

수출액 예측을 통한 수출 유망 국가와 품목 추천 앙상블 모형과 클러스터링을 활용한 기업 선호도 반영 추천 솔루션

김성범 | 배준호

AI

제9회 산업통상자원부 공공데이터 활용 빅데이터 분석 경진대회

Problem Solving 문제 해결 과정



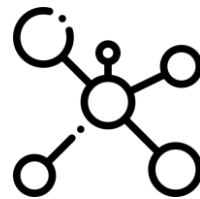
1. 분석 개요

문제 정의 및 해결방안
분석 프로세스



2. 분석 내용

데이터 탐색
파생변수 정의
데이터 전처리
모델링
모델해석



3. 클러스터링

제안 배경
모델링



4. 활용방안

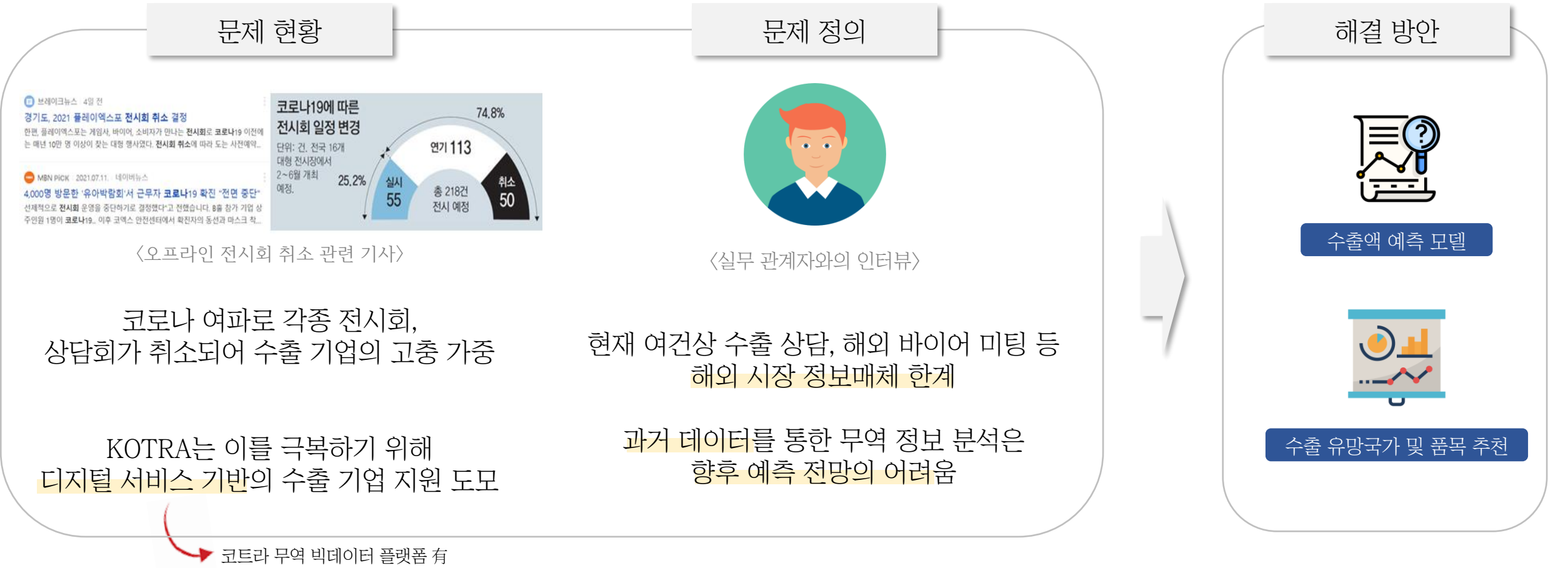
페르소나 정의
기업 선호도 조사
수출 유망 국가·품목 추천
결과 해석



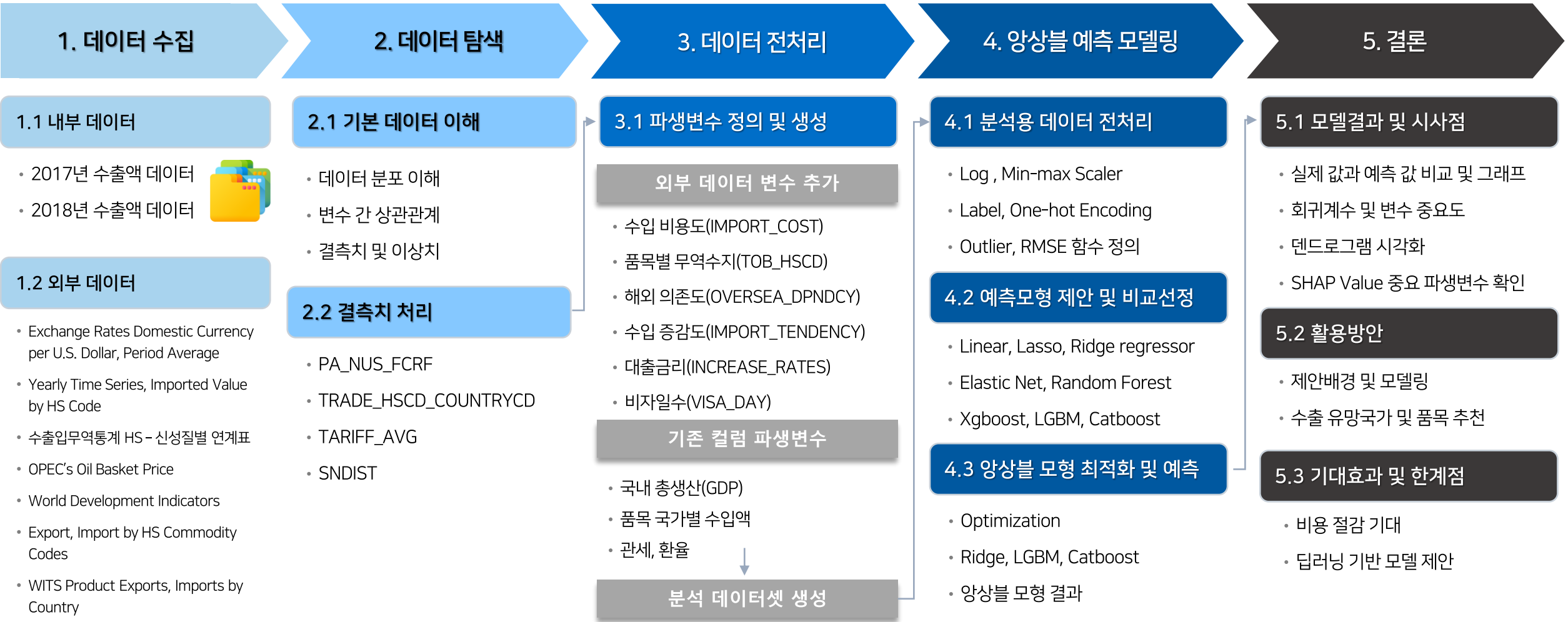
5. 결론

기대효과
한계점 및 제언

문제 정의 및 해결방안(Define & Solution)

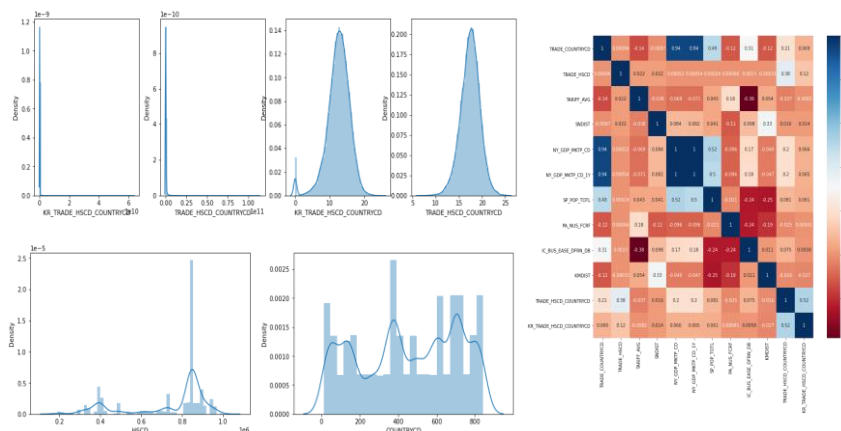


분석 프로세스(Process Flow)



데이터 탐색(Data Exploratory)

데이터 탐색



데이터 분포

상관관계

결측치 및 이상치 확인



종속 변수(수출액)를 예측하기 위해
설명력이 높은 변수의 필요성 제기

결측 값 대체

결측 변수명	변수 설명	결측 개수/비율
TARIFF_AVG	해당 품목 적용 관세	129 / 0.61%
SNDIST	품목을 수입한 국가 간 거리	22 / 0.1%
PA_NUS_FCRF	국가별 1년 환율 평균	3,488 / 16.46%
TRADE_HSCD_COUNTRYCD	품목별 1년 총 수입액	21 / 0.1%

〈학습 데이터의 결측치 개수와 비율〉

예측의 설명력을 높이기 위해 결측 값을 삭제하지 않고
데이터 참조와 함수를 통한 대체 방안 채택



02. 분석 내용

데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



파생변수 정의(Derived Variable)

