

<양식 2>

처리부서	LINC+사업단
------	----------

## LINC+사업단 캡스톤디자인 지원비 지급 신청서

과 제 명	머신러닝을 활용한 서울특별시 아파트 매매가 예측	팀 명	금의한양
팀 장 명	윤종훈	연락처	010-7338-2209

※노란 셀 (): 자동 계산(입력금지)



순번	지원항목	거래(물품)명	금액 (부가세포함)	예금주명	은행명	계좌번호
1				윤종훈	IBK기 업은행	
2						
3						
4						
합 계			0원			

### <작성요령>

\* 지원항목 : ①재료비 ②문헌/문구/인쇄 및 복사비/설문조사비 ③전시회참관비 중 해당번호 기입

#### \* 영수증 1장 금액당 1줄 입력

ex) 영수증 1장 금액이 10,000원이고 2건의 물품을 구매한 것이면 거래명에 "대표물품명 외 1건", 금액에 10,000원 기재

\* A4 사이즈보다 작은 문서는 A4 용지에 서로 겹치지 않도록 폴로 부착(스테이플러 사용 금지)

\* 본 페이지가 1페이지를 초과하지 않도록 유의

\* 줄 삭제 혹은 추가 가능(합계란은 삭제 금지)

\* 수정테이프 사용불가

### <첨부서류>

\* 필수 제출 : (순서대로) 영수증(현금영수증 또는 전자세금계산서), 거래명세서, 통장사본, 물품 사진

\* 해당시 제출 :

-전자세금계산서(영수)인 경우(선결제한 경우) : 업체로 대금 이체한 증빙서류, 업체 통장사본

-<양식2-1>구매신청사유서 : 문헌/문구/인쇄 및 복사, 설문조사(기념품) 구매 시 제출

-<양식2-2>설문조사 결과보고서 : 첨부-설문조사지 양식 1부

-<양식2-3>전시회참관 보고서 : 첨부-전시회 원본 티켓

위와 같이 신청합니다.

2020.06.04.

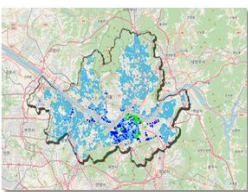
지도교수 : 산업경영공학과 성명: 허선(인)

**한양대학교 ERICA LINC+사업단 귀하**

산학협력 선도대학 육성 사업단	담 당	팀 장	단 장

## 캡스톤디자인 과제 요약서

학과(전공)	산업경영공학과		팀 명	금의한양	
참여학생명	송수빈, 유동완, 윤종훈, 이시현, 허지원				
과 제 명	머신러닝을 활용한 서울특별시 아파트 매매가 예측			수강학기	2020년-1학기
과제분야	■공학   □ 자연과학   □ 디자인   □ 인문사회				
과제유형	□ <b>C</b> ommunity 사회문제 □ <b>I</b> ndustry 산업체연계 □ <b>E</b> fusion 학제융합		□ <b>H</b> umanity&Design 인문사회디자인 ■ <b>E</b> ngineering 이.공학 창의 □ <b>S</b> tartup 스타트업지원		
작품제작유형	□ H/W 및 H/W+S/W                      ■S/W 및 문제해결				
지도교수	소 속	산업경영공학과	성 명	허 선(인)	
1. 과제개요 및 필요성	부동산은 국내 가계 자산 중 차지하는 비중이 가장 크고 증가하고 있는 추세이며, 커진 만큼 부동산 시장의 변동성에 대한 국가적, 국민적 심리적 불안을 만들고 있다. 이에 따라 부동산 가격을 예측하는 것은 많은 이해 당사자들에게 불안함을 잠재우고 최선의 선택을 할 수 있는 근거를 마련한다.				
2. 주요성과 (특허출원포함)	2013년도에서 2017년도의 실거래가를 기반으로 머신러닝을 이용하여 2018년도의 실거래가를 예측하고 제작된 웹사이트를 통해 아파트별로 예측값과 상세 정보를 열람가능하다.				
3. 결과 및 개선방향	<p>특정 조건으로 행정동에 계층적 군집화를 진행하였다.</p> <p>평가 척도로 RMSE를 사용하였다.</p> <p>( RMSE는 실제값에 예측값을 빼고 제곱을 한 뒤 데이터 수를 나눠주는 MSE에 제곱근을 한 값임)</p> <p>Class별로 RMSE를 측정한 결과, Class1 오차가 가장크고 그 다음 Class2, 그리고 Class0의 오차가 가장 적었음.</p> <p>Class1과 Class2의 오차범위가 크게 나타난 이유에는 정부 정책, 재개발, 급매물 등 여러 이유 중 강남3구라는 환경적 요인이 큰 비중을 차지하는 것으로 보임</p> <p>1. 향후 추가적인 연구를 통해 분석 지역의 범위를 넓혀 전국 단위의 아파트 매매가 분석 및 예측을 기대할 수 있다.</p> <p>2. 연구의 사각지대에 존재했던 연립,다세대 부동산에 대한 연구 또한 기대할 수 있다.</p>				

<div>4.</div> <div>최종결과물</div> <div>(최종작품</div> <div>사진/도면/모</div> <div>델/발표자료</div> <div>등)</div>	<div><table><thead><tr><th>데이터 수</th><th>class0</th><th>class1</th><th>class2</th></tr></thead><tbody><tr><td>전체</td><td>71,806</td><td>4,416</td><td>419,629</td></tr><tr><td>2013-2017</td><td>64,834</td><td>3,907</td><td>377,675</td></tr><tr><td>2018</td><td>6,972</td><td>509</td><td>41,954</td></tr></tbody></table></div> <div><div><div>Training Data</div><div><div>Sample 1</div><div>Learning Algorithm</div><div>Sample 1</div><div>Sample 2</div><div>Learning Algorithm</div><div>Sample 2</div><div>...</div><div>Sample N</div><div>Learning Algorithm</div><div>Sample N</div></div><div><div>Model Training</div><div>Testing Data</div></div></div><div><div>Test Sample Input</div><div><div>Tree 1</div><div>Tree 2</div><div>...</div><div>Tree 600</div></div><div><div>Prediction 1</div><div>Prediction 2</div><div>...</div><div>Prediction 600</div></div><div><div>Average All Predictions</div><div>Random Forest Prediction</div></div></div></div>	데이터 수	class0	class1	class2	전체	71,806	4,416	419,629	2013-2017	64,834	3,907	377,675	2018	6,972	509	41,954	<div><div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div><div>이차전지 폐배기 배출량</div></div></div>
데이터 수	class0	class1	class2															
전체	71,806	4,416	419,629															
2013-2017	64,834	3,907	377,675															
2018	6,972	509	41,954															

본 캡스톤디자인과 관련하여 제출한 영상, 사진, 문서 등의 자료는 대학 및 프로그램 홍보 등의 비영리적인 목적으로 대학 및 산학협력단에서 사용할 수 있습니다.

