 특성을 분류하여 나무형태로 만들게 된다[5]. 의사결정나무 분석에서는 CART(classification and regression trees) 알고리즘과 CHAID(chi-squared automatic interaction detection) 알고리즘이 주로 사용되는데, 본 연구에서는 목표변수가 범주형인 점을 감안하여 카이제곱통계량을 이용하는 CHAID를 적용한다. CHAID는 다른 알고리즘과 다르게 과다적합이 일어나기 전에 나무를 키워 나가는 것을 중지하며 범주형 변수에만 국한되어 사용된다는 점이 다른 알고리즘과 차별화된다. CHAID 분류과정에서는 각 입력변수가 목표 집단에서 최대한 가능한 계층을 나타내도록 검정을 통해 묶어 준다. 카이제곱 검정을 통해 가장 큰 차이를 보여주는 분할을 현재마디에서 분할자로 선택하며, 더 이상 분할이 없을 때까지 이 과정을 통해 나무를 키워나가게 된다[22]. 본 연구에서는 의사결정나무 모형 구축을 위해 유의수준, 최대가지 수, 최대깊이, 최소 범주형 크기, 분리 정확도, 리프크기, 규칙개수의 분류기준 등을 반복 설정하며 RMSE(root mean square error)가 더 이상 개선되지 않을 때까지 작업을 수행한다.

### 3.2 신경망 분석

신경망은 신경 생리학 분야에서 인간 뇌의 활동을 이해하고자 하는 시도에서 출발하여 생물학적인 프로세스를 컴퓨터를 이용하여 모형화한 것이다. 인간의 뇌를 흉내 내어 실제 자신이 가진 자료를 반복 학습의 과정을 거쳐 숨겨진 패턴을 찾는 기법으로 다양한 분야에서 활용되고 있다[19-21]. 신경망의 가장 중요한 세 가지 특징이 있는데 첫 번째로 수학적 모형이 필요 없다는 것이고 둘째로 자료의 해석보다는 예측에 더 유용하다는 점이고 마지막으로 자료의 크기가 커야 한다는 점이다[5].

본 연구에서는 다층인식자 신경망을 분석에 이용한다. 다층인식자는 입력층, 은닉층, 그리고 출력층으로 구성된 신경망이다. 입력층은 각 입력변수에 대응되는 마디로 구성되어 여러 개의 은닉층의 은닉마디로 구성된다. 각 은닉마디는 입력층으로부터 변수들이 선형결합을 비선형결합으로 처리하여 출력층 또는 다른 은닉층으로 전달하고 출력층은 목표변수에 대응하는 마디를 갖게 되어 여러 개의 목표변수를 있는 경우 여러 개의 출력마디들이 존재하게 된다[5]. 본 연구에서는 은닉층을 점차 늘리는 방법으로 반복학습을 실시하는 방법으로 모형의 예측력을 향상시키게 된다. 예측력 향상은 RMSE를 기준으

로 확인하며 은닉층 노드수를 점차 늘려도 RMSE가 더 이상 개선되지 않을 때까지 무한반복학습을 하게 한다.

### 3.3 로지스틱 회귀분석

종속변수가 범주형 변수일 때는 선형회귀모형을 대신하여 로지스틱회귀모형이 일반적으로 사용된다. 특히, 종속변수가 본 연구처럼 초기계약이 체결될 경우와 안 될 경우로 구분되는 이진형 더미변수일 경우에는 이항 로지스틱 회귀모형이 분석에 사용된다. 이 모형의 종속변수  $Y_i$ 에는  $i$ 세대가 초기에 계약될 경우와 계약되지 않을 경우로 구분하는 더미변수 값이 입력되는데 계약될 경우를 1, 계약되지 않을 경우를 0으로 가정할 경우 독립변수  $x_i$ 에 대해  $Y_i$ 가 1이 될 확률  $P(Y_i = 1|x_i)$ 는 승산(odds)  $\Omega(x_i)$ 를 구성하는 식으로 (1)식과 같이 정의될 수 있다.

$$\Omega(x_i) = \frac{P(Y_i = 1|x_i)}{1 - P(Y_i = 1|x_i)} \quad (1)$$

승산  $\Omega(x_i)$ 을 로짓변환(logit transformation)하게 되면 로짓(logit)이 (2)식과 같이 도출되게 된다.

$$\log \Omega(x_i) = \log \left[ \frac{P(Y_i = 1|x_i)}{1 - P(Y_i = 1|x_i)} \right] = x_i' \beta \quad (2)$$

(2)식은 독립변수들이 선형인 관계로 회귀분석이 가능하며, 회귀계수는 최대우도법(Maximum likelihood estimation)을 통해 추정된다.

## 4. 분석의 틀

### 4.1 데이터 마이닝 적용 절차

데이터 마이닝은 일반적으로 자료추출(sampling), 자료탐색(explore), 수정(modify), 모형설정(model), 모형평가(assess)로 구성되는 SEMMA 5단계 과정을 통해 결과를 도출하게 된다[23]. 본 연구에서는 SAS E-Miner 프로그램을 이용하여 Fig. 1의 흐름도(flow chart)에서 보이는 SEMMA 5단계 과정을 수행한다. 첫 번째 자료추출 단계는 흐름도의 Data partition에 해당하며 여기서는 분양아파트의 초기계약 전체자료를 훈련용 자료(training data) 70%와 검증용 자료(test data) 30%로 분할한다. 두 번째 자료탐색 단계는 흐름도의 Stat explore에 해당하며

