

머신러닝을 활용한 서울특별시 아파트 매매가 예측

소속 산업경영공학과

지도교수 허선 교수님

팀명 금의한양

팀원 송수빈, 유동완, 윤종훈, 허지원, 이시현



I. 배경 및 목적

■ 배경

- 주택 가격은 주택 시장, 개인, 정책당국, 금융기관에 유의미한 영향을 미치는 중요한 요소이므로 부동산 가격 예측은 다양한 이해관계자들에게 도움을 줌.
- 정책당국자와 투자자에게는 시장의 안정성 파악을 가능하게 하고, 개인에게는 자산 관리 등의 의사결정을 내리는 중요한 판단 기준을 마련.

■ 목적

- 머신러닝 기법을 이용하여 2013년부터 2017년까지의 서울특별시 아파트 실거래 매매가 데이터를 기반으로 2018년 매매가를 예측.

II. 프로젝트 세부내용

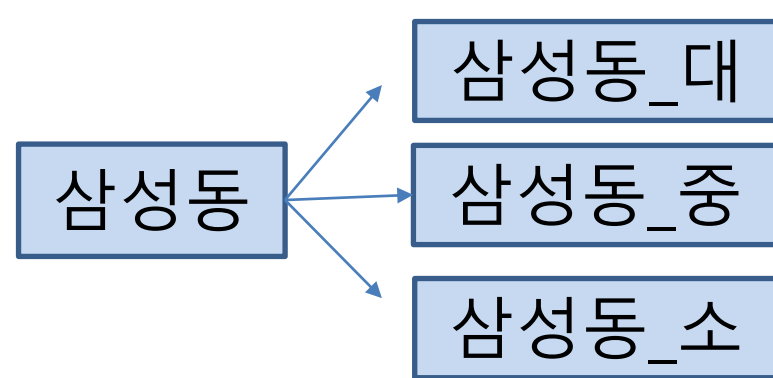
■ 데이터 전처리 및 분석

1. 상관계수 파악 및 변수 제거

변수 간 높은 상관관계를 보이는 변수 제거
- 집세지수, 코스피, 행정동 별 인구

2. 클러스터링 관련 변수설정

전용면적을 대, 중, 소로 구분하여 행정동을 세 개로 분리 후 시행



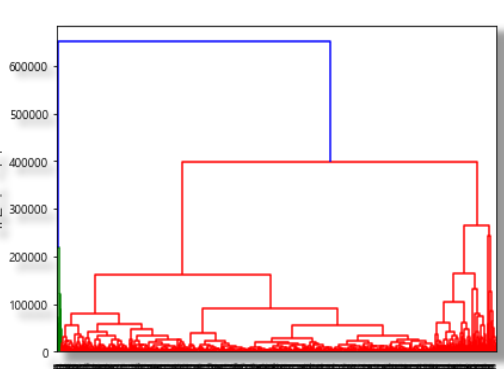
3. 결측치 및 이상치 제거

- 1) 공시지가 매칭 불가능한 데이터 제거
- 2) 클러스터링 과정 중 샘플 10개 이하 행정동 제거
- 3) 행정동 별 매매가 극단 이상점 제거
- (1Q - 3*IQR)이하 값, (3Q + 3*IQR) 이상 값 제거

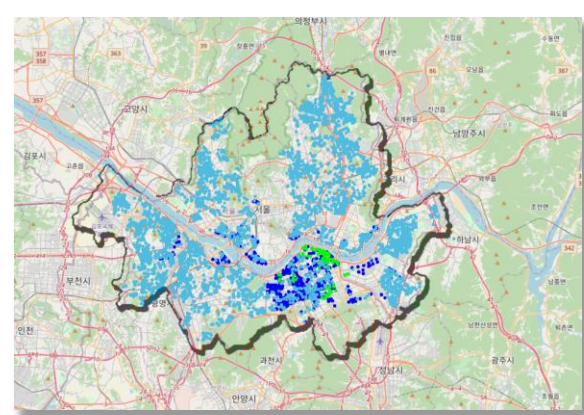
486,555개 데이터
(행정동 421개)

446,416개 데이터
(행정동 274개)