2020 .06 .04

학생(대표) : 윤종훈 (인)

<양식 6> 모든 과제 팀 작성

※본 결과보고서는 표준양식으로 캡스톤디자인 과제 취지 및 특성 등을 감안하여 양식의 추가/수정 가능

# 캡스톤디자인 결과보고서

## 머신러닝을 활용한 서울특별시 아파트 매매가 예측

송수빈, 유동완, 윤종훈, 허지원, 이시현  
지도교수 허선

## Apartment Price Estimate by Machine Learning

Subin Song, Dongwan Yoo, Jonghoon Yoon, Jiwon Heo, Sihyon Lee

Supervised by prof. Sun Her

### Abstract

The project aims to improve the utilization and accessibility of forecast results, not just research, by predicting apartment prices in Seoul and creating websites for related stakeholders to easily access forecast information.

In this project, the dedicated area and actual transaction price had a major impact on the prediction of the selling price, so it was judged that it was too much to predict all the data with a single model, so the clustering was carried out of this project. Each administrative dong was divided into large, medium, and small areas using the exclusive area of the goods traded and clustered based on the selling price. Through hierarchical clustering, clusters were divided into three. Using the existing apartment-related variables, administrative-dong-related variables, economy-related variables, and facility-related variables, the model with good performance was selected for each cluster. The clusters divided into three showed the best performance in the Xgboost, LGBM and Random Forest, respectively, and were selected as predictive models.

The hyperparameters of the models in each cluster were tuned using grid search and RMSE was used as the model evaluation scale. Using tuned hyperparameters, we predicted the sale price for 2018 and showed approximately 2000 RMSE. For the visibility and convenience of the generated results, the website was created and the details were available for inquiry, thereby enhancing the utilization of the research results.

### 요약

본 프로젝트는 서울시 아파트 매매가를 예측하고 관련 이해 관계자들이 예측 정보에 쉽게 접근하기 위한 웹사이트를 제작하면서 연구로만 끝나는 것이 아니라 예측 결과의 활용성과 접근성을 높이고자 하는 연구이다.

본 프로젝트에서는 전용면적과 실거래가가 매매가를 예측하는데 주된 영향을 미치므로, 한 모델로 모든 데이터를 예측하기에는 무리가 있다고 판단하여 군집화를 진행했다. 거래된 매물의 전용면적을 이용해 각 행정동을 대,중,소로 나누고, 매매가를 기준으로 군집화를 실시하였다. 계층적 군집화를 통해

군집을 3개로 나누었다. 기존에 생성했던 아파트 관련 변수, 행정동 관련 변수, 경제 관련 변수, 시설 관련 변수를 이용하여 군집별로 성능이 좋은 모델을 선정하였다. 세 개로 나눈 군집은 각각 Xgboost, LGBM, Random Forest에서 가장 좋은 성능을 보였으며, 이를 예측 모델로 선정하였다.

각 군집마다 선정된 모델의 하이퍼파라미터는 그리드 서치(Grid Search)를 이용하여 튜닝하였고, 모델 평가 척도로 RMSE를 사용하였다. 튜닝된 하이퍼파라미터를 이용하여 2018년 매매가를 예측했고, 대략 2000대의 RMSE를 보였다. 생성된 결과의 가시성과 편리성을 위해 웹사이트를 제작하였고, 세부 내용을 조회 가능하게 하여 연구 결과의 활용성을 높였다.

---

## 1. 과제 필요성

일반적으로 부동산은 국내 가계 자산 중 차지하는 비중이 가장 크고 증가하고 있는 추세이며, 부동산의 꽤 많은 비중을 차지하는 아파트 시장의 규모 또한 점점 커지고 있다. 커진 시장의 크기만큼 부동산 시장의 작은 파장은 국민들의 심리적 불안을 만들고, 국가의 거시경제에 큰 변동성을 주고 있다. 이에 따라 부동산 시장의 많은 이해 당사자들, 예를 들어 부동산 개발자 및 투자자, 공인중개사, 감정평가사, 금융기관 뿐만 아니라 주택을 거래하는 일반 시민 모두가 영향을 받는 대상이 되었고 그에 따라, 꾸준히 많은 주목을 받고있는 시장이다. 따라서 주택가격의 예측과 영향을 미치는 요인 분석은 미시적인 관점, 거시적 관점에서 모두 의미가 있다. 따라서, 2013년부터 2017년 서울특별시 아파트 매매가를 이용하여 2018년 서울특별시 아파트 매매가를 예측하는 머신러닝 모델을 구축하는 것을 목표로 한다. 추가적인 연구를 통해 분석 지역의 범위를 넓혀 서울특별시뿐만 아니라 전국 단위로 확장하여 아파트 매매가 분석 및 예측이 가능할 것으로 기대된다.

## 2. 과제 해결 방안 및 과정

### 2.1 데이터 수집

공공데이터포털에서 2013년도부터 2017년도까지 서울시에서 거래된 56만6천개의 아파트 매매가 API 데이터를 기초로 이용했고, 아파트 가격에 영향을 끼치는 새로운 변수들을 행정안전부, 통계청, 서울열린데이터, 공공데이터포털 등을 이용해서 생성했다.

#### 2.1.1 1차 결측값 제거

아파트 가격에 영향을 끼치는 새로운 변수 중, 공시지가 변수 데이터를 공공데이터 포털에서 제공받아 기초 데이터와 매칭시키는 과정에서 대량의 결측치가 발생하였는데, 이를 제거하였다.

#### 2.1.2 행정동 매칭

공공데이터 포털, 서울열린데이터 등 대부분의 참고할 수 있는 자료들은 행정동을 기준으로 생성되었으나 우리의 데이터에는 법정동만 있어 행정동으로 변환하는 과정이 필요했다. 이를 위해 통계청과 같은 공공 포털을 참고하여 도로명 주소를 기준으로 하여 행정동을 매칭시켰다. 매칭이 안돼 발생한 결측치는 도로명 주소를 직접 검색하여 수작업으로 채워넣는 방식으로 처리하였다.

## 2.2 변수 생성

초기의 데이터는 ['거래금액', '건축년도', '년', '도로명', '도로명건물본번호코드', '도로명건물부번호코드', '도로명시군구코드', '도로명일련번호코드', '도로명코드', '법정동', '법정동본번호코드', '법정동부번호코드', '법정동시군구코드', '법정동읍면동코드', '법정동지번코드', '아파트', '월', '일', '일련번호', '전용면적', '지번', '지역코드', '층']으로 기초적인 변수밖에 없어 매매가 예측일 위해 새로운 변수들의 생성은 필수적이었다.