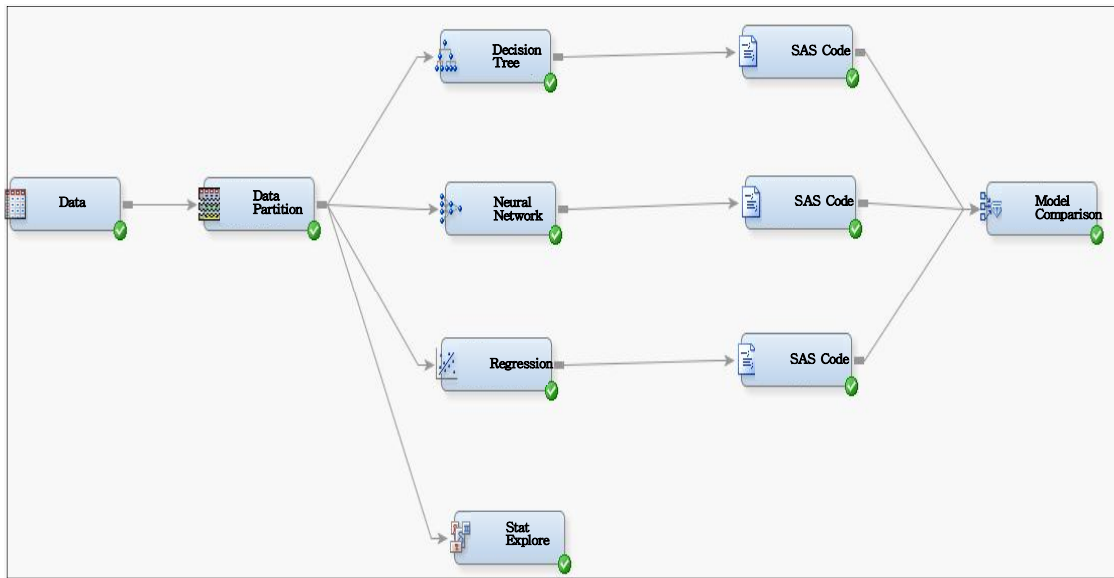


Fig. 1. SAS E-Miner flow chart

빈도분석과 카이제곱통계량을 이용하여 분양계약에 영향을 미치는 변수들을 탐색한다. 세 번째 수정 단계는 변수를 분석목적에 맞게 변환하고 결측치를 정리하는 단계인데, 본 연구의 경우 변수들이 이미 분석에서 요구하는 명목변수로 변환되어 있고 결측치가 없는 관계로 이 단계는 생략한다. 네 번째 모형설정 단계에서는 SAS E-Miner에서 제공하는 의사결정나무와 신경망, 로지스틱 회귀모형을 선택한 후 훈련용 자료에 이를 적용하여 모형을 구축하게 된다. 다섯 번째 모형평가 단계는 흐름도상의 Model comparison에 해당되며, 훈련용 자료를 통해 구축된 최종모형들을 검정용 자료를 통해 비교하게 된다.

#### 4.2 분석 자료

분석 자료는 2013년 하남시 위례신도시에서 보급자리 주택으로 공급된 A단지의 계약 자료이다. 이 단지를 사례로 선택한 이유는 현실적으로 다른 미분양 아파트 단지의 초기계약 자료를 확보하는 것이 어렵다는 점과 연구자들이 사업상황을 잘 알고 있어 자료처리와 해석 과정에서 오류를 범할 가능성이 적다는 점에 있다.

이 사업은 하남도시공사에서 자체 시행한 사업으로 민간건설사에 도급으로 시공을 담당하게 했다. 단지규모를 보면, 대지면적이 81,102㎡, 연면적이 260,146.17㎡, 최고층수는 27층, 세대수는 1,673세대이다. Table 1은 사례

Table 1. Composition of subscription

Type	Balcony type	Total units	Number of supplied units		
			Special supply	General supply	Additional supply
75A	Expansion	92	32	60	
75B	Expansion	292	84	208	
75A	Non-expansion	54	6	48	
84A	Expansion	699	176	523	
84B	Expansion	482	114	325	43
84A	Non-expansion	54	5	31	18
Total		1,673	417	1,195	61

Table 2. Contract rate of initial sale

Supply Type	Number of subscribers	Number of contractors	Contract rate (%)
Total	1,673	1,065	63.66
Special supply	417	303	72.66
General supply	1,195	717	60.00
Additional supply	61	45	73.77

단지에서 공급되는 세대를 유형별로 정리한 것이다. 세대평면은 판상형(A)과 탑상형(B), 전용면적 75㎡와 84㎡의 조합으로 4개 타입(75A, 75B, 84A, 84B)으로 구분되지만 판상형의 경우는 다시 발코니 확장형과 비확장형으로 구분되기 때문에 모두 6개 타입을 가지게 된다. 공급 세대수는 84A 판상형 발코니 확장형이 41.8%로 가장 많고 75A와 84A 판상형 발코니 비확장형이 3.2%로 가장 적다.

청약은 특별공급(신혼부부, 노부모, 생애 최초, 3자녀

이상, 기관 추천)을 먼저 하였는데, 여기서 미달 세대수가 발생하여 이를 일반공급으로 전환하여 공급하였다. 특별공급과 일반공급에서 무자격자 61세대가 당첨되었는데 이는 잔여세대 모집공고를 통해 청약자들에게 다시 공급되었다. 분양 당시 청약경쟁률은 100%를 넘었지만, 주택경기 침체가 장기화되면서 위례신도시라는 입지적 장점과 보금자리주택이라는 가격 장점에도 불구하고 계약포기가 속출하면서 608세대가 계약이 되지 않은 채로 남게 되었다. Table 2를 보면, 전체 계약률은 63.66%이며 계약률은 잔여세대공급, 특별공급, 일반공급 순으로 나타나고 있다.

### 4.3 변수 설정

Table 3은 데이터 마이닝에 사용되는 종속변수와 독립변수를 정리한 것이다. 종속변수는 초기계약 체결 세대와 미계약 세대를 구분하는 이진형 명목척도 값을 갖게 된다. ‘주택공급에 관한 규칙’에서는 분양 아파트 계약을 입주자로 선정된 날로부터 5일 경과 후 3일 이상의 기간을 정하여 해약 한다고 규정하고 있다. 본 연구에서는 이 기간 동안 계약을 체결한 세대만을 초기계약 체결 세대로 본다.