해외 수출 관점의 도메인 지식을 통한 파생 변수 생성

외부 데이터 파생 변수



데이터 출처: OPEC Basket Price,
World Development Indicators

수입 비용도(IMPORT_COST)

각 국가와 한국 거리에 따른 **실질
유가 부담** 지수를 통해 수입 비용



데이터 출처: WITS Product
Exports, Imports by Country

수입 증감도(IMPORT_TENDENCY)

전년대비 해당 품목에 대한
수입액 증감 및 **품목 점유율** 파악



데이터 출처: Export, Import by
HS Commodity Codes

품목별 무역수지(TOB_HSCD)

품목별 수출·수입액의 차이
해당 품목의 **국가 경쟁력**



데이터 출처: IMF Data

대출금리(INTEREST_RATES)

해당 국가의 대출 금리가 낮을수록
자금 조달 및 **대금 지불**의 용이성



데이터 출처: WITS Product
Exports, Imports by Country

해외 의존도(OVERSEA_DPNDCY)

해당 국가·품목의 수입·수출액 차이
즉, **해외 수입 의존성** 파악



데이터 출처: 외교부 해외안전여행

비자일수(VISA_DAYS)

비자의 경우 해외 시장 진출 시
인력 교류에 제약 조건으로 작용



02. 분석 내용

데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



파생변수 정의(Derived Variable)

특성 공학(Feature Engineering)을 통한 파생 변수 생성

내부 데이터 파생 변수



변수 개수 : 4

국내 총생산(GDP)

1인당 실질 국내총생산(GDP)
경제성장률 등 이용하여
국가의 재정능력



변수 개수 : 8

품목·국가별 수입액

해당 품목·국가에 관한
수입액을 통해 전세계로부터
수입 비율



변수 개수 : 4

관세

관세를 이용하여
해당 국가가 특정 품목에 대한
수입의 용이성



변수 개수 : 9

환율

해당 국가의 환율을 고려한 변수
국가간 정보를 결합하여
종합적인 무역 용이성 고려



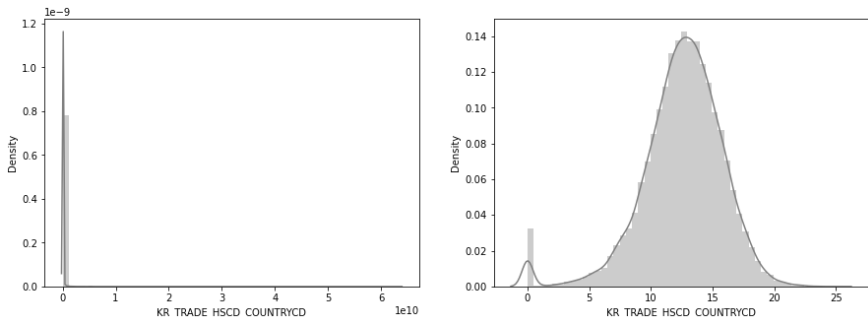
02. 분석 내용

데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



데이터 전처리(Data Preprocessing)

스케일링(Scaling)



왜도(Skewness)는 학습에 부정적인 영향

왜도 1 값 이상 변수에 로그(Log) 변환 적용하여 정규분포 변환

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

올바른 학습을 위해 수치형 변수를 일정한 범위 값 변환하는
최소최대정규화(Min-Max Normalization)을 적용

인코딩(Encoding)



원핫인코딩(One-Hot Encoding)

레이블인코딩(Label Encoding)

범주형 변수를 수치형 변수로 변환의 필요성

범주형 변수에 대해
원핫 인코딩을 적용하여 0과 1을 갖는 더미변수와
문자형을 숫자로 대체하는 레이블 인코딩 적용



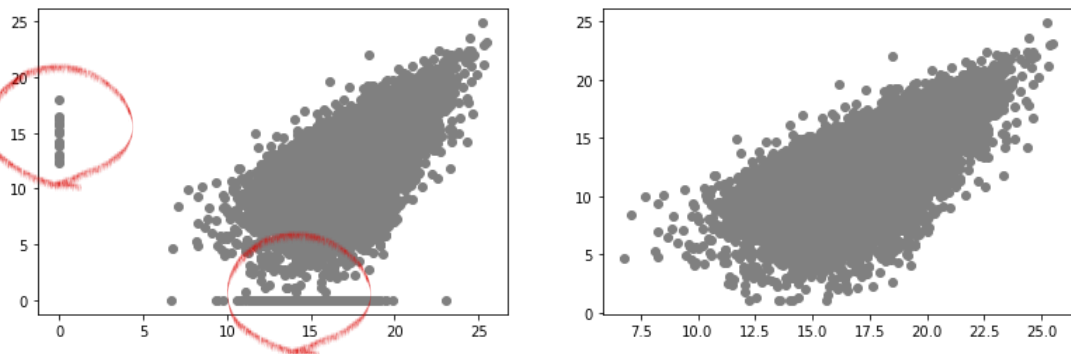
02. 분석 내용

데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



데이터 전처리(Data Preprocessing)