### 2.2.1 아파트 관련 변수

아파트 관련 변수는 아파트와 직접적으로 관련된 물리적인 변수를 의미한다. 따라서 매매가의 큰 영향을 줄 수 있는 변수이다. 생성한 변수는 공시지가, 최고 층, 8대 브랜드 변수이다. 먼저 공시지가는 정부에서 감정평가사가 각 부동산에 대해 감정평가한 금액이다. 이 부분에서는 결측값이 꽤 발생했는데 일사편리라는 부동산정보 통합 열람 홈페이지를 이용하여 수작업으로 채워넣었다. 최고 층 변수는 각각의 아파트별로 거래된 매물 중 최고로 높은 층수를 최고 층 변수로 생성하였다. 마지막으로 8대 브랜드 변수는 매년 평균 매매가가 높은 브랜드와 인터넷 기사, 설문조사 등을 참고하여 8대 브랜드를 선정하였고 변수로 생성했다.

### 2.2.1 행정동 관련 변수

앞서 매칭한 행정동을 기준으로 생성한 변수를 의미한다. 행정동은 변수를 생성할 데이터를 찾을 수 있는 단위 중 가장 작은 단위로 인구, 기초수급자 수, 65세 이상 노인 수, 외국인 수, 행정동별 매매가 상하위 10%, 주차공간의 변수를 생성했다. 먼저 인구, 기초수급자 수, 65세 이상 노인 수, 외국인 수는 행정동별 존재하는 수를 매칭하여 변수를 생성했다. 행정동별 매매가 상하위 10%는 행정동별로 매매가를 평균내어 상위 10%에 속하면 1 그렇지 않으면 0으로 표현한 변수를 생성했고 하위 10%도 같은 방식으로 변수를 생성했다.

### 2.2.2 시설 관련 변수 생성

시설 관련 변수는 해당 아파트를 기준으로 300m (100m, 300m, 500m, 1000m를 모두 시행하여 300m가 최적이라 판단했고 참고 논문 또한 300m로 지정)내에 있는 시설의 수를 기준으로 생성한 변수를 의미한다. 이를 위해서는 아파트의 위도, 경도가 필요했는데, 이는 각 아파트의 도로명 주소를 만든 후 구글 geocoding API를 이용하여 위도와 경도를 구했다. 생성한 변수는 초등학교, 중학교, 고등학교, 지하철, 환승역 여부, 병원, 시장, 도서관, 공장으로 각 아파트의 좌표와 해당 시설의 좌표를 haversine을 이용해 거리를 구하는 방식으로 생성했다. 이후 금액과의 상관관계가 높아질 수 있도록 각 변수들마다 일정한 공식을 이용해 변환과정을 거쳤다.

### 2.2.3 경제 변수 생성

경제 변수는 해당 시기에 경제의 움직임을 나타낸 변수를 의미한다. 부동산 매매가를 결정짓는 데는 거시 경제의 영향도 꽤 크므로 국내 총생산, 경제 성장률, 집세물가지수, 소비자 물가 지수, 코스피, 환율 변수들을 생성했다. 각각의 변수들은 부동산 매물이 거래되던 시기에 매칭하는 것으로

생성되었다.

#### 2.2.4 2차 결측값 제거

모든 변수를 생성한 후, 공시지가와 같은 일부 변수는 결측값이 생성되었다. 해당 자료 자체가 없는 것으로 주로 거래량이 적은 곳에서 발생하여 이는 전체에 큰 영향을 미치지 않을 것이라 판단했다. 따라서 어떠한 변수에 대해서 하나라도 결측값이 존재하는 데이터들은 제거하였다.