독립변수는[17], [24-26]에서 타당하게 나타난 요인들

중에서 사례단지의 물리적, 입지적 특성에 고려하여 도입한다. [17]에 따르면, 미분양 세대의 시장잔류시간은 평형, 층수, 세대배치, 조망과 같은 구조특성과 계약자 연령, 거주지 같은 인적 특성에 영향을 받는다. 이 같은 변수들은 본 연구의 초기계약 결정 과정에서도 영향을 줄 가능성이 높다는 점에서 구조특성과 인적 특성으로 도입한다.

[24-26]은 주택가격을 다룬 연구이지만 주택에 대한 소비자들의 선호가 주택가격에 반영된다는 점에서 본 연구에 시사점을 줄 수 있다. [24]에서는 골프장 조망이 좋은 세대일수록 주택가격이 높게 형성되고 있다는 것을 보여주고 있는데, 사례단지가 골프장 조망을 특징으로 하고 있다는 점에서 조망 여부가 계약 결과에도 영향을 미칠 것으로 보고 도입을 결정한다. [25]을 보면, 초등학교와 생활편의시설이 멀어질수록 주택가격이 떨어지고 있다. 이를 고려하여 본 연구에서도 접근성 특성으로 편의시설 접근성, 도로 접근성, 초등학교 거리를 도입한다. [26]에 따르면, 쓰레기 소각장과 같은 유해시설은 주택가격에 부의 영향을 준다. 사례단지에도 거주자들이 싫어하는 환경적으로 유해한 요인들이 있고 이들 요인들이 계약체결에 영향을 준다고 보고 독립변수로 도입한다.

구체적으로, 구조특성은 전용면적, 발코니 확장, 층수,

Table 3. Variables for data mining

Variables		Description	Criteria	
Dependent variable		ct	Whether to conclude a sale contract	Contract, Non-contract
Independent variables	Structural characteristics	ha	Exclusive area of unit	84㎡, 75㎡
		type	Whether to expand a balcony	Non-expansion, Expansion
		f	Location of floor	Base, Bottom, Top floor
		gl	Golf course view	Non-view, View
		hl	Location of unit	Center, East, West
		d	Direction	Southeast, Southwest, South
	Accessibility characteristics	da	Distance to facilities	0~125m, 125~250m, 250m~
		dl	Road accessibility	Roadside, Central location
		sd	Distance to elementary school	0~125m, 125~250m, 250m~
	Environmental characteristics	doeffect1	Lights coming out of car and noise effect	Unaffected, Affected
		doeffect2	Pump and water tank room influence	Unaffected, Affected
		doeffect3	Electric and generator room noise and soot	Unaffected, Affected
		doeffect4	Heat exchange chamber	Unaffected, Affected
		doeffect5	Infringement of the privacy of public facilities	Unaffected, Affected
		doeffect6	Just infringing on the privacy of the mall	Unaffected, Affected
		doeffect7	Outdoor noise	Unaffected, Affected
		doeffect8	Public toilet noise and odor	Unaffected, Affected
		dofillot	Low-rise pilots	Unaffected, Affected
doyuli	Model house and different grill window	Unaffected, Affected		
Human characteristics	ad	Subscriber's residential region	Gyeonggi-do, Seoul, Hanam, Others	
	ag	Subscriber's age	20's, 30's, 40's, 50's, 60 old or older	
	k	Type of subscription	General supply, Additional supply, Special supply	

골프장 조망, 세대 위치, 향으로 구성한다. 전용면적 변수(ha)는 84m<sup>2</sup>, 75m<sup>2</sup>로, 발코니 확장 변수(type)는 발코니 비확장과 확장으로, 층 변수(f)는 저층, 기준층, 최상층으로, 골프장 조망 변수(gl)는 골프장 조망과 비조망으로, 세대 위치 변수(hl)는 동쪽위치, 서쪽위치, 중앙위치로, 향 변수(d)는 남동향, 남향, 남서향으로 구분된다.

접근성 특성은 편의시설, 도로, 초등학교 접근성 변수로 구성된다. 편의시설 접근성 더미변수(da)는 0~125m, 125~250m 미만, 250~400m로 구분되며, 도로 접근성 더미변수(dl)는 도로변 위치, 중앙 위치로 구분되며, 초등학교 접근성 더미변수(sd)는 0~125m, 125~250m, 250~400m로 구분된다.

환경특성은 입주자 모집공고에서 공개된 소음과 매연 등의 부정적 요인을 검토하여 변수화한 것으로 총 10개의 환경요인이 독립변수로 선정된다.

인적특성은 청약자를 기준으로 청약자 거주지 변수(ad), 연령 변수(ag), 청약종류 변수(k)로 구성한다. 청약종류 변수는 Table 2의 특별공급, 일반공급, 잔여세대 공급으로 구분된다. 사례단지는 '보급자리주택건설 등에 관한 특별법'에 따라 서울시, 인천시, 경기도에 거주하는 무주택 세대주를 대상으로 주택을 공급하며, 하남시 30%, 경기도 20%, 수도권 50% 비율만큼 해당지역에 우선 공급한다. 청약자 거주지 변수는 이를 참조하여 하남시, 서

울시, 하남시 제외 경기도, 기타지역을 구분하는 더미변수로 설정한다.

