< KDT 6기 AI 학습 모델 웹서비스 개발 프로젝트계획서>

1. 프로젝트 개요

1-1. 제목 : Port-Luck(울산항 선박 입출항 기록 및 차량 입출문 기록을 활용한 운송 차량 규모별 화물 적 재 및 입출문 소요 시간 예측 서비스)

1-2. 팀명 : 울지마

1-3. 팀원: 곽동현, 김도영(팀장), 김민정

2. 프로젝트 필요성

2-1. 프로젝트 추진 배경

- (1) 추진 배경
 - ① 울산 본항 선박 화물 운송 목적 출입 화물차량의 실제 화적 시간과 대기 시간의 괴리로 불편 발생
 - ② 울산 본항 출입 및 화물 적재를 위한 차량의 대기 행렬로 항만 주변 도로 교통 체증 문제 심각
 - ③ 선박 입출항 및 화물차 항만 출입 데이터 분석 및 차종별 소요 시간 예측 후 운전자에 정보를 제공하여 불편 감소를 목표

(2) 필요성

- 울산항만공사 PortWise
 - ① 항만 모니터링 등 입항 선박 관련 데이터에 서비스가 집중되어 있기에, 화물차량의 운송 계획 경향에 대한 직접적인 예측 자료로 사용될 수 없음
 - ② 항만 데이터 플랫폼으로 활용되고 있기에 해운 물류 기업은 서비스의 직접적인 수혜 대상이 될 수 있으나. 화물 차량 운전 기사에게는 서비스의 혜택이 간접적임
- 부산항만공사 차량 반출입예약시스템(VBS, Vehicle Booking System)
 - ① 트럭 운전자가 사전에 반출입할 화물과 시간을 예약하는 시스템으로써 예약을 바탕으로 터미널의 차량 집중 시간을 예측할 수 있으나 예약이 이루어지기 전에는 특정 시점에의 경향성은 예측이 불가함 ② 대기하는 화물 차량의 규모(소형, 중형, 대형 등)에 따라 화물 적재 시간이 차이가 나는 점을 반영하지 못하고 있으며 이는 불필요한 교통 체증으로 이어짐
- 따라서 운송사, 화물차 기사, 항만 관계자들에 소요 시간 예측 서비스를 제공하여 운송/하역 계획 수립 시 활용 가능

2-2. 프로젝트 목적

- (1) 데이터 분석 목적 : 울산 본항 출입 차량 소요 시간 예측
- (2) 기대 성과 : 입문 후 출문까지의 소요시간 예측정보 제공을 통해 대기시간으로 인한 불편 최소화

3. 요구사항 수집 및 분석

- 3-1. 웹서비스 기능 요구사항
- (1) 시간대별 / 차종별 소요 시간 예측 및 안내
- (2) 시간대별 / 차종별 소요 시간 분포 경향성 예측 및 안내 ex) 목요일 오후 4시에는 5톤 이상 트럭의 통행량이 多 → 소요 시간 70분 증가 예상됨
- (3) 회원가입 / 로그인 / 임시 비밀번호 발급 / 비밀번호 변경 기능
- (4) 게시판 내 유저용 질문 판 / 자기 질문 조회 / 관리자 답변 및 질문 삭제 권한
- (5) 울산 본항 주변 주유소 정보 안내 및 카카오맵 api 연동
- (6) (1)-(5)의 기능 모바일 웹에서 동일하게 구현
- 3-2. 데이터 분석 요구사항
- (1) 울산 본항 선박 입/출항 기록 분석
- (2) 울산 본항 화물 운송 차량 입문/ 출문 기록 분석
- 선박의 입항 및 출항 시간 동안의 화물차 데이터 수집 및 분석
- 화물차의 입문 시간, 출문 시간, 차종 등의 데이터를 기반으로 예측 모델 구축

4. 수집 데이터

4-1. 출처

- (1) 해양수산부 해운항만물류정보시스템(PORT-MIS)
- (2) 울산항만공사
- 4-2. 데이터 세트 건수
- (1) 울산항 선박입출항현황(210401-230331)
- (2) 울산항만공사_최근 2개년 화물차 항만 출입 데이터
- (3) 해양수산부_선박운항정보

