|  |
| --- |
| **제10회「2022 빅콘테스트」데이터 분석 계획서** |

|  |  |  |  |
| --- | --- | --- | --- |
|  | \* 해당란에 ☑ 표시 | | |
| **참가리그** | □ 데이터분석리그 | | |
| **세부분야** | □ 이노베이션분야 □ 데이터분석분야 | | |
| **세부부문** | □ 루키부문 □ 퓨처스부문 □ 챔피언부문 | | |
| **개인/팀여부** | □ 개인 □ 팀(총 5명) | **개인/팀명** | 부릉부릉 |
| **지도교사명** |  | | |
| **대표ID** | dimple0112@yonsei.ac.kr | | |

※ **5장 내외로** 목차는 준수하여 자유롭게 작성

|  |  |
| --- | --- |
| **분석 주제명** | **용인시 전기차 완속 충전소 입지 선정 모델 개발** |
| **분석 배경** | 탄소 중립이 중요해지고 있는 시기에 내연기관 차량에서 전기차 차량으로의 전환은 필수불가결한 변화로 인식되고 있다. 전기차와 내연기관 차량의 가장 큰 차이점은 에너지 공급원으로, 주유소에서 휘발유, 경유, LPG가스등 을 공급받아 운행하는 내연기관 차량과는 다르게 전기차는 충전기를 통해 배터리를 충전해야 한다. 전기차 시장이 활성화되고, 전기차 이용자들의 불편함을 줄이기 위해서 꼭 필요한 일이 적절한 위치에 전기차 충전소를 설치하는 것이다. 용인시 전기차 충전소 최적입지 선정 과제는 불필요한 비용은 줄이면서 이용자들에게 더 큰 효용을 주기 위해 용인시에 설치할 완속 전기차 충전소의 최적입지를 데이터 분석을 통해 선정하는 것이다.  U+에서 제공한 전기차 앱 사용자 데이터는 좌표정보를 지니고 있어 어느 시점에 어디에서 전기차 충전 수요가 발생했는지를 확인할 수 있는 핵심 데이터이다. 거주지 기준, 활동자 기준, 거주자 활동자 기준으로 나뉘어진 데이터는 각 필요에 맞게 수요를 예측하는데 사용된다. 여기에 데이터공공포털 등 다양한 외부 데이터를 활용해 용인시내 아파트, 공영주차장, 다중이용시설, 산업단지 등 전기차 충전기 설치가 용이한 환경에 전기차 충전소를 수요에 맞게 선정하는 것을 목표로 한다. |
| **분석 내용**  **요약** | 용인시 내 전기차 완속 충전소가 자리하기 적합한 후보지를 선정한 뒤, 충전소 수요를 알 수 있는 제공된 전기차 앱 실행 데이터 및 각 후보지의 접근성과 이용성을 파악할 수 있는 추가 데이터를 기반으로 하여 변수를 선택한다. 전처리 과정을 거친 데이터셋으로 다양한 군집화 모델을 실행하고 하이퍼 파라미터를 조정한 뒤 최상의 성능을 보이는 모델을 선정한다. |
| **분석방법 및**  **계획** | 1. 충전소 입지 후보 탐색을 위한 추가 데이터  분석 주제는 용인시 지역에 완속 충전으로 충전할 수 있는 최적 충전 거점을 확보하는 것이다. 그런데 대회 주최측에서 제공한 데이터만으로는 최적의 장소를 찾기 어려워 추가 조사가 필요했다. 경기연구원에서 발행한 논문 “경기도 전기차 충전시설 적정 설치기준 연구”를 보면 공공 충전시설 설치지점 선정기준이 제시되어 있다. 정량적 지표로 충전 수요 발생 가능성이 있고 운전면허소지자수, 아파트 단지수, 종업원 300인 이상 사업체 수 등 사람이 많이 모이는 곳을 충전 수요가 많은 곳으로 선정하고 있다. 정성적 지표에 3번 이용성이 존재하는데 충전량, 충전 횟수 등 이용 선호도가 높은 설치지점으로 아파트 단지, 공공기관, 직장대단위업무시설, 대형쇼핑센터, 산업단지, 체육시설, 공원, 주유소 등이 제시되어 있다. 