## 5. 분석 결과

### 5.1 의사결정나무 분석 결과

의사결정나무 모형을 적용하여 반복적 작업을 수행한 결과 유의수준 10%, 최대가지 수 5, 최대깊이 10, 최소 범주형 크기 5, 분리정확도 4, 리프크기 5, 규칙개수 5에서 더 이상 모형이 개선되지 않았다. Table 4에서 이때의 RMSE를 보면 0.465로 나타나고 있다

Table 5의 실제치와 예측치를 비교한 결과를 보면, 정분류율은 66.52%, 오분류율은 33.48%이다. 미계약 세대

Table 4. Statistics of decision tree analysis

Obs	MSE	RMSE
1171	0.21579	0.464532

Table 5. Classification rate of decision tree analysis

	Non-contracted (%)	Contracted (%)
Non-contracted (%)	4.78	31.60
Contracted (%)	1.88	61.74

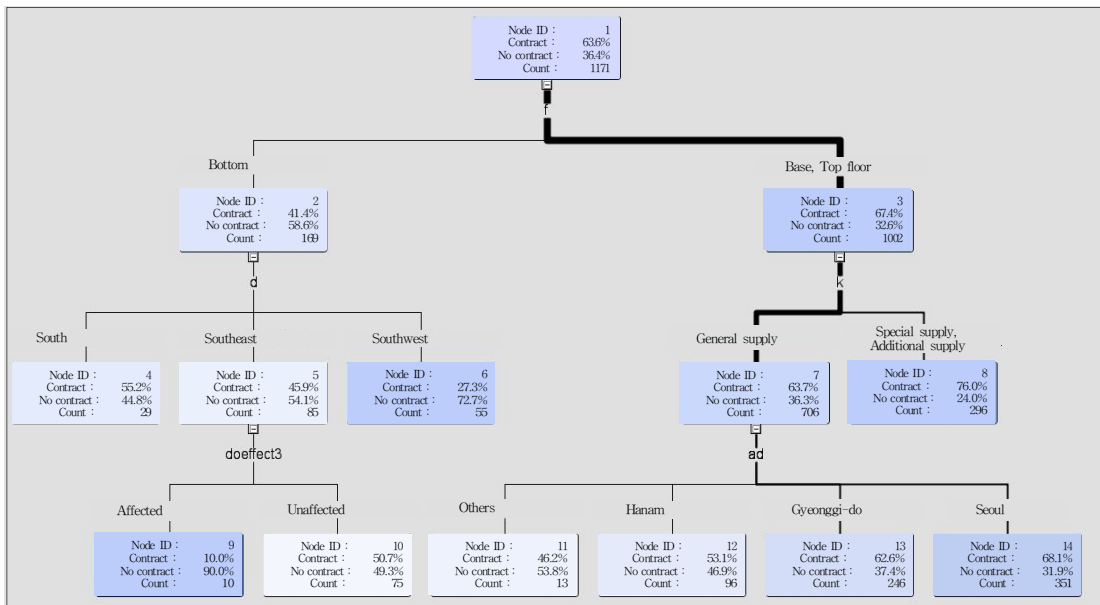


Fig. 2. Results of decision tree analysis

를 계약 세대로 잘못 인식하는 경우는 31.60%로 상당히 높게 나타나고 있다.

Fig. 2에서 10% 유의수준에서 유의한 변수들을 살펴 보면 층수가 가장 중요한 변수가 된다. 분양 아파트가 저층인 경우에는 향, 전기·발전기실 소음 및 매연이라는 환경요인이, 분양 아파트가 기준층 또는 최고층인 경우에는 청약종류, 청약지역이 분양계약 체결에 영향을 미치는 변수로 확인되었다. 이에 기초하여 초기계약 체결 시 당첨자들이 고려하는 요인을 해석하면 다음과 같다.

첫째, 당첨자들은 아파트 계약 시 층수를 가장 중요한 요인으로 고려한다. 고층은 아파트 조망 측면에서 선호되고 있고 층수가 높을수록 가격이 높게 형성된다. 결과적으로 기준층과 최고층의 계약률이 저층 세대보다 더 높게 나타났다.

둘째, 저층 당첨자들은 남서향에 비해 남향과 남동향 계약을 선호하는 것으로 나타났다. 저층 계약자들은 저층의 단점인 일조량 부족을 극복하길 원하고 이 때문에 향에 민감하게 반응하는 것으로 보인다.

셋째, 사례단지의 남동향 쪽에는 전기·발전기실이 배치되어 있어 소음과 매연이 발생할 수 있는데, 남동향 세대 중 여기에 영향을 받는 세대의 경우 계약률이 떨어지는 것으로 나타났다.

넷째, 청약종류가 계약체결에 영향을 주는 것으로 나타났다. 특별공급 당첨자의 경우 평생 1번 분양받을 권리를 사용한 관계로 계약을 포기하게 되면 다시는 특별공급을 받을 수 없게 되고 따라서 일반공급 당첨자에 비해 계약을 더 많이 체결하는 것으로 나타났다.

다섯째, 청약자의 거주지가 초기계약에 영향을 주는 것으로 확인되었다. 보급자리주택의 경우 지역우선 공급이 법적으로 정해져 있어 사례단지의 경우도 당해지역인 하남시 청약자들에게 30%, 경기도 청약자들에게 20%를 우선 배정하고 나머지 서울·인천·수도권 청약자들에게 50%를 배정하게 된다. 하남시 청약자가 당첨될 가능성이 가장 높지만 사례단지의 분양가가 하남시 관내 아파트들보다 높아 하남시 당첨자들의 계약결정에 부정적 영향을 줄 가능성이 높다. 반면, 서울시 청약자의 경우 사례단지의 분양가를 낮은 것으로 평가할 가능성이 크고 소득이 상대적으로 높아 긍정적 영향을 받게 될 가능성이 높다. 결과적으로 서울시 당첨자들의 계약률이 가장 높게 나타났고 하남시 당첨자의 계약률이 가장 낮게 나타났다.

### 5.2 신경망 분석 결과

신경망 분석은 SAS E-Miner를 이용하여 프로그램에서 지원하는 최대반복수 1,000회, 최대시간 4시간을 기준으로 히든 노드수를 점차 늘리는 반복 학습을 적용하였다. Table 6은 히든 노드 수 17에서 더 이상 RMSE의 개선이 일어나지 않는 것을 보여주고 있다. Table 7을 보면 이때의 RMSE는 0.245이며 모형은 총 33개의 입력층과 17개의 은닉층을 가지게 된다. Fig. 3은 이를 다층 퍼셉트론 구조로 도식화한 결과이다.