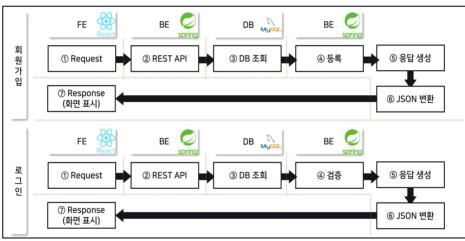
4-3. 주요 내용

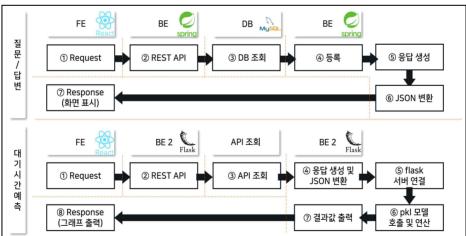
- (1) 울산항만공사 : 최근 2개년 화물차 항만출입데이터(23.04.30 기준)
 - : 화물차 항만 입출문 기록, 화물차 차종 등
- (2) 해운항만물류정보시스템(PORT-MIS)
 - : 선박 양/적하 기록, 입항 목적 등(단순 경유 및 급유는 제외)

5. 해결 현안 분석 및 주요 설계

5-1. top-down 분석 내용

<프레임워크>





5-2. 서비스 설계

- PC/모바일 서비스 설계 : 웹 및 모바일 화면 설계, 로그인, 게시판, 분석 결과 가시화
- (1) 로그인 및 회원 가입 후 중앙의 버튼 클릭 시 예측 모델 결과값 시각화 화면 전환
- (2) api의 데이터를 가져와 실시간 입/출문 소요시간 예측 그래프 시각화
- (3) DB 설계 : 웹 서비스에 필요한 테이블과 필드를 정의
- (4) 회원가입, 로그인, 게시판, 임시비밀번호 발급 및 비밀번호 변경 기능 구현

① member : 회원가입을 위한 테이블

- 사용자마다의 권한 구별 (role, enabled - ADMIN, USER)

② question : 질문을 게시판을 위한 테이블

- 전체 질문/사용자가 등록한 질문을 username(id로 쓰임)으로 구별

- 답변이 완료된 질문과 아닌 질문을 구별(answered)

- 질문이 등록된 순서대로 자동으로 부여되는 id를 통해 구별

③ answer : 답변을 위한 테이블

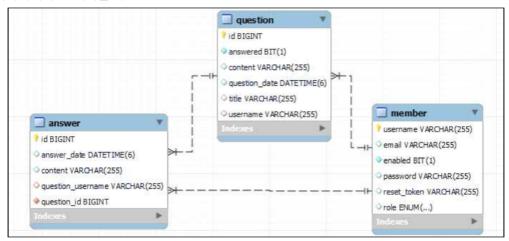
- 답변이 완료된 질문과 아닌 질문을 구별(answered)

* 구별 :

질문한 사용자의 username을 기반으로 구별됨

질문의 id를 기반으로 관리자 답변 시 개별적인 값을 불러옴

④ 데이터베이스 테이블 구조



5-3. 데이터 학습 모델 설계

사용 데이터

- (1) 울산항 선박입출항현황(210401-230331)
 - ① 데이터 개요
 - 시스템명 및 데이터 출처: 해운항만물류정보센터(new.portmis.go.kr)
 - 내용 : 21년도 4월 01일부터 23년도 3월 31일까지의 울산항 선박입출항현황
 - ② 데이터 활용
 - 활용 빈도 : 선박의 척 수와 입/출항시간, 화물차의 입/출문시간의 관계를 학습시키는 데 사용됨. 모델 학습 시 사용하며 모델 완성 후에는 사용 예정 없음
 - 비중 : 화물차의 예측 체류시간(입문과 출문까지)을 구하는 데 핵심적인 역할
 - 중요성 : 화물차와 선박의 연관관계를 예측에 필수 데이터
- (2) 데이터 획득
 - 지속성 : 해운항만물류정보센터의 지속적 업로드를 통한 연속적인 데이터 확보 가능
 - 활용범위 : 화물차 항만 출입 데이터와 연결하여 화물차의 시간별 체류시간을 학습
 - 가공 가능성 : 계선 장소 기반 울산 본항의 부두 데이터만 활용. 울산 신항 및 기타 울산항 내 모든 항구 활용 시, 각 항의 거리가 있어 예측값의 정확도/신뢰도 하락. 또한 선박 화물 종류별 고유 번호 부여 후 number 열을 추가하여 선박을 용도 별로 라벨링함. 출항일시 열을 datetime 형식 변환하여 유효한 날짜만 년도, 월, 일, 시간, 분으로 필터링 후 필요한 열만 선택하여 정리