### 2.3 EDA 및 시각화

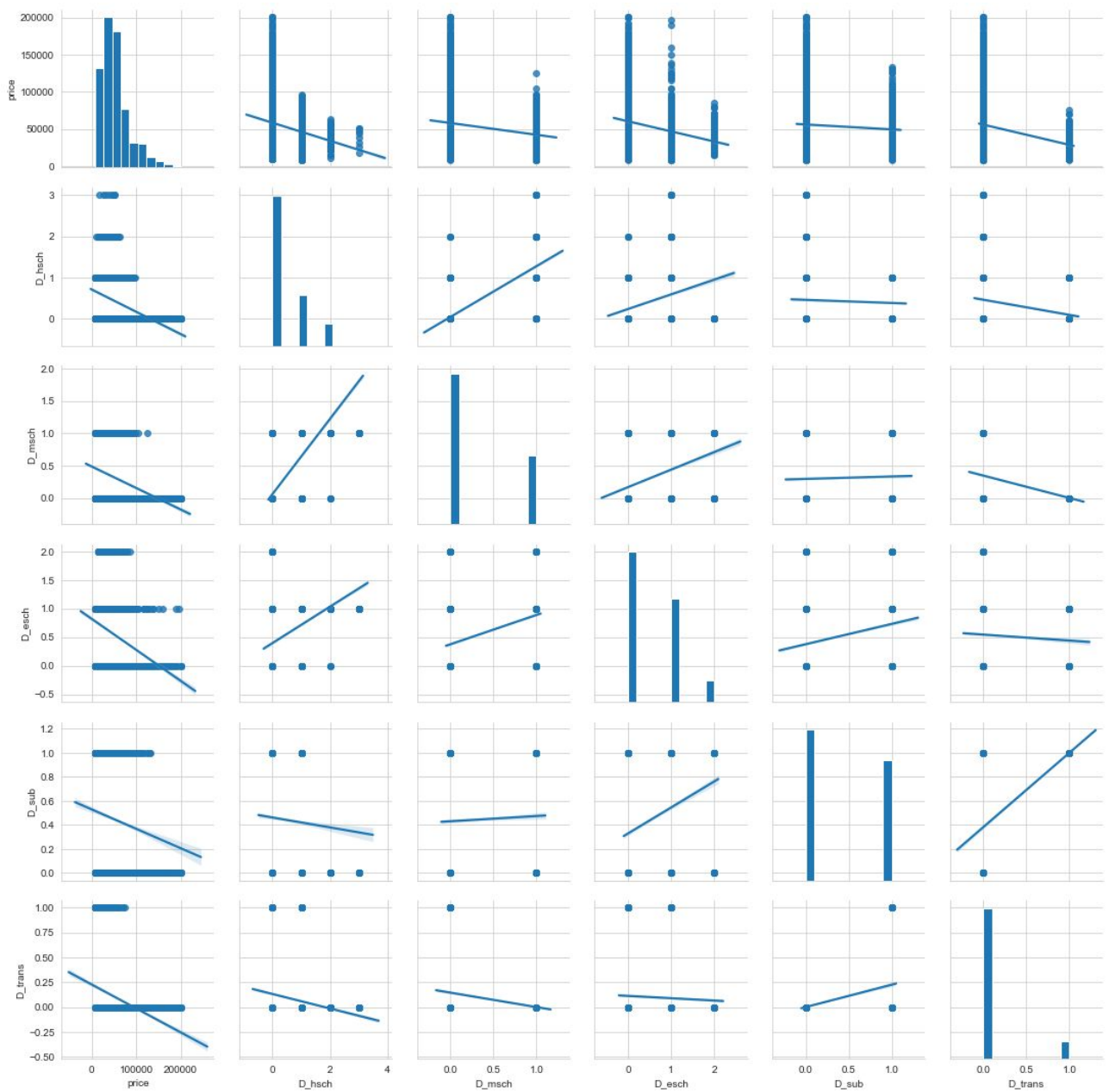


<그림. 아파트 관련 변수들간의 상관관계 그래프>

건축년도가 현재와 가까워질수록 층수가 높아진다. 2000년대 초반에는 20층 이상의 층수가 높은 층수가 많이 지어졌지만, 현재는 15층~20층 정도의 아파트가 주로 지어지고 있다. 이를 볼 때, 2000년대 초반에는 높은 아파트를 선호했다면 현재와 가까워질수록 15층~20층의 아파트를 선호했다고 볼 수 있다. 또한, 2000년대 초반에는 큰 평수의 아파트를 많이 지었고, 2010년전후로부터 전용면적이 100 정도로, 평수로 따지면 30평대 아파트를 더 많이 지은 것으로 보아, 2000년대 초반과는 다르게 최근

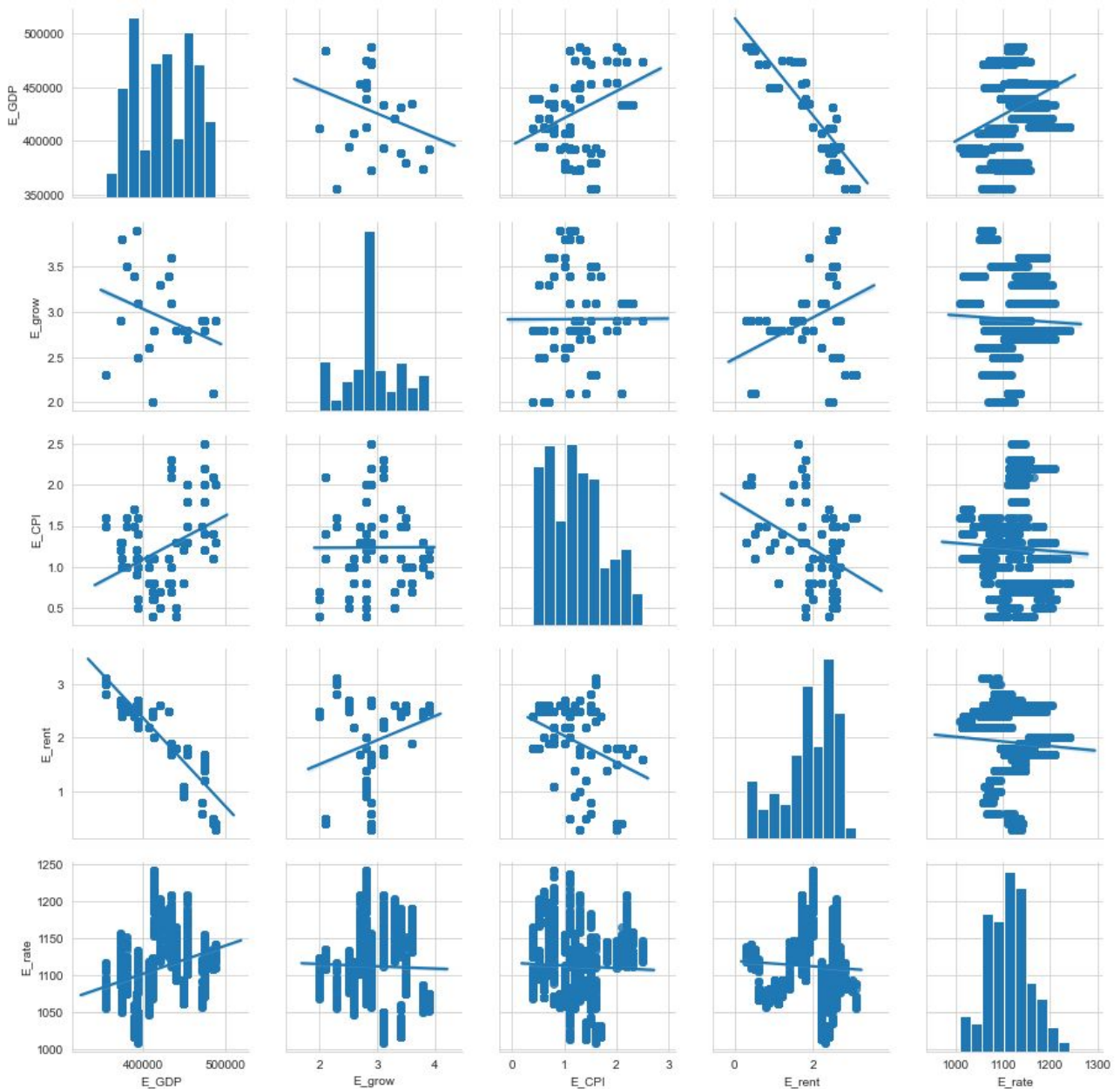


들어 작은 평수를 더 선호하는 추세로 변하고 있음을 알 수 있다. 마지막으로, 건축연도가 현재와 가까운 아파트일수록, 즉 최근에 지어진 아파트 일수록 거래건수가 많은 것을 알 수 있었다.