Table 6. RMSE by the number of hidden nodes

Number of hidden nodes	RMSE	Number of hidden nodes	RMSE
1	0.297182	15	0.109308
2	0.294620	16	0.105892
3	0.270709	17	0.100769
4	0.218617	18	0.102477
5	0.195559	19	0.101623
6	0.204099	20	0.101623
7	0.234842	21	0.100769
8	0.222032	22	0.100769
9	0.149445	23	0.103330
10	0.164816	24	0.100769
11	0.133222	25	0.100769
12	0.140051	26	0.100769
13	0.117848	27	0.100769
14	0.111870		

Table 7. Statistics of neural network analysis

Obs	MSE	RMSE
1171	0.060056	0.245063

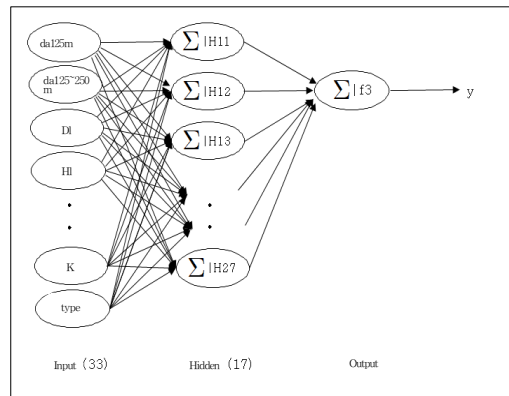


Fig. 3. Structure of multi-layered neural network



Table 8. Classification rate of Neural network analysis

	Non-contracted (%)	Contracted (%)
Non-contracted (%)	29.80	6.58
Contracted (%)	5.29	58.33

Table 8에서 신경망분석의 실제치와 예측치를 비교한 결과를 보면, 정분류율은 88.13%, 오분류율은 11.87%로 나타났다. 실제 계약되지 않은 세대를 계약으로 잘못 예측하는 경우는 상당히 낮은 수준인 6.58%로 나타났다. 이를 통해 신경망을 이용한 반복훈련을 통해 상당히 우수한 모형이 개발되었음을 알 수 있다.

5.3 로지스틱 회귀분석 결과

Table 9의 로지스틱 회귀분석 결과를 보면, RMSE는 0.471로 나타나고 있다. Table 10에서 실제치와 예측치를

Table 9. Statistics of logistic regression analysis

Obs	MSE	RMSE
1171	0.219661	0.470686

Table 10. Classification rate of logistic regression

	Non-contracted (%)	Contracted (%)
Non-contracted (%)	6.83	29.55
Contracted (%)	4.18	59.44

비교한 결과를 보면, 정분류율은 66.27%, 오분류율은 33.73%로 나타나고 있다. 실제 계약되지 않은 세대를 계약으로 잘못 예측하는 경우는 신경망분석 결과보다 높은 29.55%로 나타나고 있다.

Table 11의 로지스틱 회귀분석 결과를 보면, 구조특성에서는 층수와 세대 위치 변수가 유의한 것으로 나타났

Table 11. Results of logistic regression analysis

Variables	Description	Criteria	Coefficient	Std error	$\chi^2$	$Pr>\chi^2$	
Constant			0.4494	0.7833	0.33	0.5661	
Structural Characteristics	ha	Exclusive area of unit	75㎡	0.1829	0.1119	2.67	0.1021
	type	Whether to extend a balcony	Non-expansion	-0.1396	0.1761	0.63	0.4279
	f	Location of floor	Base	0.4233	0.1255	11.37	0.0007
			Bottom	-0.5761	0.1643	12.3	0.0005
	gl	Golf course view	Non-view	-0.1916	0.1249	2.35	0.125
	hl	Location of unit	Center	0.2405	0.1387	3.01	0.0829
East			-0.0289	0.1764	0.03	0.8697	
d	Direction	Southeast	-0.0833	0.1318	0.4	0.5272	
		Southwest	-0.1217	0.1653	0.54	0.4615	
Accessibility characteristics	da	Distance to facilities	0~125m	0.2193	0.1514	2.1	0.1473
			125~250m	0.0244	0.2263	0.01	0.9141
	dl	Road accessibility	Roadside	0.0425	0.1777	0.06	0.8111
	sd	Distance to elementary school	0~125m	0.1418	0.2118	0.45	0.5033
125~250m			0.1038	0.1421	0.53	0.4651	
Environmental Characteristics	doeffect1	Lights coming out of car and noise effect	Unaffected	-0.0926	0.1927	0.23	0.6309
	doeffect2	Pump and water tank room influence	Unaffected	0.2201	0.2059	1.14	0.2851
	doeffect3	Electric and generator room noise and soot	Unaffected	0.1023	0.1378	0.55	0.4578
	doeffect4	Heat exchange chamber	Unaffected	-0.1554	0.1199	1.68	0.1949
	doeffect5	Infringement of the privacy of public facilities	Unaffected	0.1904	0.3398	0.31	0.5752
	doeffect6	Just infringing on the privacy of the mall	Unaffected	0.2956	0.3331	0.79	0.3748
	doeffect7	Outdoor noise	Unaffected	-0.0571	0.1618	0.12	0.7242
	doeffect8	Public toilet noise and odor	Unaffected	-0.3618	0.26	1.94	0.164
	dofillot	Low-rise pilots	Unaffected	-0.2716	0.203	1.79	0.181
doyuli	Model house and different grill window	Unaffected	-0.00521	0.217	0.00	0.9808	
Human Characteristics	ad	Subscriber's residential region	Gyeonggi-do	0.1363	0.1429	0.91	0.3402
			Others	-0.3235	0.3302	0.96	0.3273
			Seoul	0.3239	0.1402	5.34	0.0209
	ag	Subscriber's age	20's	0.2256	0.2333	0.93	0.3336
			30's	0.0541	0.1136	0.23	0.6338
			40's	0.0548	0.1291	0.18	0.6714
			50's	-0.046	0.1644	0.08	0.7797
k	Type of subscription	General supply	-0.4075	0.1483	7.55	0.006	
		Additional supply	0.1968	0.2715	0.53	0.4686	

다. 층수는 의사결정나무 분석에서도 유의했던 변수로 로지스틱 회귀분석에서도 1% 유의수준에서 유의한 것으로 나타났다. 결과를 보면, 최상층에 비해 기준층 계약이 더 잘 체결되는 데 반해 저층은 계약이 잘 체결되는 않는 것으로 나타났다. 세대 위치는 10% 유의수준에서 유의하며, 세대위치가 서쪽인 경우보다 중앙인 경우가 계약체결이 잘 되었다. 이는 중앙에 위치한 세대가 측면에 위치한 세대보다 향이 좋고 난방비 측면에서 더 유리하기 때문으로 보인다.