4. 군집화



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



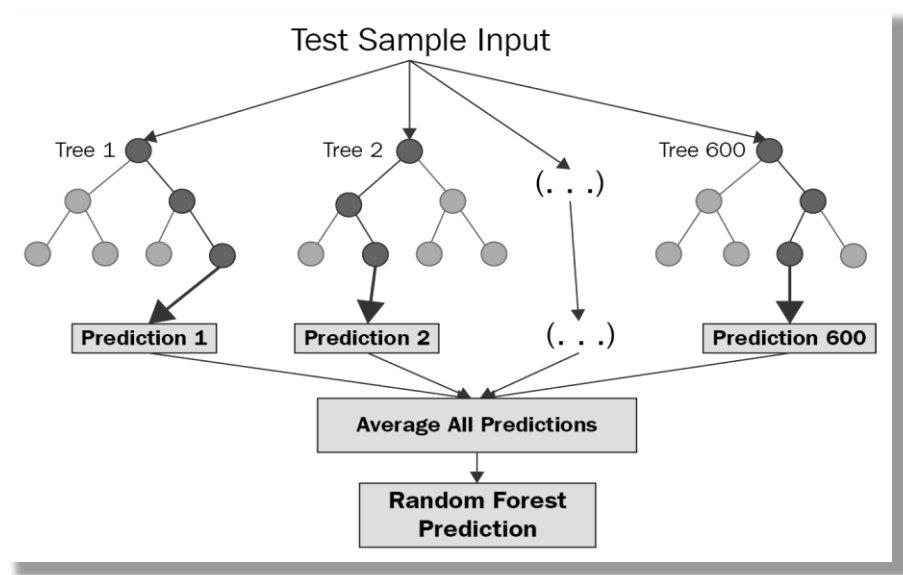
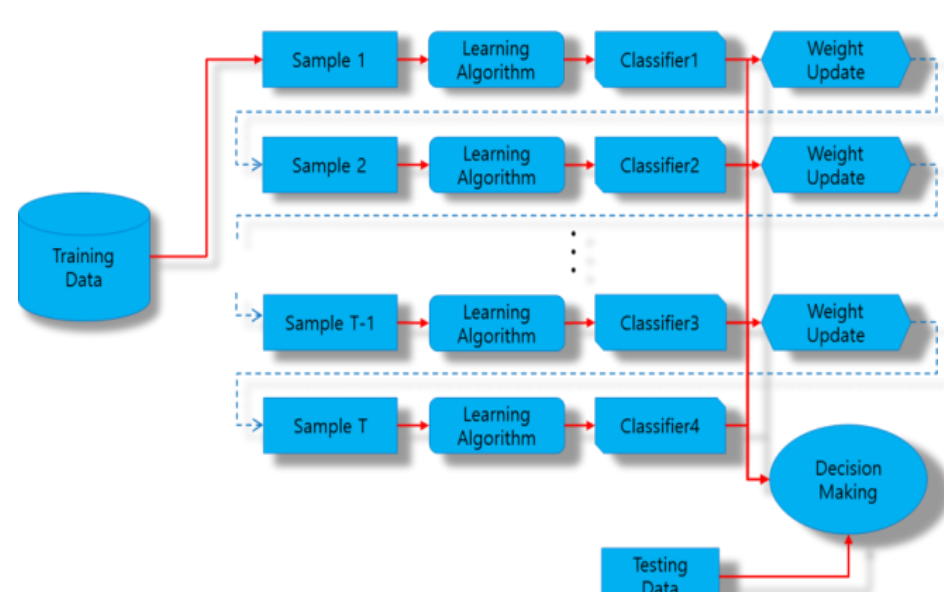
행: 클러스터링 기준 변수
열: 년도 별 행정동 별 가격 평균

■ 예측 모형 설명 및 결과

Boosting기반
XGboost, LGBM

앙상블 모형

Random Forest, Extra trees



III. 데이터 수집 및 설명

■ 데이터 출처



■ 데이터 설명

행정동 관련 변수

- 행정동 별 기초수급자 수
- 행정동 별 인구
- 행정동 별 65세 이상 인구
- 행정동 별 외국인 인구
- 행정동 별 상위 10% 가격 평균
- 행정동 별 하위 10% 가격 평균
- 행정동 별 주차공간 수

아파트 관련 변수

- 거래가
- 거래 년도
- 건축 년도
- 전용면적
- 아파트별 주차공간 수
- 층 수
- 공시지가
- 8대 브랜드 아파트
- 최고층

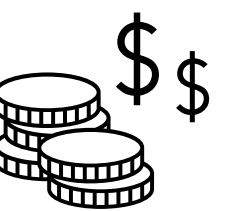


시설 관련 변수

- 반경 300m 내의 고등학교 개수
- 반경 300m 내의 중학교 개수
- 반경 300m 내의 초등학교 개수
- 반경 300m 내의 지하철역 수
- 반경 300m 내의 환승역 수
- 반경 300m 내의 병원 개수
- 반경 300m 내의 시장 개수
- 반경 300m 내의 도서관 개수
- 반경 300m 내의 공장 개수

경제 관련 변수

- 국내 총생산
- 경제 성장률
- 소비자물가지수
- 집세물가지수
- 코스피
- 환율



IV. 결론 및 활용 방안

■ 데이터 분석 결과

모델 성능 척도

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

정부 정책

급매물

재개발

환경적 요인

■ 활용 방안 및 기대



1. 프로젝트 결과를 조회할 수 있는 웹사이트 제작 :

<https://whdzl1213.wixsite.com/hanyang>

2. 향후 연구를 통해 서울특별시 뿐만 아니라 전국 단위로 확장하여 군집분류 후 매매가 분석 기대.

3. 연립·다세대 부동산 관련 연구, 시세정보를 제공하지 않는 소형 단지 아파트에 관한 연구 또한 추가로 연구 가능성 제시.



한양대학교