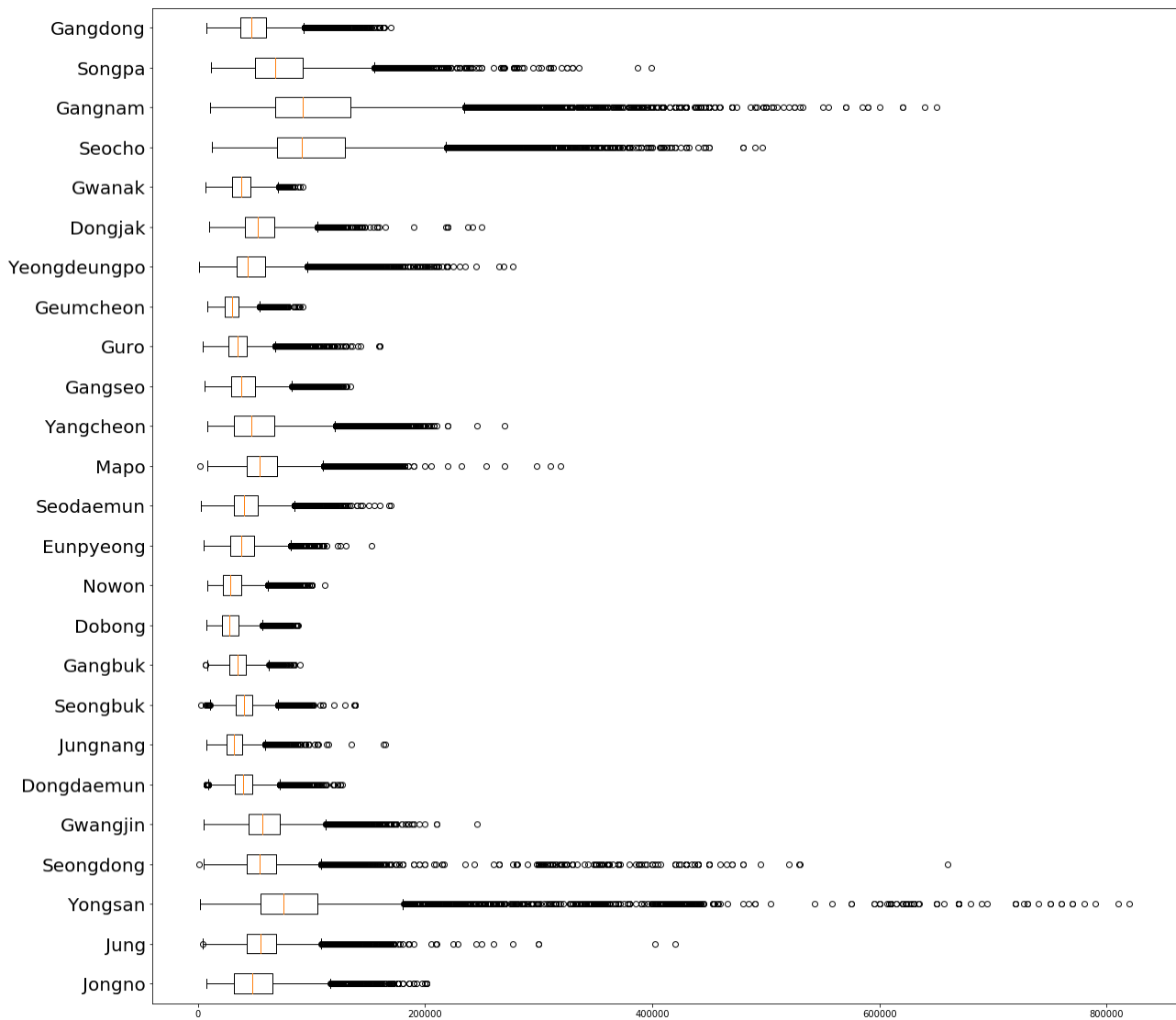
<그림. 시설 관련 변수와 가격의 상관관계 그래프>

300m내에 환승역이 있을 경우, 10억 아래의 분포를 보이고, 300m내에 역이 있을 경우, 13억 보다 아래에 값이 존재함을 알 수 있었다. 300m내에 초등학교가 있는 경우, 2개 있을 경우 10억 아래이지만, 1개 있을 경우 최대 20억까지 나가는 경우가 있다. 300m내에 중학교가 있는 경우, 이때는 1개일 경우밖에 없으며, 300m내에 고등학교가 1개, 2개, 3개 있을 경우 모두 10억 아래에 값이 존재하고, 고등학교 수가 많아질수록 값의 분포도 줄어든다.



<그림. 경제 관련 변수간의 상관관계 그래프>

국내 총생산과 일별평균환율은 환율이 1075에서 1150 사이일 때는 양의 상관관계를 보인다. 국내 총생산과 소비자물가는 아주 약한 상관관계를 가진다. 국내 총생산이 증가할수록 소비자 물가도 증가한다. 국내 총생산과 집세증감률은 강한 음의 상관관계를 갖는다. 국내 총생산과 경제성장률은 경제성장률이 2.5에서 3.25 사이일 때 음의 상관관계를 가지며, 이는 국내 총생산이 늘면 경제성장률은 감소하는 것으로 해석할 수 있다.



<그림. 구별 가격 분포 확인>

용산구, 성동구, 강남구, 서초구에서 거래된 매매가가 다른 구들과 비교했을 때, 압도적으로 높음을 알 수 있다. 이 중 용산구에서 60억대의 매매가가 형성되는 이유는 한남동의 한남더힐 아파트 때문임을 알게되었다.

## 2.4 데이터 전처리(이상치, 다중공선성 제거)

첫번 째로, 군집화를 위해 묶은 행정동별 가격 분포에서 극단 이상치를 삭제하였다. 모델의 합리성을 위해 4분위수  $1Q-3IQR$ ,  $3Q+3IQR$ 에 해당하는 극단 이상치를 제거하였다. 그 결과, 행정동421개에 해당하는 48만655개에서 행정동 274개와 446,416개가 남았다.

두번 째로, 다중공선성을 피하기 위해 변수간의 상관관계를 파악한 후 상관관계가 0.7이 넘는 집세지수, 코스피, 행정동별 인구 변수를 제거하였다.

## 2.5 클러스터링 데이터 생성

각 행정동별로 년도별 매매가가 유사하게 움직이는 행정동을 찾기 위해 그리고 전체 데이터는 편차가 매우 커 작게 나누기 위해 클러스터링이 필요했다. 이를 위해서는 클러스터링을 하기 위한 데이터가 필요했다. 데이터의 행은 행정동을 기준으로 해야 했는데, 하나의 행정동에도 편차가 매우

켰기 때문에 행정동을 분리해줄 필요가 있었다. 행정동을 나누는데 가장 좋은 방법은 전용면적을 기준으로 나누는 것이라 판단하여 대중소로 나누었다. 동일한 기준으로 대중소를 나눌 것인지 아니면 각 행정동별로 기준을 찾아 나눌 것인지에 대하여 판단해야 했는데, 동일한 기준으로 나눌 경우 행정동별로 편차를 고려하지 않은 결과가 나올 것으로 판단되어 각 행정동별 전용면적의 30%, 70% 지점의 값을 기준으로 분리하였다. 전용면적이 30%이하면 행정동\_소, 70% 이상이면 행정동\_대, 그리고 그 사이 값이면 행정동\_중으로 구분해주었다. 구분 결과를 행으로 하고 열을 연도별 평균 금액으로 하여 클러스터링 데이터를 생성하였다.

