**[별첨서식 3] 최종보고서 양식**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **종합설계 최종보고서 요약서** | | | | | | | |
| **과 제 명** | 국문 : Open API를 이용한 부동산 정보 제공 서비스 개발  (영문 : Development of Real Estate Information Service Using Open API) | | | | | | |
| **지도교수** | 박동진 교수님 | | | | | | |
| **참여업체** | 업체명 | | 담당자명 | | 연락처 | | |
|  | |  | |  | | |
| **과제수행팀**  **※책임학생 명시** | 학번 | | | | 성명 | | 연락처(이메일) |
| 201302133 | | | | 최현제 | | 01053295285  (nouu30133@naver.com) |
| 201502071 | | | | 김윤환 | | 01086815827  (gwi5827@naver.com) |
| **과제수행기간** | 2020 년 09 월 20 일 - 2020 년 11 월 20 일 ( 2 개월) | | | | | | |
| **과제유형** | 산업체 연계형 과제 □, 학생 창작형 과제 ☑ | | | | | | |
| **소요비용** | 총액 | 천원 | | 전공지원금 | | 천원 | |
| 업체지원금 | | 천원 | |
| 공주대학교 산업시스템공학과 규정에 맞추어 졸업작품 과제를 성실하게 수행하고 이에 따른 최종보고서를 제출합니다.    2020 년 11 월 20 일    (책임학생) 최현제 (인)  (지도교수) 박동진 (인)  **공주대학교 산업시스템공학과 학과장 귀하** | | | | | | | |

**◎ 요약내용**

|  |  |
| --- | --- |
| 과제의  목적 및  필요성 | 기존의 부동산 어플리케이션은 다수의 이용자들을 위해 유저 인터페이스 기반으로 부동산 정보 제공을 한다. 이 방식은 이용자들에게 가시적이고 조작하기 쉬운 UI를 제공함으로서 고객들이 원하는 정보들을 이해하기 쉽게 전달해준다. 하지만 어플리케이션은 통일 된 정보 항목으로 이용자 개개인이 필요로 하는 정보들의 니즈를 충족시킬 수 없고 각 어플리케이션마다 인터페이스 조작이 다르기 때문에 조작을 익혀야 한다는 단점이 있다. 이러한 부동산 어플리케이션의 두가지 단점들을 파훼하고자 해당 프로젝트를 기획하였다. |
| 과제수행  내용/방법  (회로도 flow chart 등 첨부) | 부동산 특정 정보를 알고 싶은 고객이 의뢰를 하는 것을 가정으로 한 상황에 대한 시뮬레이션을 하였다. 기존 부동산 어플리케이션과 같이 다수의 불특정 고객들을 대상으로 인터페이스를 제공하는 기능과 다르게 웹페이지에 구독 기능을 만들어 임의의 사용자가 이메일을 입력하고 구독 버튼을 누르면 데이터베이스에 저장되게 만들었다. 서비스를 웹페이지에 구독 기능을 넣기 위해 파이썬 플라스크를 사용 하였으며 유지 보수를 원활히 하기 위해 MVC 아키텍쳐 기반으로 구성하였다. |
| 결과물에 대한 기술 | 임의의 의뢰자가 19월 1월부터 20년 9월 까지 충청도의 전역의 아파트 매매 현황과 이를 기반한 매매가의 예측이라는 가정하에 시뮬레이션을 시행했다. 해당 데이터의 추출은 공공데이터 사이트의 xml 기반 Open API로 해결 하였으며 데이터 전처리 및 예측은 파이썬의 각종 라이브러리를 이용하여 해결하였다. |

**◎ 시작품 사진(전,후,좌,우)**

|  |  |
| --- | --- |
|  |  |
|  |  |

**종합설계 최종보고서**

1. 과제개발 목적의 필요성

(1) 목적 및 필요성

**[기존 부동산 정보 제공 어플리케이션이 가지는 정보의 한계를 생각하다.]**



해당 이미지는 모 회사의 부동산 정보 제공 서비스 어플리케이션이다. GUI 기반으로 정보를 제공해주고 있으며, 데이터 베이스의 SQL처럼 필터 기능을 이용하여 자신이 원하는 조건의 부동산 정보를 쉽게 찾을 수 있다. 하지만 해당 필터 조건의 정보는 한정적이며, 이용자들 각각의 원하는 정보를 모두 제공할 수 없는 단점이 있다. ‘어떻게 하면 위와 같은 부동산 기본정보 뿐만 아니라 이용자 각자가 추구하는 부동산 정보를 100% 충족시킬 수 있을까?’ 에 대한 생각을 하며 다음과 같은 프로젝트를 기획하였다.