<사용 부두 목록(울산 본항)>

순서	부두명	순서	부두명	순서	부두명	순서	부두명
1	1부두 01	13	6부두 04	25	SK5부두	37	일반부두 04

2	2부두 01	14	6부두 05	26	SK6부두	38	일반부두 05
3	2부두 02	15	7부두 01	27	SK7부두	39	일반부두 06
4	2부두 03	16	8부두 01	28	SK8부두	40	일반부두 07
5	3부두 01	17	8부두 02	29	SK부이 02	41	일반부두 08
6	3부두 02	18	9부두 01	30	SK부이 03	42	자동차부두 01
7	4부두 01	19	SK1부두 11	31	UTT부두	43	자동차부두 02
8	4부두 02	20	SK1부두 12	32	가스부두	44	자동차부두 03
9	5부두 01	21	SK2부두 01	33	남화부두	45	염포부두 01
10	6부두 01	22	SK2부두 02	34	일반부두 01	46	염포부두 02
11	6부두 02	23	SK3부두	35	일반부두 02	47	염포부두 03
12	6부두 03	24	SK4부두	36	일반부두 03		

<선박 종류별 변수로 사용할 고유 번호 부여>

번호	선박 종류	번호	선박 종류	번호	선박 종류
1	LNG 운반선	10	산물선(벌크선)	19	일반화물선
2	LPG 운반선	11	석유제품 운반선	20	자동차운반선
3	견인용예선	12	선박용도	21	철강재 운반선
4	관공선	13	세미(혼재)컨테이너선	22	케미칼 운반선
5	급유선	14	시멘트운반선	23	케미칼가스 운반선
6	기타 예선	15	신조선	24	폐기물 운반선
7	기타 유조선	16	압항 예선	25	풀컨테이너선
8	기타선	17	여객선		
9	모래운반선	18	원유운반선		

이 외에 입항일시와 출항일시가 비정상적인 데이터 중 상/하위 15% 삭제

(2) 울산항만공사_최근 2개년 화물차 항만 출입 데이터

① 데이터 개요

- 시스템명 및 데이터 출처: 공공데이터 포털_울산항만공사

- 내용 : 21년 4월 1일부터 23년 4월 30일까지의 울산항 초소 입/출문 현황

② 데이터 활용

- 활용 빈도

: 선박 입/출항 일정과 화물차 입/출문의 관계를 모델에 학습시키는 데 사용 후 별도 사용 용도 없음

- 비중 : 화물차의 예측 체류시간(입문 후 출문까지의 소요 시간)

- 중요성 : 다양한 차량이 입/출문까지 걸리는 시간을 예측하기에 모델 학습의 필수 데이터

<차량 무게별 크기 분류 (단위:ton)>

크기 분류	해당 열 이름 : 내용				
1	2.5톤 미만 해당 열의 차명 모두 1톤 미만				
소	1톤 미만				
_	2.5톤 이상				
중	1톤 이상-5톤 미만				
	5톤 이상				
대	컨테이너 차량				
	특수차량(중장비)				

③ 데이터 획득

- 지속성 : 23년 3월 31일 기준 2개년 데이터로 한정

- 활용범위 : 화물차와 선박의 관계를 예측하는 모델 학습시키는 데 사용

- 가공 가능성
- : 차량 입출문 가능 초소명 중 울산 본항만 사용
- : 차종의 개수를 줄여서 1~5톤 등 범위로 줄이고, 대-중-소 분류 열을 만들어 차량을 분류
- : 차명에 오타 및 똑같은 차량인데 다르게 분류된 부분 정리
- : 출문시각에서 입문시각을 감산하여 화물차의 체류시간을 계산하여 새로운 열로 추가
- : 데이터 전처리를 통한 예측 정확도 향상 가능 (입/출문 시간 누락 데이터 삭제, 화물차량 규모 라벨링 등)
- : 입문시각, 출문시각 열의 비어있는 행 삭제
- : 실제 하역 과정 절차의 시간이 최소 20-30분 소요이기에 예측 정확도/신뢰도 향상을 위해 입문과 출문 시각의 차이가 너무 작은 하위 15% 행 삭제