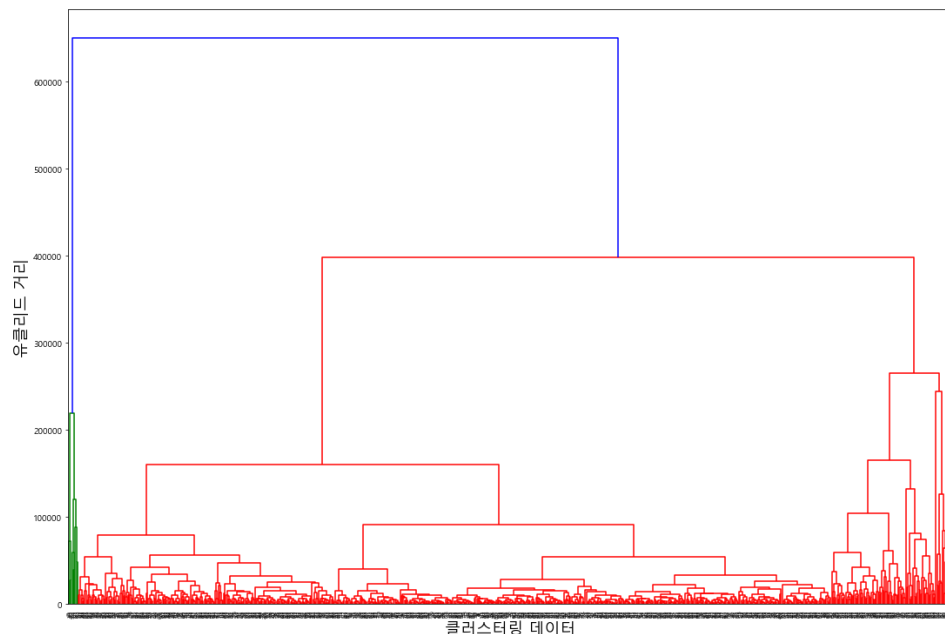
```
final_data.head(3)
```

	area	2013	2014	2015	2016	2017	2018
0	화곡제1동_중	20649	20419	21370	23340	24780	25795
1	화곡제1동_소	13742	13224	12024	14917	12395	18195
2	화곡제1동_대	28344	29397	30233	32109	34384	38011

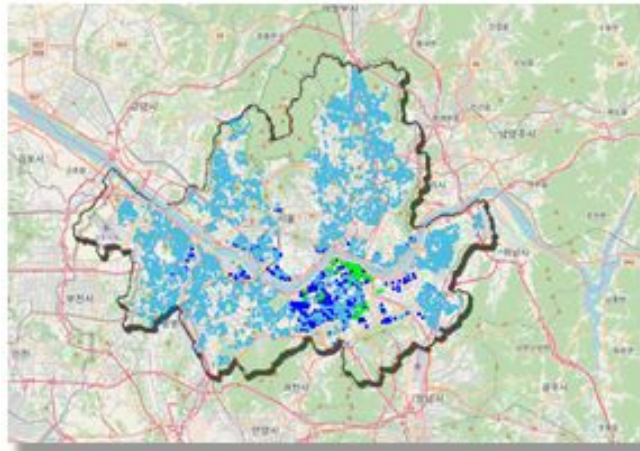
<그림. 클러스터링 데이터 예시>

### 2.5.1 클러스터링

클러스터링 데이터를 이용하여 계층적 군집화를 할 것인지 비계층적 군집화를 할 것인지 판단해야 했다. 대중적인 K-평균 군집화(K-mean Clustering)의 경우 K 클러스터의 수를 특정하기에 어려움이 있어 클러스터의 개수를 먼저 입력하지 않아도 되는 계층적 군집화(Hierarchical Clustering)를 이용했다. 계층적 군집화 후 나온 결과인 덴드로그램을 보고 클러스터의 개수를 3개로 판단하여 진행하였다.



<그림. 계층적 군집화 시각화>



<그림. 군집화된 행정동 지도 시각화>

### 2.5.2 3차 결측값 제거

클러스터링 과정 중 대중소로 나눈 행정동의 년도별 평균을 낼 때 샘플의 수가 10개 이하이면 데이터가 지나치게 편중될 수 있다고 판단했고, 이를 결측값이라 생각하고 해당 행정동을 제거하였다.

## 2.6 모델 선정

서울시의 경우 각 지역마다 매매가의 편차가 매우 크기 때문에 하나의 모델로는 예측하기 힘들다고 판단되었고, 군집화 결과 나누어진 데이터들은 서로 비슷한 가격 추이를 보이는 데이터라고 판단하여 군집별로 머신러닝을 이용하여 클래스마다 예측 모델을 구축했다. 모델은 Ridge, Lasso, Elastic Net, Random Forest, Extratrees, Adaboost, Gradientboost, Xgboost, LGBM 등의 여러 모델을 이용하였고 평균적으로 가장 성능이 좋은 Random Forest, Extratree, Xgboost, LGBM 4개 모델을 선정했다.

### 2.6.1 모델 데이터 생성

클러스터링 결과를 행정동의 대중소가 3개로 나뉘었다. 이를 이용하여 원본 데이터를 3개의 class로 나누고, 훈련 및 검증 데이터인 2013~2017년과 테스트 데이터인 2018년으로 나누었다.

<표. class별 데이터 수>

데이터 수	class0	class1	class2
전체	71,806	4,416	419,629
2013~2017	64,834	3,907	377,675
2018	6,972	509	41,954

### 2.6.1 모델 기법 적용

앞서 선정한 4개의 모델을 이용하여 단순한 몇 번의 예측 결과 각 군집마다 좋은 성능을 보이는 모델이 달랐다. 0~2번 군집은 각각 Xgboost, LGBM, Random Forest가 좋은 성능을 보여 해당 모델로 선정하였다.

### 2.6.2 하이퍼 파라미터 튜닝

각 군집마다 선정된 모델의 하이퍼파라미터를 그리드 서치(Grid Search)를 이용하여 튜닝하였다. 모델 평가 척도로 RMSE를 사용했는데, Grid Search CV결과 RMSE와 TEST 결과 RMSE의 차이가 작은 파라미터를 선정하여 과대적합의 위험성을 줄이면서 튜닝을 마쳤다.

### 2.6.3 예측

튜닝된 하이퍼파라미터를 이용하여 2018년 매매가를 예측했고, 대략 2000대의 RMSE를 보였다. 생성된 결과는 가시성과 편리성을 위해 사이트에 올려 조회할 수 있도록 하나의 데이터로 합쳐 생성했다.

