1. **중고폰 적정 가격 예측 챗봇** (2017.08)
   * **중고폰 적정 가격 예측 서비스 챗봇 ‘얼마고’ 구현**

**■ 팀원 및 역할**총 6명  
리서치 & 디자인 / 데이터 수집, 가공 / 데이터 분석 알고리즘 구현 및 검증/ 챗봇 개발 / 시각화 **■ 본인 역할**데이터 전 처리 및 분석 알고리즘 구현 **■ 프로젝트 기간 및 소속 기관**2017.08.01 ~ 2017.08.27 / 경기 빅데이터 전문가 분석 과정 **■ 사용 언어 & 프레임워크**R, Python, Django, AWS **■ 프로젝트 목적**

45만건의 중고 거래 데이터를 분석하여 중고폰 모델별 적정 가격을 예측하고, 이를 카카오톡 API를 통해 구현  
**■ 프로젝트 세부 내용:**① 데이터 수집 단계  
- 45만건의 중고 품목 유통 현황 데이터 및 통신 물가 지수 데이터를 확보.  
- 통신사 3사 홈페이지에 개시된 핸드폰 정보를 크롤링하여 스마트폰 표준 데이터  
(ex) 모델명, 저장 용량, 출고가 등)를 구축  
② 데이터 가공 단계  
- 45만건의 중고 품목 유통 현황 데이터에서 5만건의 중고폰 관련 데이터 추출 후 불필요한  
데이터들은 제거  
- 중고 품목 유통 현황 데이터의 각각 레코드에 해당하는 핸드폰 모델명을 검색어 사전 기반으로  
추출  
- 이후 EDA와 시각화를 통해 데이터의 성질과 특성을 파악 (ex) 상관관계 등)  
- 필요 변수에 한 해 Data 표준화 및 정규화 진행  
③ 데이터 분석 및 알고리즘 구현  
- feature 중 중고 판매 게시글을 가장 중요한 피처로 생각하였고, 중고 판매 게시 글에 따른 등록  
제품의 흠결 등급을 매기기 위한 NLP 진행  
- 흠결 등급의 label은 같은 제품 간 가격을 z값으로 표준화 후 rank로써 평가하여 labeling  
- label된 텍스트 데이터를 tf-idf, doc2vec, 단순 빈도수 등의 워드 임베딩 과정으로 수치화  
- 수치화된 데이터를 분류 알고리즘을 적용, label에 분류하는 모델 학습  
- 이를 통해 가격 예측 모델의 제품의 흠결 등급 변수 추가  
- 가격 예측의 최종 모델 변수는 판매자의 중고 제시 가격, 출고가, 흠결 등급, 물가 지수 등을 활용  
- 최종 모델은 여러 예측 알고리즘을 시도 후, 가장 설명력이 높은 Random Forest 모델 선택 매우  
높은 수준의 RMSE와 R2 수치를 나타냄

④ 모델 검증 단계

- k-fold 교차 검증을 통한 과적합 우려 없음을 확인  
 ⑤ 챗봇 개발  
 - 예측 프로세스를 카카오톡 API를 활용, 버튼형 챗봇으로 구현  
 - 이 과정에서 서버와 데이터베이스는 직접 구축