

프로젝트 진행 보고서

[작성자]

이름 : 김상겸

휴대폰 : 010-9212-0252

이메일 : vhf1030@naver.com

[기본 정보]

과정명	프로그래밍 기반 얼라인 프로젝트
트랙	데이터/AI
파트	파트2

[프로젝트 개요]

제목	수산물 가격 예측
한 줄 설명	머신러닝을 활용한 시장별 어종별 수산물 가격 예측
배경 또는 기획 의도	수산물 시장의 가격 변동을 예측하여 소비자에게 최적의 구매 시점 판단 기회를 제공
참여 인원	김상겸, 원정환, 지서연
본인 역할	프로젝트 협업 및 버전 관리: GitHub를 활용한 코드 관리 및 팀원 간 협업 데이터 수집: 수산물 가격정보 및 검색 트렌드 데이터 수집 데이터 전처리 및 분석: 결측치 보완, 원-핫 인코딩, 트렌드 지표 생성 등 모델 개발 및 검증: 수산물 가격 예측을 위한 머신러닝 모델 설계 및 성능 평가 서비스 구축: Django 기반의 API 개발 및 예측 모델을 웹 애플리케이션과 연동
진행 기간	2025년 1월 3일 ~ 2025년 2월 4일
팀프로젝트 링크	https://github.com/vhf1030/ml_pisces

[기술 스택]

활용한 기술 스택	python, sklearn, django, git, notion
선택 이유	문법이 쉽고 간결하며 머신러닝 및 서비스 개발 분야에서 많이 사용됨

[데이터 셋]

출처	수산물 가격 정보: 인어교주해적단 경제지표: 야후파이낸스 해양 기상 정보: 기상청 검색 트렌드: 네이버랩
크기	2017년부터 2025년 1월까지 수집된 일별 데이터
특징	일자별 시장별 어종별 가격 기상 정보, 환율, 검색어 트렌드 등 다양한 변수 추가 범주형 변수는 원-핫 인코딩 처리, 결측치 보완 및 정규화 수행

[프로젝트 진행 단계]

<ol style="list-style-type: none">1. 문제 정의 및 데이터 수집2. 데이터 전처리 및 탐색적 데이터 분석 (EDA)3. 모델 개발 및 검증4. 결과 해석 및 인사이트 도출5. 최소 기능 제품 서비스 구축

[프로젝트 세부 과정]

<ol style="list-style-type: none">1. 문제 정의 및 데이터 수집<ul style="list-style-type: none">- 수산물 가격의 변동성이 커서 소비자와 유통업자가 합리적인 의사 결정을 내리기 어려운 문제를 해결하고자 프로젝트 기획2. 데이터 전처리 및 탐색적 데이터 분석<ul style="list-style-type: none">- 결측치 처리(선형 보간법 및 평균 대체)- 범주형 변수 원-핫 인코딩 적용 (시장명, 어종 등)- 가격 변화 패턴 분석 및 상관관계 분석 수행3. 모델 개발 및 검증<ul style="list-style-type: none">- LightGBM, RandomForest, XGBoost 등 여러 모델을 테스트하여 성능 비교- 최적 모델 선정(LightGBM) 및 하이퍼파라미터 튜닝(AutoML)- 평가 지표: RMSE, R2 기준으로 성능 검증4. 결과 해석 및 인사이트 도출<ul style="list-style-type: none">- 날씨 및 환율이 일부 어종의 가격에 미치는 영향 확인- 시장별 가격 패턴이 있음을 발견하여 시장별 맞춤형 예측 모델 고려5. 최소 기능 제품 서비스 구축<ul style="list-style-type: none">- Django 기반의 웹 서비스 개발 및 배포- 서비스에 학습된 머신러닝 모델 연동

[프로젝트 결과]

https://github.com/vhf1030/pisces

[프로젝트 회고]

잘 한 점	모델 성능 최적화: 여러 머신러닝 모델을 비교하면서 최적의 성능을 확보 데이터 전처리 및 활용: 다양한 외부 데이터를 통합하여 더 정교한 예측
-------	--

	<p>모델 개발</p> <p>MVP 완성: Django 기반 API 서비스까지 구축하여 실제 사용할 수 있는 형태로 구현</p>
개선할 사항	<p>과적합 문제 해결 필요: 피처 엔지니어링 과정에서 너무 많은 Lag Feature를 생성하여 모델이 일부 데이터에 과적합된 경향이 있음</p> <p>서비스 기능 부족: 소비자 및 유통업자가 직접 활용할 수 있도록 기능 추가 필요</p> <p>운영 최적화 필요: API 연동, 스케줄링 및 DB 연동을 통한 효율적인 데이터 관리</p>