(2) 활용성 및 기대효과

**[프리미엄 회원을 위한 서비스 개발]**

이용자들 각자가 원하는 부동산 정보들을 100% 충족시키기 위한 아이디어로 의뢰 방식의 웹 사이트를 제작하였다. 해당 방식으로 부동산 정보 제공 서비스 업체들이 프리미엄 회원을 위해 부동산 정보 의뢰 방식을 도입하여 이용자들이 원하는 정보를 100% 충족 할 수 있는 서비스를 만드는 기대효과를 가졌으면 한다.

**2. 과제수행 과정**

(1) 과제수행 과정

한명의 고객이 19년 1월부터 20년 9월까지 충청도의 아파트 매매 건수 상위 20개와 충청도 아파트의 가격 예측을 요청하는 상황을 가정하여 프로젝트를 진행하였다. 데이터는 공공데이터의 국토교통부\_아파트매매 실거래자료 Open API XML 형식의 자료를 기반으로 데이터를 추출했다. 19년 1월부터 20년 9월까지의 아파트 매매 건수 상위 20개의 정보는 파이썬의 pandas 라이브러리를 통해 데이터를 전처리를 했으며, plotly 함수를 통해 시각화를 하여 상위 20개의 데이터를 그래프로 구현하였다.

충청도 아파트 가격을 예측하는 정보는 파이썬 머신러닝 라이브러리인 사이킷런을 활용하였으며, 처음에는 LinearRegression() 모델을 사용하여 회귀분석을 실시했으나 훈련데이터의 오차를 줄이는데 급급한 나머지 과적합이 발생하여 Ridge(), Lasso(), ElasticNet() 모델을 사용하여 규제를 주어 해당 문제를 해소하였다. 또한 최적의 하이퍼 파라미터를 찾기위해 GridSearCV 클래스를 사용하여 5폴드 교차 검증을 통해 휴리스틱한 파라미터를 구하는 과정과 타겟 데이터와 피쳐 데이터를 로그로 만들어 정규분포로 만들고 이상치를 제거하는 과정을 통해 예측값과 실제값의 오차 평가지표를 지속적으로 줄여나갔다.

(2) 과제 수행의 문제점 및 처리결과

(2) - 1. 타겟 데이터를 거래금액으로 두고 LinearRegression() 모델을 사용하여 다중 선형 회귀분석을 실시한 결과 MAE와 RMSE의 평가지표가 아주 큰 숫자가 나와 Value Type Error를 발생하였다. 이는 위에서도 설명했듯이 훈련데이터에 초점이 맞추어져 과적합이 발생하는 상황이 발생하였으며, L2규제와 L1규제를 적용하여 회귀계수와 오차값의 균형을 맞추는 것으로 회귀의 평가지표를 줄여나갈 수 있었다.

(2) - 2 타겟 데이터와 피쳐 데이터중 수치형 데이터에 대해 로그를 적용하여 정규분포로 만들어 회귀분석을 원활히 할 수 있게 하고, 최적의 하이퍼 파라미터를 구하고, 이상치를 제거하는등 평가지표를 줄여나가기 위해 노력을 했다. RIDGE()의 파라미터 Alpha를 0.5을 넣은 모델의 MSE가 약 3300으로 가장 좋은 수치가 나왔으나 더 이상 이 평가지표를 줄일 수 없었다. 수치형 데이터가 있는 필드를 추가하여 피쳐데이터를 추가해 평가지표를 줄여나가야 하는 상황이며, 기존에 추출한 데이터에 어떻게 새로운 필드를 적용해야 할지가 최대의 관건이다.