여기에 제시된 곳들을 최대한 많이 포함하여 충전소 입지 후보를 선정하기로 하여 용인시 거주자의 생활권을 파악하고자 용인시 사이트에서 용인시 내 아파트 단지 현황을 얻고, 도시공사가 제공하는 주차장 현황, 한국석유공사가 제공하는 주유소 가격 정보들을 ‘공공데이터포털’에서 얻고, 대규모 점포 현황과 주차장 현황을 ‘경기데이터드림’에서 얻었다. 중복을 피하기 위해 기존 전기차 충전소 위치를 파악하려고 ‘경기데이터드림’에서 충전소 현황을 얻었다. 그리고 사설주차장 데이터는 구글에서 크롤링하여 얻었고 용인시의 산업단지 데이터는 구글지도를 보고 엑셀로 한 행씩 입력하였다.   * 1. 충전소 입지 후보 탐색을 위한 데이터 전처리   추가데이터 자료의 주소 데이터를 기반으로 카카오 API를 이용하여 위도, 경도를 얻는 과정에서 오류가 생기는 주소를 구글 지도와 도로명 주소 사이트로 확인하여 수정했다. 그리고 모두 문자로 되어 있고 의미가 없다고 생각되는 열은 제거하였다. 용인 대규모 점포에서 점포 면적을 알 수 없었던 행은 나머지 점포들의 평균값으로 대체하였다**.**  2. 충전 수요 예측을 위한 추가 데이터  앞서 살펴본 후보들 중 최적 입지를 선정하기 위한 모델을 개발하기 위해 사용할 데이터는 다음과 같다.   1. 용인시 내 기존 전기차 충전소 데이터(출처: 공공데이터포털) 2. 용인시 아파트 데이터(출처: 공공데이터포털) 3. 용인시 법정읍면동별 전기차 보급 대수 데이터(출처: 예정) 4. 용인시 법정읍면동별 자동차 보급 대수 데이터(출처: 예정) 5. 기타 추가로 필요하다고 판단 되는 데이터   이번 데이터 분석 과제에서 충전기 수요를 예측하기 위해 사용할 데이터 중 가장 중점을 두고 살펴볼 데이터는 빅콘테스트에서 제공한 전기차 앱 사용자 수 데이터이다. 전기차 충전 수요를 가장 잘 반영함과 동시에 위경도 및 행정구역 정보도 담고 있어 위치 별 수요를 예측하는 목적에 부합한다. 주어진 3가지 데이터는 각각 거주자 기준, 활동지 기준, 거주자 활동지 기준으로 나뉘는데, 주거단지 지역의 수요를 파악할 때는 거주자 기준 데이터를 중점적으로 활용할 예정이다. 다중이용시설, 산업단지 등 비거주지 지역의 수요를 파악할 때는 활동지와 거주자 활동지 기준 데이터를 중점적으로 활용할 예정이다.  모델링을 통해 얻은 최적입지와 기존 충전소 입지를 비교해 같은 지역 혹은 기준 거리 이내에 기존 충전소가 있다면 다른 입지를 새롭게 선정할 수 있다. 이 과정에서 필요한 데이터가 용인시 기존 전기차 충전소 데이터이다.  용인시 법정읍면동별 자동차 보급 대수와 전기차 보급 대수는 미래의 전기차 충전수요에 어떤 변화가 발생할지 예측하기 위한 데이터이다. 법정읍면동별 전기차 대수/법정읍면동별 자동차 대수를 통해 현재 전기차 보급률이 지역별로 어느정도 인지 파악할 수 있다. 다양한 원인에 의해 지역별로 보급률이 다르므로, 미래 수요를 예측하기 위해서 이 비율을 활용할 예정이다. 가령 타지역보다 보급률이 낮은 지역은 앞으로 전기차 보급률이 빠르게 상승할 가능성이 높다고 볼 수 있다. 그렇지 않더라도, 만약 해당 지역의 충전 인프라의 부족이 전기차 보급률이 낮은 원인이라면, 보급률을 반영한 수요예측을 통해 해당 지역의 충전 인프라를 보급해 전기차 보급률을 상승을 이끌어낼 수도 있다.  2.2 수요 예측을 위한 데이터 전처리  우선 거주지역에 전기차 충전소를 보급할 때 가장 먼저 고려해야 할 부분은 주차장 접근성이다. 특히 최근 아파트들은 대부분 입주민을 제외한 차량의 출입을 통제하고 있어 수요예측 시 아파트 단지 내 충전소의 수요는 해당 아파트 주민들에 한해서 발생하는 점을 반영해야 한다. 이에 거주자 기준 전기차 앱 사용자 데이터의 좌표데이터를 통해 해당 데이터가 아파트 거주민의 수요인지, 비아파트 거주민의 수요인지 구분하는 작업이 필요하다. 이를 위해 카카오API를 활용, 좌표데이터를 기반으로 주소를 얻은 뒤, 얻은 주소데이터와 용인시 아파트 데이터의 주소를 비교했다. 