이상치(Outlier)

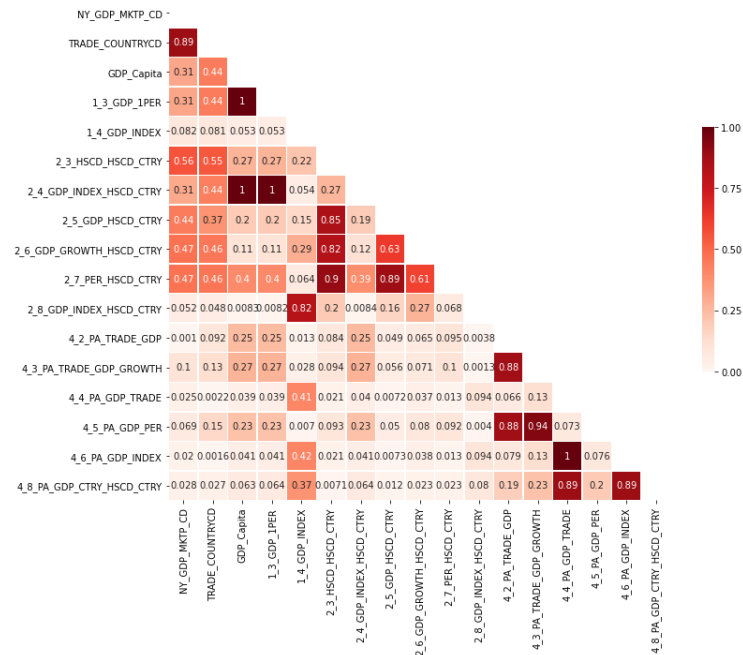


이상치(Outlier)는 학습 모델에 있어 부정적인 영향

국가의 해당 품목 수입금액은 있으나
한국으로부터 수입한 금액이 없다는 것을 의미

한국 수출액의 규모와 추세를 정확히 예측하기 위해
이를 이상치라고 정의하여 처리

상관분석(Correlation)



변수 간 상관성이 높으면 다중 공선성 문제 발생 가능

따라서, 설명 변수 간 상관계수를 확인

강한 상관성(0.8) 이상을 보이는 변수 제거



02. 분석 내용

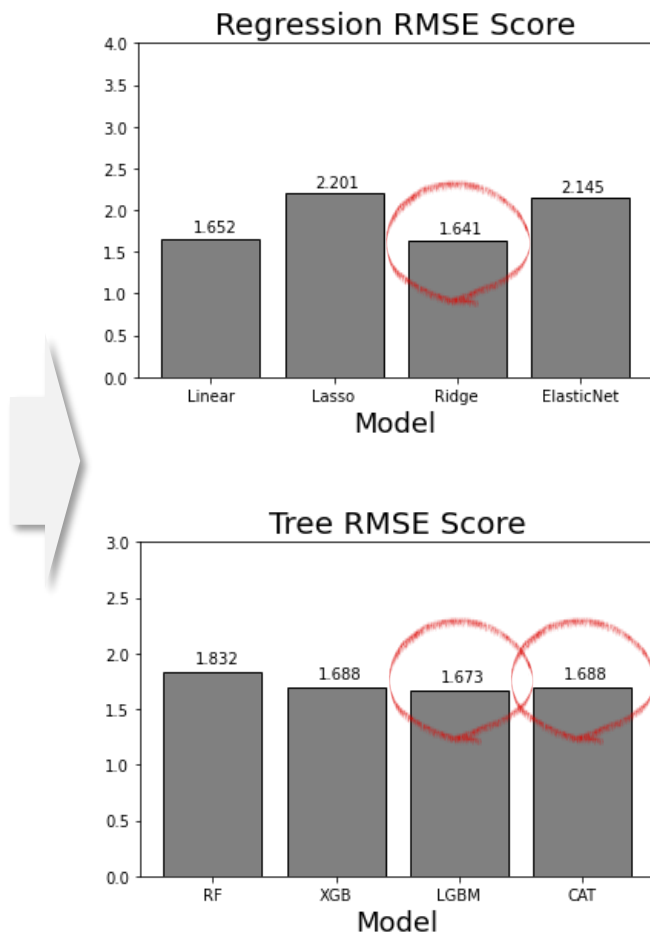
데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