(3) 해양수산부 선박운항정보

- ① 데이터 개요
- 시스템명 및 데이터 출처: 공공데이터_포털_해양수산부
- 내용 : 조회 시작/종료 날짜, 항구를 입력했을 때 선박의 일정 공지
- 획득 방법 : 공공데이터 포털 활용 신청을 통한 인증키 발급
- ② 데이터 활용
- 활용 빈도 : 화물차의 입항 월, 일을 입력받고 해당 일자에 입항하는 선박의 척 수 카운트
- 비중 : 입항 선박의 수가 증가할수록 항구에 출입하는 차량의 대수 역시 비례하여 증가하기에, 예측 과정에서 선박의 척 수는 필수 요소임.
- 중요성 : 실시간 조회를 통해 실시간으로 예측하기 위한 필수 데이터
- 데이터 변수명(feature값) 및 예측값(target명)

변수명	입문시각_연도	입문시각_월	입문시각_일	입문시각_시간	차종	선박_갯수
target명	걸린시간					

데이터 전처리 및 모델 학습

(1) 데이터 전처리

- ① 울산항 중 울산 본항으로 차량 출입 초소명과 선박 입출항 부두 수를 제한 : (울산항으로 검색했을 때 울산본항, 울산신항, 온산항, 미포항 등 울산의 모든 항구가 포함됨. 이에 <laidupFcltyNm>을 사용하여 부두를 구별하고, 울산 본항에 해당하는 부두에 입/출항하는 선박만 카운트 (제1부두, sk1부두, 그 외 기타 등)
- ② 차종 및 차명 재라벨링 후 이에 기반하여 대-중-소 열을 추가로 만들어 차량을 분류
- ③ 누락 데이터의 열 개수 확인 후 삭제 및 상/하위 15%인 차량 입/출문 시각 값 삭제
- ④ 선박의 화물(석유, 차량, 컨테이너, 동/식물 등)에 따라 하역하는 속도 및 통관절차에 걸리는 소요 시간의 차가 있으므로 선박의 용도별 가중치를 적용

<선박 용도별 가중치>

선박 종류	가중치	선박 종류	가중치	선박 종류	가중치
LNG 운반선	0.774623	LPG 운반선	0.823479	견인용 예선	0.463642
관공선	1.017279	급유선	0.550433	기타 예선	0.701309
기타 유조선	0.987432	기타선	0.846952	모래운반선	0.404514
산물선(벌크선)	1.028598	석유제품 운반선	0.419844	세미(혼재)컨테이너선	0.624957
시멘트 운반선	0.498442	신조선	0.534387	압항 예선	0.799037
여객선	0.000000	원유운반선	0.769412	일반화물선	0.959023
자동차운반선	0.521344	철강재 운반선	0.926885	케미칼 운반선	0.922364
케미칼가스 운반선	0.796628	폐기물 운반선	1.227192	풀컨테이너선	1.500000

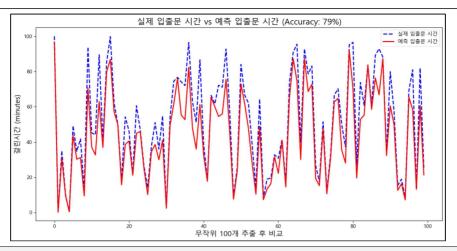
(2) 학습

- ① Scikit-learn의 라이브러리 RandomForestRegressor를 사용한 모델 학습 및 평가
- ② 사용 모델 선정과정

모델비교	RandomForest, XGboost, extraTree							
	RandomForest_model							
선정 이유	초기 예측 시도 중 LSTM 및 GRU 모델을 사용하고자 했으나, GRU 모델 사용 시 학습 시간이 너무 길거나 오류 및 예측의 정확성 측정이 어려움. 이에 시계열 예측이 가능한 대체재를 찾다가 세 머신러닝 모델 중 오차범위가 가장 좁은 randomForest를 예측 모델로 선정							
accuracy (정확도)	3번의 모델 평가를 통해서 train MAE(평균 절대 오차) : 12.89, test MAE : 15.37의 결과를 얻음. 즉 본 내용에서는 54분이 걸린다는 예측에서 약 15.37분의 오차범위를 가짐							
R2 score	0.7984277344							
최적의 hyper- para -meter (hp로 통칭)	n_estimators : 200 - rf모델에서 사용할 트리의 개수 min_samples_split : 2 - 노드 분할을 위한 최소 샘플 수(샘플 2개 이상 분할 시도 설정) min_samples_leaf : 1 - 리프 노드(맨 끝 노드)가 되기 위한 최소 샘플 수 max_features : sqrt - 각 트리 분할 시 고려할 최대 특성의 수 결정('sqrt'는 특성 개수의 제 곱근만큼의 특성을 고려/'log2'나 정수값을 사용하기도 함.) max_depth : 20 - 트리의 최대 깊이(깊어질수록 복잡한 모델이지만 과적합 위험 있음.) bootstrap : False							
	xgboost_model(XGBRegressor)							
accuracy (정확도)								
R2 score	0.6834532454							
최적의 hyper-par a -meter	min_samples_leaf : 1 max_features : sqrt							
	extraTree_model							
accuracy (정확도)	3번의 모델 평가를 통해서 train MAE(평균 절대 오차) : 14.35, test MAE : 16.04의 결과를 얻음. 즉 본 내용에서는 54분이 걸린다는 예측에서 약 16.04분의 오차범위를 가짐							
R2 score	0.7553674335							
최적의 hyper-par a -meter	4. max_features : sqrt							
(3) 축력	5. max_depth : 20 6. bootstrap : False							