이 과정을 통해 약 29만개의 데이터 중 약 16만개의 데이터가 용인시 아파트 거주민의 수요데이터임을 확인할 수 있었다.  활동지 기준 데이터는 특별한 추가 정보 없이 그대로 활용할 예정이고, 거주자 기준 활동지 데이터는 용인시민들이 용인시 내에서 충전소를 이용하는지, 외부에서 충전소를 이용하는지를 확인하기 위한 용도로 활용한 후, 용인시내 수요예측을 위해 용인시로 구분되는 데이터를 활동지 데이터와 함께 비거주지 지역의 수요예측을 위한 데이터로 활용할 예정이다.  3. 모델 구축 및 학습  수요예측 모델을 구축하기 위해 우선 법정읍면동별 전기차 등록대수/법정읍면동별 자동차 등록대수로 얻은 법정읍면동별 전기차 보급률을 계산한다. 전기차 보급률을 가중치로 두고 전기차 앱 사용자 수 데이터를 통한 수요예측 모델을 구현한다. 이때 아파트 거주민 수요는 제외한 후 모델을 구축할 예정이다. 수요 예측을 위해 고려하고 있는 방안은 크게 2가지로, 첫번째는 용인시를 가상의 격자로 나누어 수요를 예측하는 방안이다. 가상의 격자 안에 분포하는 수요데이터를 수합하고, 이를 클러스터링 기법을 도입해 전기차 충전 수요를 가장 잘 나타내는 군집점들을 산출한다. 이 점들과 충전소 입지 후보들을 비교해 가장 가까운 지점을 충전소 최적입지로 선정한다. 이때 선정된 최적입지를 기존 충전소 입지와 비교해 해당 지역의 수요를 기존 충전소가 충족한다면 추가적인 설치가 불필요하다는 결론을 내리고, 만약 수요를 충족하지 못한다면 새로운 충전소를 설치하는 결론을 내린다. 혹은 앞선 방법과 반대로 기존 전기차 충전소를 기반으로 격자점 내부의 수요를 클러스터링 기법을 통해 군집점을 얻은 뒤, 각 군집의 수요 중 기존의 전기차 충전소가 충족하는 수요를 제외하고 남은 수요를 토대로 최적입지를 선정하는 방법 또한 사용 가능하다.  가상의 격자로 나누지 않고 수요를 예측하는 방안도 고려하고 있는데, 이러한 이유는 가상의 격자로 나누면서 발생하는 정보 단절을 방지하기 위함이다. 다만 가상의 격자로 나누지 않고 클러스터링 기법을 도입하기 위해선 몇 개의 군집점을 사용해야 수요를 적절히 분배할 지 예측하기가 상당히 어렵다는 단점이 있다. 이 방안을 도입하기 위해선 하이퍼파라미터인 군집점의 개수를 튜닝하는 과정이 무엇보다도 중요하고 많은 시간이 필요하다. |
| **분석결과 활용 및 시사점** | 현재 보급된 전기차수에 비해 설치된 완속 충전소의 대수는 부족할 뿐 아니라 지역 내에 불균형하게 배치되어 있기 때문에 해당 분석은 이러한 문제를 해결하는 데에 중요한 열쇠가 될 것이다. 이 모델을 기반으로 완속 충전소들을 설치한다면 각 충전소로의 접근성을 최대한으로 향상시킴으로써 전기차를 이용하는 시민들의 삶의 질 향상에 기여할 수 있을 것으로 기대된다. 더 나아가, 높아진 접근성은 이용률 증가로도 이어질 것이기에 방치되어 낭비되는 전기 에너지를 최소화하는 효율적인 운용이 가능해질 것이다.  또한, 본 분석은 특정 지역의 고유한 특성을 고려하는 입지선정방향과 분석 모델을 제시한 것으로 데이터 수집과 전처리가 준비되어 있다면 다른 지역에도 이 모델을 적용하여 전기차 완속 충전소의 건설에 많은 도움을 줄 수 있을 것이다. 관점을 달리 하여 보면, 각 지방자치단체에서 본 모델은 전기차 충전소 대수의 부족 및 불균형한 배치를 보완하고자 새로운 충전소들을 설치하고자 할 때의 좋은 근거 자료로서 활용될 수 있을 것이기에 행정 업무상으로도 유용하다고 할 수 있다. 수요와 공급을 파악해 시설 필요 지역을 파악하는 모델 역시 급속 및 중속 충전소와 같은 다른 종류의 충전소로도 확장시켜 적용할 수 있을 것으로도 기대된다. |

※ 제출자료는 평가에 반영 예정