접근성 특성 변수들은 모두 유의하지 않는 것으로 나타났다. 환경특성의 경우 모집공고에 적혀 있는 넘비시 설들이 계약체결 결정에 악영향을 줄 것이라는 예상과 달리 모두 유의하지 않는 것으로 나타났다. 이는 시행사가 단지의 단점을 입주자 모집공고에 작게 표현함으로써 청약자들이 관심 있게 보지 않은 결과인 것으로 추정된다. 이는 시행사 입장에서는 마케팅의 성공이라고 할 수 있지만 청약자 입장에서는 정보가 제대로 전달되지 않아 발생한 일종의 정보비대칭 현상이 될 수 있다. 실제로 당첨자들이 입주자 모집공고에 적혀 있는 사항을 숙지하지 않은 채 계약을 체결하고 입주 시점에서 시행사에 항의를 하거나 소송을 진행하는 경우를 자주 보고 있다.

인적특성 변수의 경우 청약연령은 유의하지 않았으나 청약자 거주지와 청약종류에서 유의한 결과가 나타났다. 청약자 거주지 변수의 경우 서울시 청약자가 하남시 청약자보다 계약을 체결할 가능성이 더 높은 것으로 나타났다. 하지만 의사결정나무에서 유의한 차이를 보여주었던 경기도 거주자의 경우 하남시 거주자와 유의한 차이를 보이지 못하였다. 의사결정나무 분석에서 보이듯이, 청약종류 변수와 관련하여 특별공급 당첨자에 비해 일반공급 당첨자들의 계약체결 확률이 더 낮게 나타났다.

5.4 최종모형 비교

Table 12는 훈련용 자료 1,171개를 이용하여 최종 개발된 의사결정나무, 신경망, 로지스틱 모형을 검정용 자료 502개에 적용하여 그 결과를 비교한 것이다. RMSE를 통해 예측력을 비교해 보면, 신경망 모형 0.245, 의사결정나무 모형 0.465, 이항 로지스틱 회귀모형 0.471 순으로 신경망 모형이 가장 우수한 것으로 나타났다. 이들 최종 모형들을 검정용 자료에 적용하고 RMSE를 통해 그 결과를 비교해 보면, 로지스틱 회귀모형 0.221, 의사결정나무 모형 0.227, 신경망 모형 0.345로 로지스틱 회귀모형의

Table 12. Comparison of final models

Models	Data	Obs	RMSE	Classification rate(%)	Misclassification rate(%)
Decision tree	training	1,171	0.464532	66.52	33.48
	test	502	0.226941	64.54	35.46
Neural network	training	1,171	0.245063	88.13	11.87
	test	502	0.344829	60.56	39.44
Logistic regression	training	1,171	0.470686	66.27	33.73
	test	502	0.220897	67.13	32.87

가장 우수한 것으로 나타났다. 정분류율을 보면, 훈련용 자료에서는 신경망 모형 88%, 의사결정나무 67%, 로지스틱 모형 66% 순으로 신경망 모형이 가장 우수한 것으로 나타나고 있다. 하지만 검정용 자료에서는 로지스틱 67%, 의사결정나무 65%, 신경망 모형 61% 순으로 로지스틱 모형이 가장 우수하게 나타나고 있다.

RMSE를 보면 의사결정나무와 로지스틱 모형의 경우 훈련용 자료에 비해 검정용 자료에서 더 양호한 결과를 보이는 데 반해 신경망의 경우 반대로 모형의 적합성이 떨어지고 있다. 정분류율에서도 의사결정나무와 로지스틱 모형의 경우 훈련용과 검정용 자료의 차이가 크지 않지만 신경망의 경우 차이가 많이 나고 있다. 이 같은 결과는 신경망 모형이 훈련용 자료에는 최적화되었지만 검정용 자료에는 최적화되지 않았다는 것을 의미한다. 데이터 마이닝 절차에 따라 전체 자료는 훈련용 자료 70%와 검정용 자료 30%로 분할되어 분석에 이용되었으며, 무작위로 추출된 관계로 이질성이 강한 자료라고 할 수는 없다. 신경망 모형이 훈련용 자료의 관측치에 민감하게 반응하면서 모형을 구축하다보니 검정용 자료의 작은 변화에도 적합도가 떨어지게 된 것으로 보여진다. 이번 연구 결과에 한정해서 본다면 로지스틱 모형의 활용성이 신경망보다 더 우수하다고 할 수 있다.

Table 13을 통해 모형별 초기계약 결정요인을 비교해 보면, 층수, 청약자 거주지, 청약종류 변수의 경우는 의사

Table 13. Comparison of determinants by model

Decision tree	Logistic regression
<ul style="list-style-type: none"> <li>• Location of floor</li> <li>• Subscriber's residential region</li> <li>• Type of subscription</li> <li>• Direction</li> <li>• Electric and generator room noise and soot</li> </ul>	<ul style="list-style-type: none"> <li>• Location of floor</li> <li>• Subscriber's residential region</li> <li>• Type of subscription</li> <li>• Location of unit</li> </ul>

결정나무와 로지스틱 모형에서 모두에서 유의하며, 향, 전기·발전기실의 소음 변수는 의사결정나무 모형에서만, 세대 위치 변수는 로지스틱 모형에서만 유의하게 나타났다. 이 같은 결과는 특정 모형만 사용하는 것 보다 두 가지 모형을 병행 사용하는 것이 초기계약 결정요인 분석에 유리하다는 것을 시사한다.