(3) 출력

- ① Joblib을 이용한 pkl 파일 형식으로의 모델 저장 및 모델 로드 후 결과값 출력
- ② flask 서버 구현을 통한 프론트엔드와의 연결 및 결과값을 직관적으로 볼 수 있도록 시각화
- ③ 실제값과 예측값 그래프 비교 및 차이



6. 시스템 구성도 설계

6-1. 프로젝트로 구성할 시스템의 모듈(서브시스템) 정의

6-2. 구성할 시스템 구현에 필요한 개발도구/프레임워크를 기술

FE(Front End): react, node js, tailwind, jsBE(Back End): spring boot, MySQL, flask

- DA(Data Analysis) : ipython notebook, python, google colab

6-3. 각 서브시스템(모듈)의 인터페이스 또는 데이터 교환/공유 format을 기술

- 데이터 교환/공유 format : github, google drive, Kakaotalk

7. 구성원 및 역할 분담

이름	역할	설명		
		- 데이터 전처리 및 분석		
곽동현	DA	- 모델링 파일 추출 및 결과값 후처리		
		- FE(flask 연결) 보조		
	FE	- 웹 사이트 개발 및 페이지 구현		
71 E 04		- 모바일 사이트 개발 및 페이지 구현		
김도영		- 일정 관리 및 회의 주관		
		- 보고서		
	BE	- 로그인/회원가입/임시 PW 발급, 변경/QnA 게시판 기능 구현		
김민정		- 데이터(DB), DA 초반 데이터 정리		
		- 보고서, PPT		

8. 프로젝트 추진 일정

날짜	세부 계획
240624	팀 결성, 공모전 정리 및 참여 공모전 결정, 프로젝트 제목 및 프로젝트 목적
240024	정리, 필요 데이터 및 데이터 분석 방향 논의
	공통 : 주제 재선정, 데이터 전처리, 계획서 작성 및 검토, 역할 분담, 화면 구성,
	계획서 작성, 데이터 전처리 완료 및 DB import 시작
240701-0707	FE : UI/UX 디자인 시작(BE와 REST API 상의)
	BE : DB 설계 및 테이블 생성, 로그인, 회원가입 기능 구현, 데이터 전처리
	DA : 데이터 전처리
	공통 : 세부 구현 기능 확정, 참가 서류 작성 완료 및 제출 (7/14)
0.40700.0745	DA : 데이터 분석 및 전처리
240708-0715	BE : DB 관리 및 데이터 저장, 게시판 기능 구현
	FE : 웹/모바일 서비스 화면 구현 및 일정 관리

DA : 모델 구현 및 결과값 출력

240716-0723 BE : BE 기능(게시판, pw 찾기, 삭제 등) 구현 및 수정

FE : 웹/모바일 가시성 개선 및 BE/DA 기능 작동 시연

240724-0726 PPT 및 보고서 작성, 발표 준비 및 작업 데이터 백업

9. 프로젝트 기대 효과

9-1. 웹서비스 기능 / 화면 기대 효과

(1) 서비스

- ① 기존 서비스는 선박의 운영 시간에 집중되어 실질적으로 항구를 이용하는 차량 소지자들에게는 효용이 없었음. 그래서 기존 선박운항 자료와 이전에 고려되지 않았던 차종별 차량 입/출문 자료를 기반하여 서비스 대상을 해운 관계자를 넘어 항구 입출문하는 모든 차량 이용객으로 확장.
- ② 웹 환경 및 모바일 환경에서의 직관적인 화면 디자인을 통하여 서비스 제공 계획.
- ③ 실제 서비스 화면
- 웹 서비스 화면