## 2.7 사이트 생성

클라우드 기반의 무료 홈페이지 제작 사이트 WIX를 이용하였다. 아파트 정보와 연구 결과를 표로 정리한 후, WixDataQuery를 이용하여 사이트에 적용하였다. 홈페이지에 '아파트 예측 가격 열람 서비스' 버튼을 만든 후, 누르면 예측 가격을 열람할 수 있는 페이지로 가는 기능을 추가하였다. WixDataQuery로 불러온 최종 데이터를 기반으로 동적 페이지를 제작하였다. '시,군,구'와 '읍,면,동' 버튼을 통해 찾고자 하는 아파트의 지역을 선택할 수 있고, 아파트 이름 검색 기능을 추가하였다. 해당 페이지에서 '상세정보' 버튼을 누르면 해당 아파트관련 상세한 정보와 위치를 알려주는 Google지도를 제작하였다.





## 상세정보

골든빌(99-1) 253765 만 원



## <그림. 웹 사이트 조회 방법>

### 2.7.1 프로젝트 적용

최종 데이터를 콘텐츠로 웹사이트를 제작함으로써, 부동산 이해관계자가 정보에 쉽게 접근할 수 있도록 용이성을 고려하였다. 웹사이트를 제작함으로써 시간제약없이 정보를 제공할 수 있고, 새로운 결과값을 간단하게 업데이트 가능하다.

## 3. 개념설계 및 상세설계(계산)

### 3.1.1 모델 XGboost

머신러닝 앙상블 부스팅 모델에서 병렬처리와 최적화를 장점으로 내세우는 Gradient Boosting 알고리즘의 단점인 느리고, 과적합 문제를 해결하기 위해 만들어진 여러 모델 중 하나의 모델이다. XGboost는, 앙상블 부스팅의 특징인 가중치 부여를 경사하강법으로 진행하는 방식의 모델이다. CART(Classification And Regression Tree)를 기반으로 하여, 분류와 회귀 둘 다 가능하다. 또 GBM의 문제점을 해결하기 위해 만들어진 모델이므로 GBM보다는 빠른 연산이 가능하다. 그리고 과적합 방지가 가능한 규제가 포함되어있다. XGboost의 수식은 다음과 같다.

- Model: assuming we have K trees

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F}$$

- Objective

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Training loss

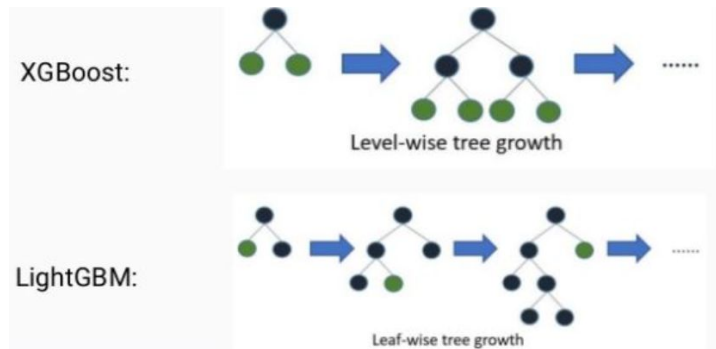
Complexity of the Trees

<그림. XGBoost의 수식>

### 3.1.2 모델 LGBM

Gradient Boosting(GBM)은 가중치를 경사하강법으로 진행한다. 기존의 알고리즘의 단점을 보완한 XGBoost가 존재한다. XGBoost는 좋은 성능을 보여주지만, 학습이 오래 걸리는 단점을 가지고

있다. 이를 보완하기 위해 LightGBM이 탄생하였다. 대용량 데이터 처리가 가능하고, 다른 모델들보다 더 적은 메모리를 사용하기 때문에 비교적 빠른 연산이 가능하다. 기존의 트리모형은 트리의 깊이(depth)를 줄이기 위해서 균형 트리(Level wise)를 이용하여 분할하지만, LGBM은 트리의 균형이 아닌 리프 노드를 지속적으로 분할하면서 진행한다.

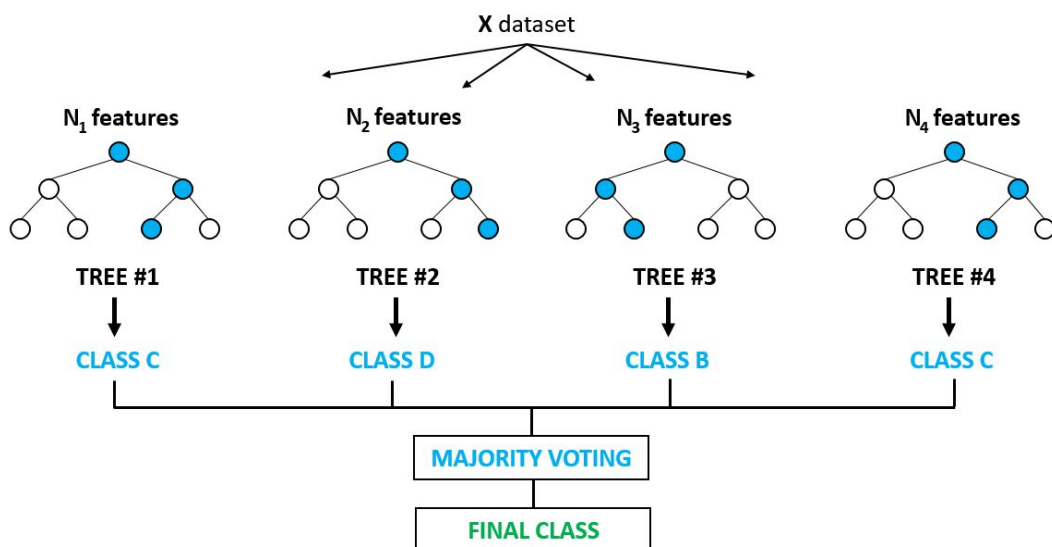


<그림. XGBoost와 LightGBM 차이>

### 3.1.3 모델 Random Forest

앙상블 중 하나로 같은 알고리즘으로 여러 개의 분류기를 만들어서 보팅으로 최종 결정하는 알고리즘이다. 앙상블 알고리즘 중 비교적 빠른 수행 속도를 가지고 있다. Random Forest의 기반 알고리즘은 결정 트리로서, 결정 트리 분류기가 전체 데이터에서 배깅 방식으로 각자의 데이터를 샘플링(부스트랩)해 개별적으로 학습을 수행한 뒤 최종적으로 모든 분류기가 보팅을 통해 예측 결정하게 된다.

하나의 트리는 계층 구조로 이루어진 노드와 엣지들의 집합으로 이루어진다. 중간 노드는 두개 또는 몇 개의 노드로 분리되는데 이때 분리의 기준은 엔트로피 지수, 지니 계수, 카이제곱 통계량 등이 있으며 변수들 중 이 기준을 최소화하는 값으로 나눈다.

