**예측모델 기말프로젝트 중간점검**

1. **프로젝트 제목**

서울특별시 공공자전거 따릉이 대여소 유형별 일별 대여 건수 예측

1. **조원**

|  |  |  |
| --- | --- | --- |
| 학번 | 이름 | 학과 |
| 2020011135 | 소규성 (팀장) | 산업경영공학과 |
| 2020011132 | 정의석 | 산업경영공학과 |
| 2020011136 | 김지나 | 산업경영공학과 |

1. **프로젝트 주제**

서울시 공공데이터 중 공공자전거 이용정보(시간대별)[[1]](#footnote-1)를 활용해 (출근용/레저용/장거리용) 등 대여소를 유형화하고, 각 유형별로 미래의 일별 대여 건수를 예측한다.

1. **프로젝트 현재까지 진도상황**

|  |  |  |
| --- | --- | --- |
| **태스크** | **상세설명** | **진도** |
| 주제 설정 | KAGGLE, 공공 데이터 등을 탐색하여 시계열 모형(ARIMA 등)을 적합하기에 좋고 분석에 의미가 있을 만한 주제 탐색 | 100% |
| 데이터 수집 | 서울특별시 공공데이터 홈페이지 내 “공공자전거 이용정보" 데이터 수집 완료  - 월별로 각 대여소에서 대여된 공공자전거의 이용정보 포함  - 고객정보(성별/연령대) 정보 또한 포함하고 있어 유형화에 유의미한 변수로써 활용 가능 | 100% |
| 데이터 파악(EDA) | 예측을 수행하기에 앞서 데이터의 형태, 분포 등을 파악하여 분석 방향성 및 주제 구체화 | 70% |

1. **앞으로의 계획**

[Task 별 상세설명]

|  |  |
| --- | --- |
| **태스크 분류** | **상세설명** |
| 데이터 파악 (EDA) | 예측을 수행하기에 앞서 데이터의 형태, 분포 등을 파악하여 분석 방향성 및 주제 구체화 |
| 분석 방향성 수립 | 데이터 탐색을 바탕으로 이용 패턴별 군집화의 필요성 확인  이용 시간을 기준으로 대여소 클러스터링 후, 클러스터 별 대여량 예측 |
| 분석 수행 | K-means, hierarchical, DBSCAN 등의 기법으로 clustering 후 클러스터별 대여량 예측을 위한 ARIMA, Support vector regression 등의 모델 구축. 2-stage로 진행 |
| 결과 산출 및 인사이트 도출 | 대여소 유형별 미래 시점의 대여 건수를 예측하고, 유형별 이용 행태가 미래에 어떤 차이를 보일 지 파악하여 공공자전거 배치 등 현실적 문제에 확장 가능한 인사이트 도출 |

[Task 로드맵]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **태스크 분류** | **11월 3주** | **11월 4주** | **12월 1주** | **12월 2주** | **12월 3주** |
| 데이터 파악 (EDA) |  |  |  |  |  |
| 분석 방향성 수립 |  |  |  |  |  |
| 분석 수행 |  |  |  |  |  |
| 결과 산출 및 인사이트 도출 |  |  |  |  |  |
| 발표 (Presentation) |  |  |  |  |  |

1. https://data.seoul.go.kr/dataList/OA-15245/F/1/datasetView.do# [↑](#footnote-ref-1)