모델 제안(Model Suggestion)

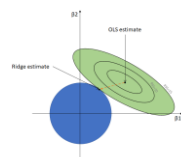
회귀 모델

Linear Regression
Lasso
Ridge
Elastic Net
Random Forest
Xgboost
LGBM
Catboost



데이터 특성과 모델 설명

Ridge



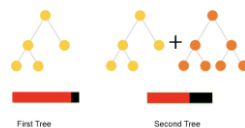
데이터 특성 상 더미 변수로 인해 **강한 희소성(Sparse)**
L2 Regularization을 이용하여 과적합(Overfitting) 방지

LGBM



2만개의 데이터에 메모리와 속도 측면에서 **효율 필요**
수직적 확장(Leaf-wise)의 방식으로 고효율적 알고리즘

Catboost



500개 품목과 43개의 국가 중심의 **범주형 데이터 특성**
범주형 변수에 적합한 카테고리(Categorical) 알고리즘



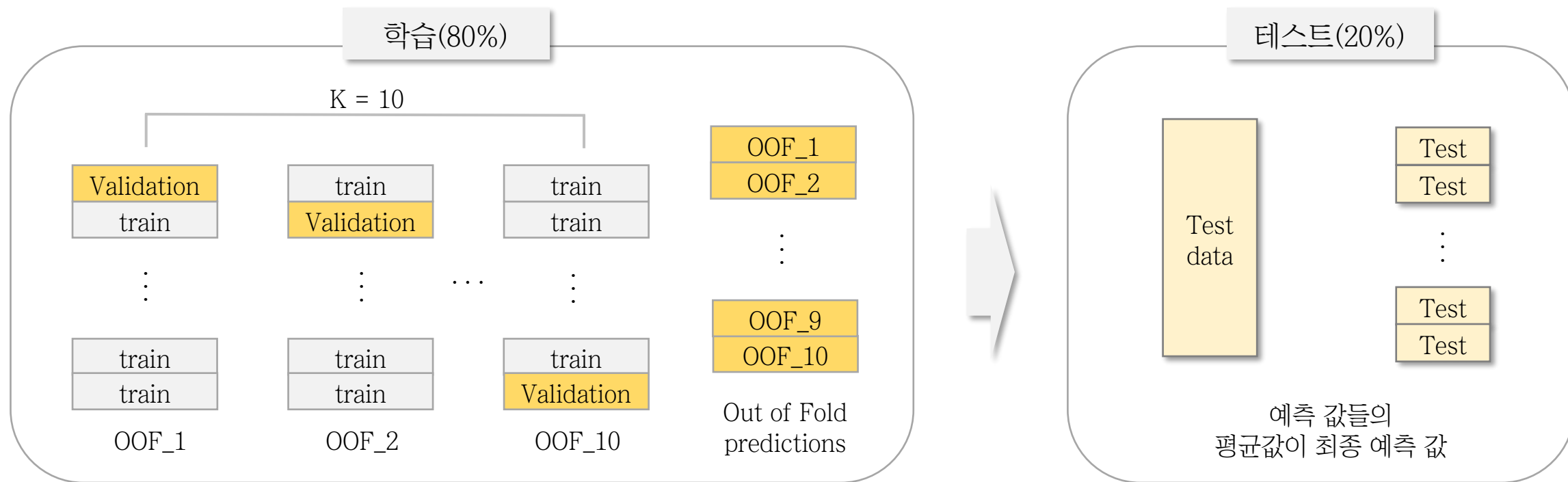
02. 분석 내용

데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



학습 방법(Learning Method)

OOF(Out of Fold) 검증을 통한 예측



K-Fold로 학습을 수행한 뒤 예측을 테스트 데이터에 10번 만큼 수행한 후 개별 예측 값을 평균하여 최종 예측
각 모델마다 앙상블(Ensemble)의 효과를 주어 과적합(Overfitting)을 회피