## 6. 결론

본 연구에서는 하남시 위례신도시 미분양 아파트 단지의 세대별 계약 자료에 데이터 마이닝 기법인 의사결정나무와 신경망, 로지스틱 회귀모형을 적용하여 세대별 초기계약 예측모형을 개발하였다. 계약 자료는 훈련용 70%와 검정용 30%로 분할되어 분석에 이용된다. 연구 결과를 정리하면 다음과 같다.

첫째, 훈련용 자료를 대상으로 모형을 개발한 결과, RMSE와 정분류율 모두에서 신경망, 의사결정나무, 로지스틱 모형 순으로 예측력이 떨어지는 것으로 나타났다. 그러나 최종모형을 검정용 자료에 적용한 결과에서는 반대로 로지스틱 모형, 의사결정나무, 신경망 순으로 예측력이 떨어지는 것으로 나타났다. 신경망 모형의 경우 훈련용 자료에 최적화된 모형을 만들어 낼 수 있지만 검정용 자료에 이를 적용할 경우 예측력이 떨어지는 것으로 나타났다. 이는 로지스틱 모형의 활용성이 신경망 모형보다 더 우수하다는 것을 의미한다. 그러나 이 같은 결과는 이번 연구에 한정된 것으로 훈련용과 검정용 자료의 추출을 달리하거나 다른 미분양 아파트 단지 자료를 분석에 이용할 경우 그 결과가 달라질 수 있다는 점에서 절대적인 의미를 부여할 수는 없다.

둘째, 의사결정나무 분석과 로지스틱 회귀분석을 통해 분양아파트의 초기계약에 영향을 미치는 구조특성과 환경특성, 인적특성을 찾아낼 수 있었다. 의사결정나무 분석을 통해 층수, 향, 전기·발전기실의 소음, 청약자 거주지, 청약종류가 유의하게 나타났다. 또한 로지스틱 회귀분석을 통해 층수, 세대위치, 청약자 거주지, 청약종류가 유의하게 나타났다. 층수, 청약자 거주지, 청약종류처럼 두 모형에서 공통으로 유의한 요인도 있지만 향, 전기·발전기실의 소음, 세대 위치 변수처럼 한 모형에서만 유의한 경우가 있어 두 모형을 같이 사용하는 것이 초기계약 결정요인 분석에 유리함을 알 수 있다. 이 같은 결과는 특정 모형만 사용하는 것보다 두 가지 모형을 병행 사용

하는 것이 초기계약 결정요인 발굴에 더 효과적임을 의미한다.

본 연구는 내용과 방법론 양쪽 모두에서 학술적 기여를 하고 있다. 내용 측면에서 보면, 거시적 관점의 기존 연구와 달리 미시적 관점에서 사례연구를 했다는 점에서 차별화가 된다. 기존의 연구들이 거시적 차원에서 단지 단위의 초기계약을 결정요인을 다루는 연구라면 본 연구는 미시적 관점에서 미분양 아파트 단지를 사례로 세대별 초기계약 결정요인을 분석하는 연구이다. 방법론적 측면에서 본 연구는 데이터 마이닝의 적용 범위를 주택 분양 예측까지 확장함으로써 융복합 분야의 발전에 기여하고 있다. 아파트 단지의 세대별 초기계약 예측 모형 개발은 기존의 데이터 마이닝 연구에서 다루어 적지 않은 주제로 본 연구를 통해 처음으로 적용 가능성을 확인하였다.

주택 미분양 문제는 사회적 관심이 크기 때문에 신문과 방송 등에서 많이 다루어지고 있지만 대부분 거시적 차원에서 미분양 주택 수 증가를 경제 여건 하에서 진단하는 수준에 그치고 있다. 본 연구는 미시적 연구라는 점에서 이와는 맥락을 달리한다. 아파트 미분양이 사회적 문제라는 점에서 세대별 초기계약 결과를 예측할 수 있는 모형을 개발한다는 것은 사전에 미분양에 대한 대비를 가능하게 한다는 점에서 그만큼 사회적 비용을 줄여주는 효과가 있다. 더불어 이 연구 결과는 주택사업자들의 분양 마케팅을 지원할 수 있다는 점에서 실용적 차원의 의미를 가진다.

본 연구는 하남시 위례신도시 미분양 단지라는 특정 사례를 분석대상으로 했다는 점에서 한계를 가진다. 다른 분양 단지에 적용할 수 있는 일반화된 모형을 개발하기 위해서는 미분양 단지 자료의 추가 확보가 필요하다. 이 경우 단지의 위치와 주변 환경 같은 단지의 입지특성을 독립변수에 추가할 수 있기 때문에 모형의 예측력을 향상시킬 수 있다. 후속 연구에서는 방법론 측면의 개선도 필요한데 본 연구에서 사용하지 않은 앙상블(ensemble) 기법인 배깅(bagging)과 부스팅(boosting)의 사용도 고려할 필요가 있다.

