|  |  |  |  |
| --- | --- | --- | --- |
| **과정명** | **빅데이터 기반 AI 개발 전문가** | | |
| **팀명** | 3조 | **팀장** | 이보윤 |
| **팀원** | 박재현, 김동욱, 조재경, 최서윤 | | |

|  |  |
| --- | --- |
| **프로젝트명** | 머신러닝을 활용한 선박 대기 시간 예측을 위한 선박 항차 데이터 분석 |
| **분석 배경** | **❑ 분석 배경**  **코로나19 이후 해운 물류 현황**  - 코로나19 대유행으로 세계 각국 항만 운영에 대한 제약 발생  - 물동량 증가로 선박 대기 시간이 길어지고, 이로 인한 물류 지연 발생에 따른 새로운 해결방안 제시 필요  **항만 정체의 심각성**  - 해운 물류 분야에서 항만 정체 심각성 대두 (세계 물동량의 85%)  - 선박 대기 시간 문제는 물류 지연 야기 → 비용 증가의 원인  **환경적 영향**  - 선박 대기 시간 증가로 연료 소비 증가 → 온실가스 배출량 증가로 직결  - 선박 대기 시간 감소로 환경에 긍정적 영향 |
| **전체 일정** | [2024.04.01 ~ 2024.04.08]  - 프로젝트 주제 선정 및 데이터 수집(1일) : 2024.04.01  - 데이터 전처리(2일) : 2024.04.02~2024.04.03  - EDA 및 특성추출(1일) : 2024.04.04  - 모델 생성 및 검증(3일) : 2024.04.05~2024.04.07  - 발표(1일) : 2024.04.08 |
| **역할 분담** | **❑ 데이터 전처리 및 분석 모델링**   |  |  | | --- | --- | | 역할 | 담당자 | | - 데이터 수집  - 데이터 전처리  - 특성 선택 및 추출  - 머신러닝 모델 선택  - 모델 훈련  - 모델 평가  - 모델 개선 | 전원  김동욱  박재현  이보윤  조재경  최서윤 |   **❑ 서류작업 및 발표**   |  |  | | --- | --- | | 역할 | 담당자 | | 프로젝트 수행계획서 작성 | 이보윤, 최서윤 | | 프로젝트 회의록 작성 | 이보윤, 최서윤 | | 프로젝트 발표 PPT 제작 | 박재현, 조재경 | | 발표 | 김동욱, 박재현, 조재경 | |
| **비즈니스 이해 및**  **수행 목표** | **❑ 프로젝트 시나리오**  [그림 1] 해운 물류 항만 항차 프로세스    [그림 1]의 해운 물류 항만 항차 과정 중 항구 입항 과정에서 선박이 항구에서 대기하는 시간을 예측하여 물류 및 운송 효율성 향상 목표  [그림 2] 분석 시나리오    [그림 2]의 분석 시나리오를 토대로 분석 모델링을 진행 예정  - 데이터 수집 : 선박 항차 데이터 수집 및 저장  - 데이터 전처리 : 수집된 데이터 전처리 → 모델 학습에 적합한 형태로 가공  - 특성 추출 : 수집된 데이터로부터 유의미한 특성 추출  - 특성 변환 : 추출된 특성을 모델이 이해할 수 있는 형태로 변환  - 학습 : 전처리된 데이터를 사용하여 머신러닝 모델 학습  - 검증 : 학습된 모델의 성능을 검증 및 평가  **❑ 프로젝트 아키텍처**  [그림 3] 프로젝트 아키텍처    **❑ 프로젝트 수행 목표**  **1. 대기 시간 최적화**  선박의 항구 대기 시간 최소화를 통한 물류 및 운송 효율성 향상 목표  **2. 정확한 예측 제공**  다양한 요인을 고려하여 신뢰할 수 있는 선박 대기 시간 예측 모델 개발로 항구 운영자 및 물류 관리자에게 정확한 정보 제공  **3. 효율적인 자원 할당**  선박 도착 및 출발 시간을 더욱 정확하게 예측함으로써 항구 내 자원의 효율적 할당 및 운영 비용 절감  **4. 서비스 수준 향상**  물류 및 운송 관련 기업 및 단체들에게 대기 시간 예측 서비스를 제공함으로써 고객 만족도 및 서비스 수준 향상  **5. 장기적인 이점 확보**  정확한 대기 시간 예측을 통해 더욱 효율적인 운송 및 물류 체인을 구축함으로써 장기적인 경쟁력 확보 및 기업 가치 향상 |