- 모바일 서비스 화면



- 9-2. 데이터 분석 예측 성과 및 파급효과
- a) 분석 예측 성과
- : 울산항 선박 입출항 기록 및 차량 입출문 기록을 활용한 화물차량 규모에 따른 특정 시점의 화물 적재 및 입출문에 소요되는 시간의 예측 정보를 제공하는 반응형 웹 서비스 개발 (모바일 및 PC 서비스 개발)
- b) 파급효과
- ① 화물차량 운전자 관점 : 불필요한 대기 시간을 줄이고 효과적인 운송 계획 수립 가능
- ② 주변 교통 관점 : 울산항 입문을 위해 대기하는 화물차량 수를 줄여 주변 교통 체증 개선에 기여
- 10. 참고 문헌
- 10-1.

DA	sklearn.metrics	https://white-joy.tistory.com/10
----	-----------------	----------------------------------

	MAE	
	label encoder	https://gongboogi.tistory.com/5
	RandomizedSea rchCV	https://kimmaadata.tistory.com/40
	RandomForest Regressor	https://direction-f.tistory.com/17
	ExtraTree Regressor	https://erdnussretono.tistory.com/49
	모델 학습	랜덤포레스트와 딥러닝을 이용한 노인환자의 사망률 예측(정보처리학회 논문지)
	모델 저장	https://terryvery.tistory.com/63
flask	(REST API)	https://justkode.kr/python/flask-restapi-1/ https://velog.io/@mingming_eee/Flask-day2 https://tutorials.pytorch.kr/intermediate/flask_rest_api_tutorial.html
REACT	UI/UX	- 이인제, ≪소플의 처음 만난 리액트 - 리액트 기초 개념 정리부터 실습까지≫, 한빛미디어, 2022 - 유영진, ≪코드로 배우는 리액트≫ - 회사에서 개발할 때 쓰는 리액트와 스프링 부트≫, 남가람북스, 2023 - 김기수, ≪코딩 자율학습 HTML + CSS + 자바스크립트 - 기초부터 반응형 웹까지 초보자를 위한 웹 개발 입문서≫, 길벗, 2022 - https://mui.com/material-ui/react-rating/ - https://recharts.org/en-US/examples - https://stackoverflow.com/questions/72108791/how-to-use-linear-g radient-in-rechart - https://apis.map.kakao.com/web/sample/ - https://www.flaticon.com/kr/
FIGMA	전체 이미지	- 이상효, ≪Do it! 실무 순서로 배우는 프로덕트 디자인 with 피그마≫,
	구상	이지스퍼블리싱, 2023
	로그인,	- 누구나 끝까지 따라 할 수 있는 스프링부트(채규태, 루비페이퍼)
	회원가입	- 데이터베이스 SQL 프로그래밍「MySQL 실습」MySQL/Workbench
	JWT filter	실습에 의한 DB 실무 능력 (홍봉희, 부산대학교출판문화원)
spring boot	Authorization	 성낙현의 JSP 자바 웹 프로그래밍 2판(성낙현, 골든 래빗) https://github.com/LeeYeongin https://velog.io/@qiwisil_227/%EC%98%88%EC%A0%9C-%EC%BD%94%EB%93%9C-%EB%A1%9C%EC%A7%81-%EC%9D%B4%ED%95%B4%ED%95%98%EA%B8%B0 https://velog.io/@hyex/HTTP-Authorization-header%EC%97%90-Bearer%EC%99%80-jwt-%EC%A4%91-%EB%AC%B4%EC%97%87%EC

개인 깃허브 및 노션 링크

(1) 곽동현: https://github.com/etoto06/KDT6_Al_Project.git (2) 김도영: https://github.com/DoyoungKim273/K6_AlFINAL (3) 김민정: https://github.com/mjmmjmmjjjmj/Port-luck-BE-

(4) 팀 노션(KDT_06_AI):

https://selective-cowbell-c20.notion.site/3eb904f790304cdca80e91ab4139a017?v=0d2421d1a4cd49c18b0311b14f7300c9

(5) 팀 자료 공유용 드라이브 : https://url.kr/2h43n4