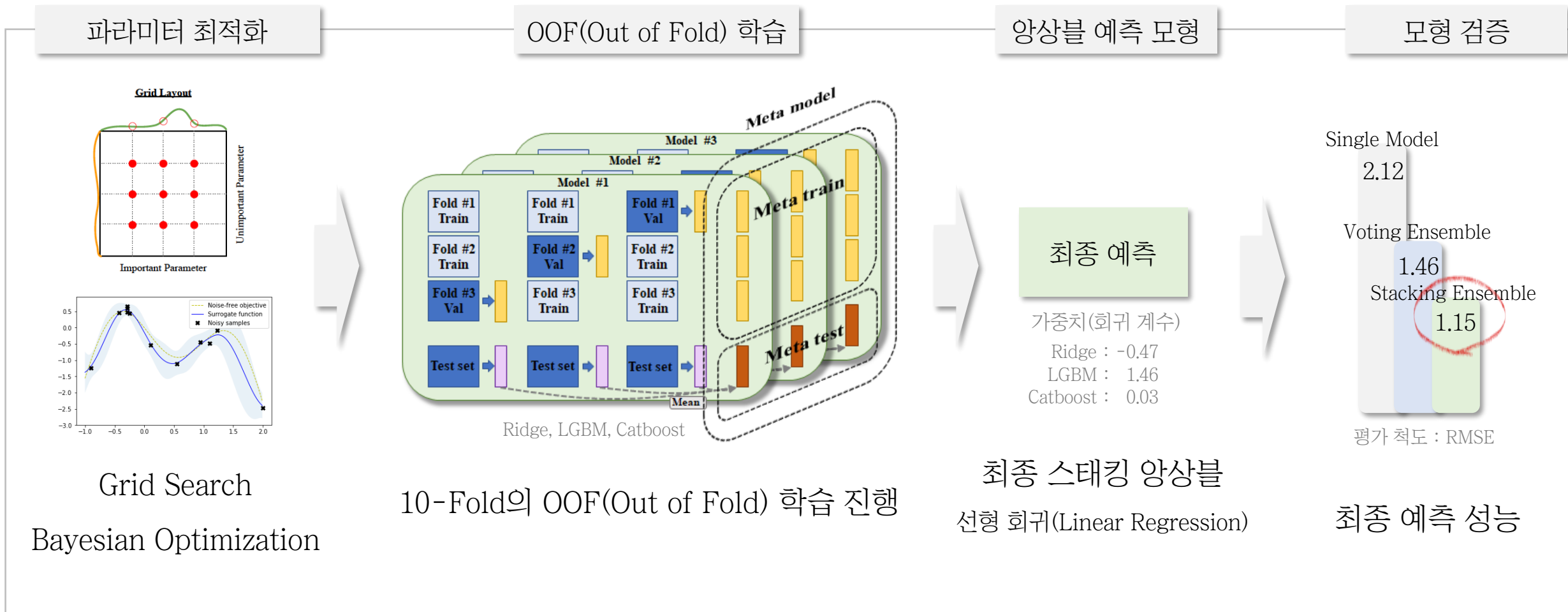
02. 분석 내용

데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



앙상블 예측(Ensemble Prediction)

OOF(Out of Fold)를 통해 학습용 데이터 생성 후 최종 모델 학습





예측 결과(Prediction Result)

테스트 데이터 셋

로그 적용 실제 값과 예측 값					실제 2018년 한국 수출액	
Ridge	LGBM	Catboost	Actual_log	Stacking_log	Actual_value	Predict_value
14.36933	14.61473	14.52873	15.11129	14.75029	3653845.00000	2546656.93341
21.74515	21.83725	21.07773	22.12470	22.06328	4061035672.00000	3819106247.02447
9.74074	9.55222	10.02636	9.15377	9.36201	9449.00000	11636.80440
12.08858	12.63590	12.52811	12.35819	12.87323	232859.00000	389738.67147
14.49703	14.59196	14.91440	14.46261	14.65266	1909995.00000	2309778.96375

〈16,661건의 학습을 통해 예측한 4,166건의 데이터〉

예측용 데이터 셋

					2019년 한국 수출액	
Ridge	LGBM	Catboost	Stacking_log	Predict_value		
10.93170	11.62267	11.35252	11.90737	148355.45798		
15.84631	16.11026	16.69460	16.28589	11826938.40903		
10.29765	12.04857	11.55061	12.85671	383352.43837		
13.92693	14.40837	14.36328	14.65339	2311451.83414		
15.32867	14.96008	15.56094	14.80368	2686300.86700		

〈2019년 한국 수출액의 21,189건 데이터〉

Ridge, LGBM, Catboost 개별 예측 모델과 이를 스택킹 앙상블(Stacking Ensemble)한 결과

로그(Log) 변환한 예측한 값과 실제 값이 거의 유사(RMSE : 1.15)