| **수집**  **데이터 목록** | **❑ HD현대 제공**   |  |  |  | | --- | --- | --- | | **Feature Name** | **Description** | **단위** | | ARI\_CO | 도착항 소속국가 |  | | ARI\_PO | 도착항 항구명 |  | | SHIP\_TYPE\_CATEGORY | 5대 선종 |  | | DIST | 정박지와 접안지 사이 거리 | km | | ATA | 실제 정박 시간 | hour | | ID | 선박 식별 일련번호 |  | | BREADTH | 선박의 폭 | m | | BUILT | 선박의 연령 | year | | DEADWEIGHT | 선박의 재화중량톤 수 | ton | | DEPTH | 선박의 깊이 | m | | DRAUGHT | 흘수 높이 | m | | GT | 용적톤수 | GT(m^3) | | LENGTH | 선박의 길이 | M | | SHIPMANAGER | 선박 소유주 |  | | FLAG | 선박의 국적 |  | | U\_WIND | 풍향 u벡터 | m/s | | V\_WIND | 풍향 v벡터 | m/s | | AIR\_TEMPERATURE | 기온 | ℃ | | BN | 보펀트 풍량 계급 |  | | ATA\_LT | 현지 정박 시간 | hour | | PORT\_SIZE | 접안지 폴리곤 영역의 크기 | Km^2 | | CI\_HOUR | 대기시간 | hour | |
| **데이터**  **확보 방안** | **❑ 데이터 수집처**  - DACON(인공지능 경진대회 플랫폼)  **❑ 수집기간**  - 2014.09 ~ 2023.02  **❑ 데이터 가격**  - 무료 |
| **분석 방안**  **수립** | **❑ 사용할 패키지**  Pandas, Numpy, Matplotlib, scikit-learn, MariaDB 등    **❑ 모델 구축을 위한 단계적 분석 절차**  **1. 데이터 전처리**  - 수집한 데이터 정제 및 전처리  - 결측치, 이상치, 중복치 처리 후 필요한 데이터 형식으로 변환  - 날짜 및 시간 데이터 → datetime 형식 변환  - 범주형 데이터 → 원-핫 인코딩 또는 레이블 인코딩 변환  **2. 특성 공학**  - 선박 대기 시간에 영향을 미치는 요인을 고려하여 새로운 특성 생성 또는 기존 특성 변환  - 상관성이 없는 특성 제거하여 차원 축소  **3. 모델 선택**  - 대기 시간 예측에 적합한 머신러닝 모델 선택  - 회귀 문제로 접근하므로 선형 회귀, 의사결정 트리, 랜덤 포레스트, 그래디언트 부스팅 등의 모델 고려  **4. 모델 학습**  - 선택한 모델 학습  - 학습 데이터 사용 → 모델 훈련  - 검증 데이터 사용 → 모델 성능 평가  **5. 모델 평가**  - 학습된 모델 성능 평가  - 평가 지표 → 평균 절대 오차(MAE),  **6. 모델 튜닝**  - 모델 성능 향상을 위한 하이퍼파라메터 튜닝 및 모델 구조 수정  **7. 최종 모델 선정 및 예측**  - 평가 지표를 토대로 MAE 수치가 가장 낮아지는 방향으로 최적의 모델 선정 및 예측 |
| **분석 결과**  **활용 방안** | **❑ 기대 효과**  **1. 운영 최적화**  - 선박 운항 일정 최적화  - 예측된 대기 시간을 고려하여 출발 시간 조절 및 항로 변경으로 선박 운항에 효율성 극대화  **2. 리스크 관리**  - 대기 시간 예측을 통해 지연으로 발생하는 비용 감소  **3. 자원 할당**  - 예측된 대기 시간에 따라 인력, 연료 등의 자원을 적절하게 배분함으로써 비용 절감 및 생산성 향상  **4. 고객 서비스 향상**  - 예측 결과 고객 서비스에 활용  - 정확한 예측을 통해 고객들에게 출발 및 도착 시각을 더 정확하게 안내함으로써 고객 만족도 향상  **5. 정책 개발**  - 항로 혼잡도 예측, 이를 기반으로 한 정책 개발로 선박 운영 환경 개선 |