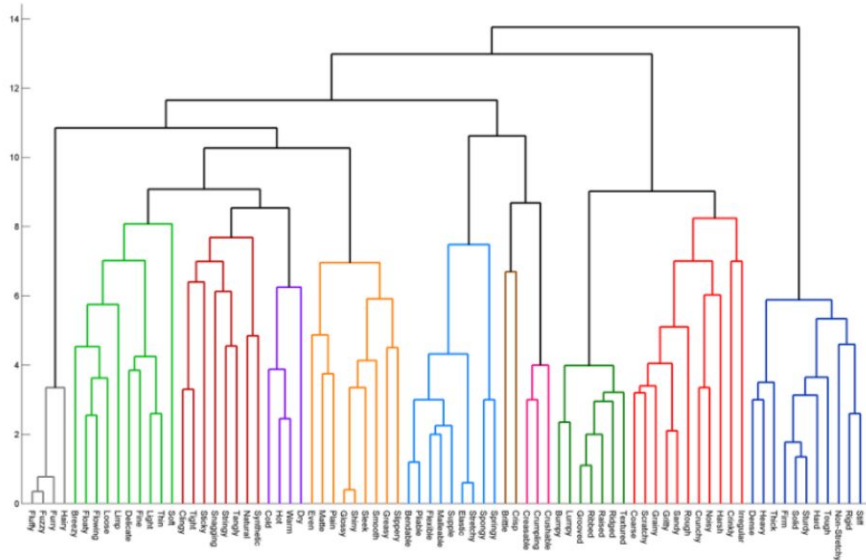


<그림. 랜덤포레스트>

### 3.2 계층적 클러스터링



계층적 클러스터링이란 계층적 트리 모형을 이용해 개별 개체들을 순차적, 계층적으로 유사한 개체 내지 그룹과 통합하여 군집화를 수행하는 알고리즘이다. 개체들이 결합되는 순서를 나타내는 트리형태의 구조인 덴드로그램 덕분에 K-평균 군집화와 달리 군집 수를 사전에 정하지 않아도 학습을 수행할 수 있다. 아래 그림과 같은 덴드로그램을 생성한 후 적절한 수준에서 트리를 자르면 전체 데이터를 몇 개 군집으로 나눌 수 있게 된다.



<그림. 계층적 클러스터링의 덴드로그램>

#### 4. 기대효과 및 개선 방향

부동산은 지속적인 관심을 받고 있는 분야인만큼 아파트 예측 모델은 다양하게 활용될 수 있다.

연구대상이 아파트로 한정되었던 본 프로젝트의 한계를 넘어 상대적 빈도수가 적어 연구의 사각지대에 존재했던 연립, 다세대 부동산에 대한 연구 가능성을 제시한다. 추가적으로, 각 지역의 거래 데이터를 구하면, 분석 지역의 범위를 넓혀 서울특별시뿐만 아니라 전국 단위로 확장하여 아파트 매매가 분석 및 예측이 가능할 것이다. 정확한 예측을 위해 아파트 매매가와 관련된 요인을 세부적으로 연구할 필요가 있고, 아파트를 군집화하는 기준에 대해 추가적인 논의가 필요하다.

## 5. 결론

급변하는 부동산 시장에서 아파트의 매매가를 예측하는 것은 어려운 일이었다. 하지만, 아파트에 영향을 주는 요인을 규명해가며, 어떤 요인이 오차에 영향을 미쳤는지 분석하고 결과를 추론해 나가는 과정에서 프로젝트의 의의가 있었다.

## 비용 분석

항목	세부항목	소요비용
재료비		

시제품가공비		
기타 경비		

## 참고문헌, 부록

김진형(2014), 주택 매매가격 및 전세가격에 영향을 미치는 거시경제지표 분석, 부산대학교 대학원 석사학위논문.

유하연(2015), 통계 모형을 이용한 서울시 아파트 매매가 예측 분석, 석사학위논문

이태형(2019), 인공지능경망을 활용한 주택가격지수 예측에 관한 연구: 서울 주택가격지수를 중심으로, 박사학위논문

배성완,유정석(2018), 기계 학습을 이용한 공동주택 가격 추정: 서울 강남구를 사례로, 부동산학연구

최철(2016), 서울시 아파트 매매가격 결정 요인에 관한 연구: 서울시 매매가격 상위 5개구를 중심으로, 석사학위논문

주지혜(2018), ARIMA 모형을 활용한 단독주택 실거래가 상승률 예측, 박사학위논문

김호진(2019), 거시경제 변수와 주택정책이 아파트가격에 미치는 영향, 석사학위논문

박성호(2018), 기계 학습을 이용한 아파트 가격결정요인 분석 : 부산 지역을 사례로, 석사학위논문

유석종 (2019), 나이브 베이즈 분류를 활용한 부동산 추천 기법 연구

김경민 (2017), 데이터마이닝을 이용한 부동산펀드 위험 특성 분석. 부동산분석

민성욱(2017), 딥러닝(Deep Learning)을 이용한주택가격 예측모형 연구-서울시 주택 매매 실거래가를 중심으로, 박사학위논문

김근석(2015), 부동산 실거래자료에서의 이상치 탐색 방법

유창영(2010), 아파트 브랜드가 아파트 판매가에 미치는 영향분석, 석사학위논문

<그림. XGBoost수식>, <https://brunch.co.kr/@snobberys/137>

<그림. XGBoost와 LightGBM 차이>,  
<https://www.slideshare.net/GabrielCyprianoSaca/xgboost-lightgbm>

<그림. 랜덤포레스트>,  
<https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57>

<그림. 계층적 클러스터링의 덴드로그램>,  
<https://ratsgo.github.io/machine%20learning/2017/04/18/HC/>

보고서작성자:  
 송수빈 산업경영공학과  
 E-mail: muzmuz961009@naver.com  
 TEL:010)7206-5877  
 유동완 산업경영공학과  
 E-mail: applyoo@naver.com  
 TEL:010)4153-0104  
 윤종훈 산업경영공학과  
 E-mail: [gbswhdons1213@gmail.com](mailto:gbswhdons1213@gmail.com)  
 TEL:010)7338-2209  
 허지원 산업경영공학과  
 E-mail:  
 TEL:  
 이시현 산업경영공학과  
 E-mail: :sihyeon3523@gmail.com  
 TEL:010)5416-9524  
 지도교수  
 -허 선 교수님  
 E-mail:hursun@hanyang.ac.kr TEL, FAX: 031)400-5265

#### 팀원간 역할 분담

성명	역할	참여도(%)
송수빈	전처리 및 클러스터링	100%
유동완	데이터 수집 및 변수 생성	100%
윤종훈	사이트 제작 및 모델링	100%
허지원	사이트 제작 및 문서 정리	100%
이시현	데이터 시각화 및 모델링	100%