실제 값으로 확인하였을 때 규모와 추세를 반영



02. 분석 내용

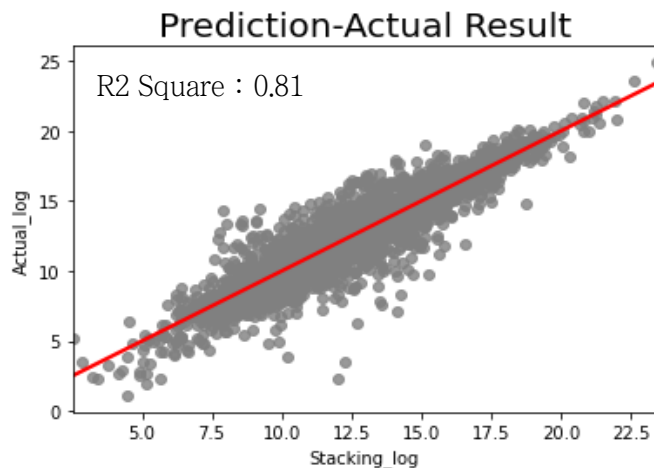
데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



회귀 분석(Regression Analysis)

설명 변수가 종속 변수(수출액)에 미치는 영향력과 크기 확인

선형 그래프(Linear Graph)

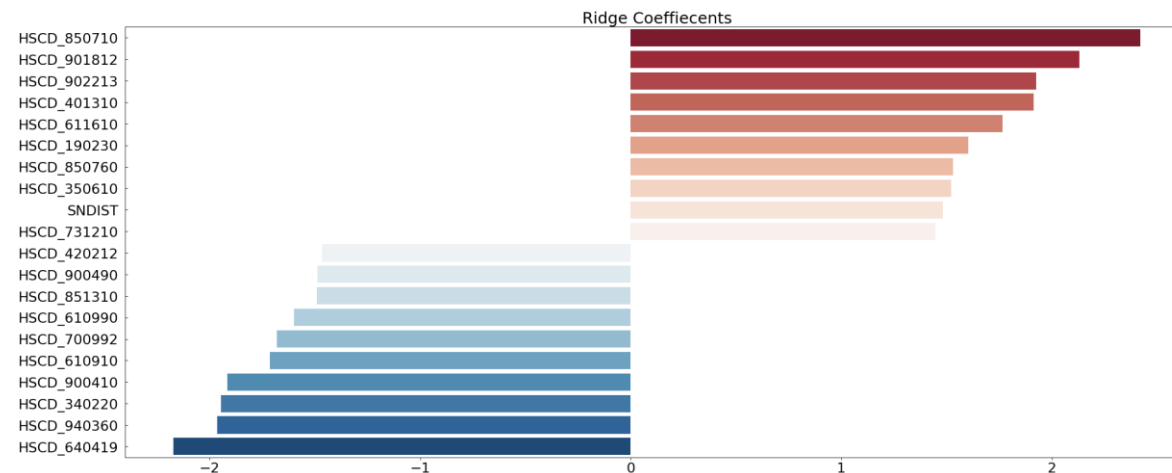


〈로그 실제 값과 예측 값의 선형관계〉

예측한 값과 실제 값의 선형성 확인

양상불 예측 모델이 규모와 추세를 반영 설명

회귀 계수(Regression coefficient)



〈Ridge 회귀 계수 상·하위 10개〉

영향력이 큰 변수는 HSCD(품목코드)

특정 품목에 따라 한국 수출액에 영향을 미치는 것으로 해석



02. 분석 내용

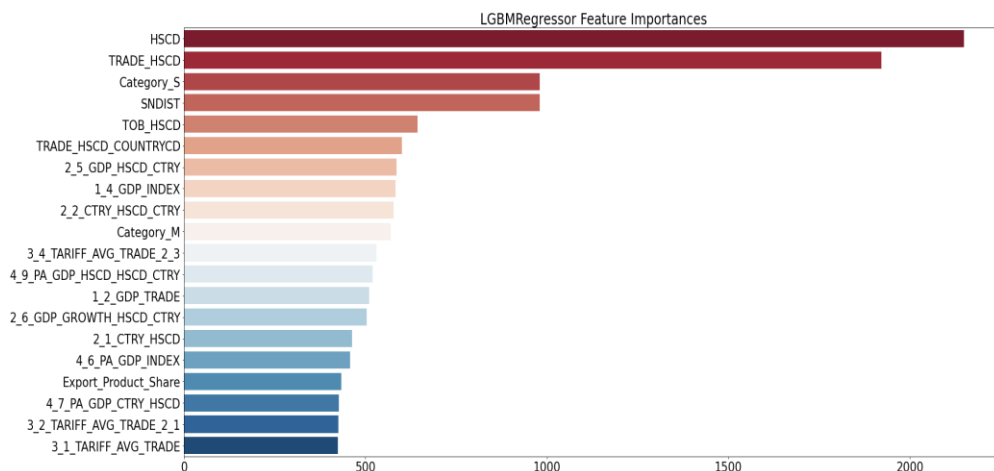
데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