## REFERENCES

- [1] M. S. Baik & J. C. Shin. (2011). A Study on the Determinants of Initial Sales Rate for New Apartment Housing. *Journal of the Korean Urban Management*

- Association*, 24(1), 213-237.
- [2] M. S. Baik & J. C. Shin. (2011). A Study on the Condominium Sales Marketing Activities and Initial Sales Rate. *Journal of the Korea Real Estate Analysts Association*, 17(3), 25-43.
- [3] H. S. Kwon & D. W. Bang. (2015). A Study on the Cause of Difference between New Apartment Subscription Rate and Initial Pre-sale Contract Rate. *Housing Studies Review*, 23(3), 111-143.
- [4] G. S. Linoff & M. J. Berry. (2018). *Data Mining Techniques: For Marketing, Sales, and Customer Relation management*. Seoul : Hankyngsa.
- [5] M. H. Huh & Y. G. Lee. (2008). *Data Mining Modeling and Cases*. Seoul : Hannarae Publishing Co.
- [6] Y. H. Kim & S. S. Ahn. (2006). A Study on the Characteristics of Fast-food Restaurant's Customers. *Tourism & Leisure Research*, 18(2), 191-209.
- [7] W. J. Kim, Y. S. Choi & D. H. Yoo. (2018). Development of Win-Loss Prediction Models and Strategies for Improving Winning Rate of the Korean Professional Baseball Teams Using Data Mining Techniques. *Korea Journal of Sport Management*, 23(3), 88-104.
- [8] J. Y. Oh & S. H. Choi. (2018). An Analysis of the Characteristics of Companies introducing Smart Factory System Using Data Mining Technique. *Journal of the Korea Convergence Society*, 9(5), 179-189.
- [9] H. J. Chun. (2015). A Study on Korean Household Income Using Data Mining. *Journal of the Korea Planning Association*, 50(2), 227-241.
- [10] K. M. Kim & C. K. Kim. (2017). Forecasting of Investment Characteristics of Global REITs Using Data Mining. *Korea Real Estate Review*, 68, 44-56
- [11] J. Y. Lee, M. H. Choi & S. Y. Lee. (2007). A Study on the Forecasting Model of Apartment Price Based on Data Mining. *Journal of the Korea Planning Association*, 42(4), 135-148.
- [12] H. J. Chun. (2017). A Study on the Determinants of Housing Price Using Data Mining. *Residential Environment: Journal of The Residential Environment Institute of Korea*, 15(3), 35-46.
- [13] B. C. Kim, Y. Kim, M. Kim & J. S. Lee. (2018). Using Data Mining Techniques to Model Housing Rental Price near Universities in Seoul. *Journal of the Korean Institute of Industrial Engineers*, 44(4), 259-271.
- [14] A. R. Hong, J. P. Ko & S. J. Yoo. (2010). A Study on the Forecasting Model of the Investment Characteristics of Seoul Office Buildings based on Data Mining. *Seoul Studies*, 11(2), 51-68.
- [15] K. S. Mun, J. G. Choi & H. S. Lee. (2015). An Analysis for Price Determinants of Small and Medium-Sized Office Buildings Using Data Mining Method in Gangnam-Gu. *International Journal of Contents*, 15(7), 414-427.
- [16] P. W. Huh, S. Y. Kim, Y. S. Hong & G. E. Shim. (2014). A Study on the Determinants of Office Building Property Management Method in Seoul. *Seoul Studies*, 15(3), 41-57.
- [17] H. S. Lee. (2004). A Study on Preference Characteristics for Each Condominium in a Same Site on Initial and Re-sales Markets Using Survival Analysis. *Journal of the Korea Planning Association*, 39(3), 81-93.
- [18] T. Y. Kim & C. M. Lee. (2005, November). Comparative Study on Renter's Choice with Data Mining Techniques. *2005 Spring Congress of Korea Planning Association*.
- [19] H. Byeon. (2017). Exploring Influence Factors for Peer Attachment in Korean Youth Based on Multi-Layer Perceptron Artificial Neural Networks. *Journal of the Korea Convergence Society*, 8(10), 209-214.
- [20] T. S. Ki & S. H. Lee. (2017). A Prediction Scheme for Power Apparatus using Artificial Neural Networks. *Journal of Convergence for Information Technology*, 7(6), 201-207.
- [21] H. Yoon. (2018). Classificatin of Normal and Abnormal Heart Sounds Using Neural Network. *Journal of Convergence for Information Technology*, 8(5), 131-135.
- [22] S. Y. Lee. (2003). *A Study on Data Application Using Data Mining*. Master dissertation. Yonsei University, Seoul.
- [23] B. S. Kim, W. S. Bae, K. H. Seok, D. H. Cho. & K. L. Choi. (2018). *SAS EM 14.1*. Seoul : Kyowoo.
- [24] J. H. Kim. (2008). The Effect of the Landscape Visibility of a Golf Course on Apartment Price: A Case Study of the Residential Area around Hansung C.C. in Yong-in City. *Journal of The Korean Regional Development Association*, 20(4), 69-88.
- [25] H. G. Sung & J. Y. Kim. (2011). The Impacts of Time-Varying Accessibility of Facilities on Housing Price Change by the Modified Repeat Sales Model - The Case of Subway Line 9 in Seoul. *Journal of the Korean Society of Civil Engineers*, D31(3D), 477-487.
- [26] M. K. Oh & J. H. Cho. (2018). Newly Improved Incineration Plant's Impacts on Nearby Apartment Sale Prices with Interrupted Time Series Analysis. *Journal of Korea Planning Association*, 53(3), 145-159.

김 지 영(Kim, Ji Young)

[정회원]



- 2010년 2월 : 가천대학교 도시계획학과(공학사)
- 2016년 9월 : 가천대학교 환경대학원 도시계획전공(공학석사)
- 2018년 3월 ~ 현재 : 가천대학교 대학원 도시계획학과 박사과정

- 2011년 7월 ~ 현재 : 하남도시공사 대리
- 관심분야 : 도시계획, 도시개발, 주택, 주거복지
- E-Mail : orgin1004@naver.com

이 상 경(Lee, Sang Kyeong)

[정회원]



- 1990년 2월 : 서울대학교 도시공학전공(공학사)
- 1992년 2월 : 서울대학교 도시공학전공(공학석사)
- 1996년 2월 : 서울대학교 도시공학전공(공학박사)

- 2007년 3월 ~ 현재 : 가천대학교 도시계획학과 교수
- 관심분야 : 도시계획 및 개발, 도시정책, 주택 및 부동산시장
- E-Mail : skyllee@gachon.ac.kr