(2) - 3 웹사이트의 구성을 처음에는 이메일 구독 기능 하나만을 배치하여 구독한 이용자에게 필요한 사항을 이메일로 질문하는 방식으로 만들었다. 하지만 비효율적이라고 생각하여 웹페이지에 형식 태그를 이용하여 설문지를 배치하는 방식으로 달리 하려고 계획중이며 추후 수정할 예정이다.

(2) - 4 이용자는 우리들이 가지고 있는 기존의 데이터를 기반으로만 질문하는 사람들이 아니기 때문에 기존 데이터에 없는 다른 요구사항을 찾아 달라고 할 수 있다. 기존 데이터 뿐만 아니라 다른 데이터를 어떻게 찾아야 할지 대책을 강구해야 할 것이다.

**4. 운용 및 시험**

(1) 운용 및 시험 요구조건

데이터 전처리 및 예측은 Anaconda3 Jupyter notebook으로 수행하였으며 웹페이지 프론트엔드와 백엔드는 VS Code로 수행하였다. 고객들의 이메일과 세션 로그 기록은 각각 MYSQL DB와 MONGO DB를 이용하였다.

(2) 운용 및 시험결과

데이터 전처리 부분에서는 특별한 이상 없이 해당 조건에 맞는 데이터를 추출하였지만 예측 부분에서는 기존 데이터가 수치형 피쳐 데이터의 부족하여 회귀 평가지표를 좀처럼 줄일 수 없었다. 데이터 전처리 후 여러 가지 모델을 써서 다중 회귀 분석을 수행한 결과 Alpha = 0.5인 Ridge 모델을 사용한 RMSE(Root MEAN Squared Error)가 약 3300정도가 나왔으며, ‘기존 데이터에 어떠한 수치형 피쳐 데이터를 넣어 평가지표를 줄일 수 있느냐?’가 이후 핵심 과제이다.

(3) 과제수행 사진

 

5. 결론

한명의 고객이 19년 1월부터 20년 9월까지 충청도의 아파트 매매 건수 상위 20개와 충청도 아파트의 가격 예측을 요청하는 상황을 가정하여 프로젝트를 진행하였다. 데이터는 공공데이터의 국토교통부\_아파트매매 실거래자료 Open API XML 형식의 자료를 기반으로 데이터를 추출했으며, 19년 1월부터 20년 9월까지의 아파트 매매 건수 상위 20개의 정보는 파이썬의 pandas 라이브러리를 통해 데이터를 전처리를 했다.

충청도 아파트 가격을 예측 수행은 파이썬 머신러닝 라이브러리인 사이킷런을 활용하였으며, RIDGE()의 파라미터 Alpha를 0.5을 넣은 모델의 MSE가 약 3300으로 가장 좋은 수치가 나왔으나 더 이상 이 평가지표를 줄일 수 없었다. 수치형 데이터가 있는 필드를 추가하여 피쳐데이터를 추가해 평가지표를 줄여나가야 하는 상황이며, 기존에 추출한 데이터에 어떤 새로운 필드를 적용해야 할 지 최대의 관건이다.

또한 서비스를 제공하기 위해 웹사이트의 백엔드는 파이썬의 플라스크 프레임워크와 MVC 아키텍처를 기반으로 사용하였으며, 최종적인 웹페이의 형태는 웹페이지에 Form 형식의 설문지를 구현하여 이용자가 어떠한 부동산 정보를 원하는지 조사한 후 메일로 원하는 정보를 제공하는 것을 목표로 구현하고자 하며, 현재는 이메일 구독 기능만을 구현한 상태이다.

과제 수행의 최대 문제점을 지적한다면 이용자들은 우리가 가지고 있는 기존 Open API, XML 형식의 데이터만이 아닌 다른 정보도 알고싶기를 원하는 사람이 대부분일 것이다. 다른 정보를 어디서 어떻게 추출해야 할지가 상당한 미지수이며 해당 방법을 모색하기 위해 노력을 기울여야 할 것이다.