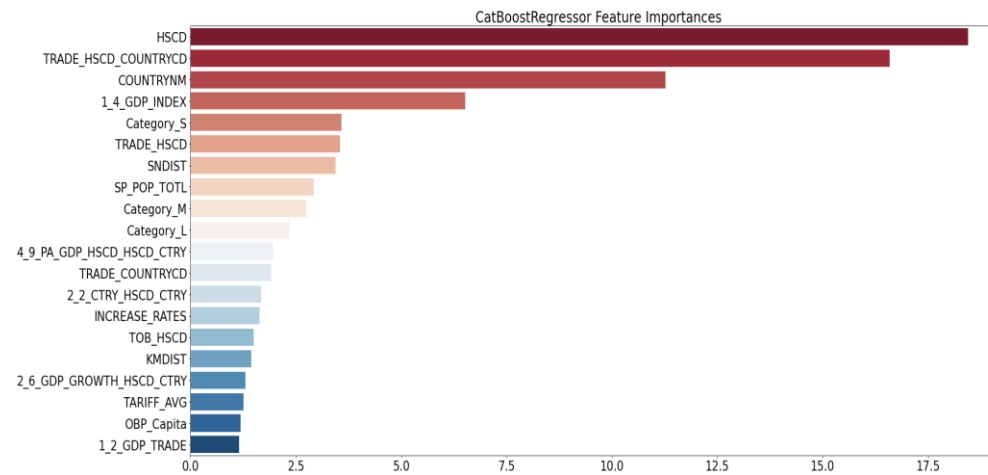
변수 중요도(Feature Importance)

종속 변수(수출액)를 예측하는데 자주 사용된 변수 확인

LGBM의 변수 중요도



Catboost의 변수 중요도



〈높을수록 자주 사용된 변수를 의미함〉

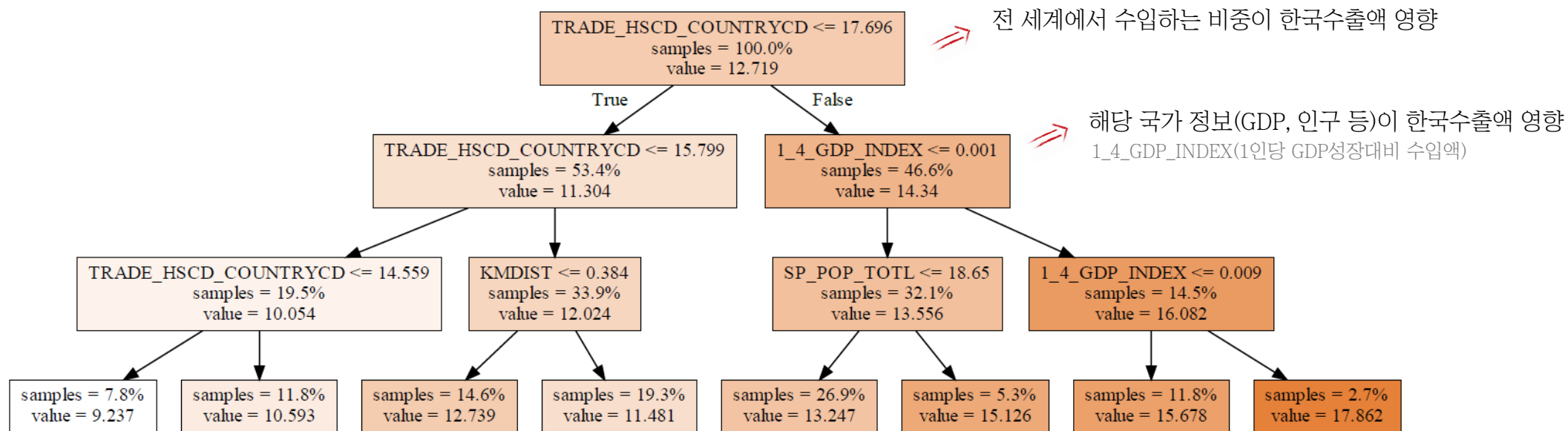
공통적으로 품목코드(HSCD), 해당 국가의 특정 품목수입액, 국가, 거리 순으로 변수 중요도 확인

특정 품목과 국가의 정보가 한국으로부터 수입하는 패턴 확인



덴드로그램(Dendrogram)

의사결정 나무(Decision Tree)를 통해 종속 변수(수출액)를 설명 변수 패턴 확인



〈직관적인 시각화로 변수 해석의 용이성을 갖는 의사결정 나무(Decision Tree)〉

특정 국가가 수입하는 특정 품목의 수입액과 국내총생산(GDP), 인구, 거리가 종속 변수를 설명
즉, 국가 정보 변수가 한국 수출액에 영향을 의미



02. 분석 내용

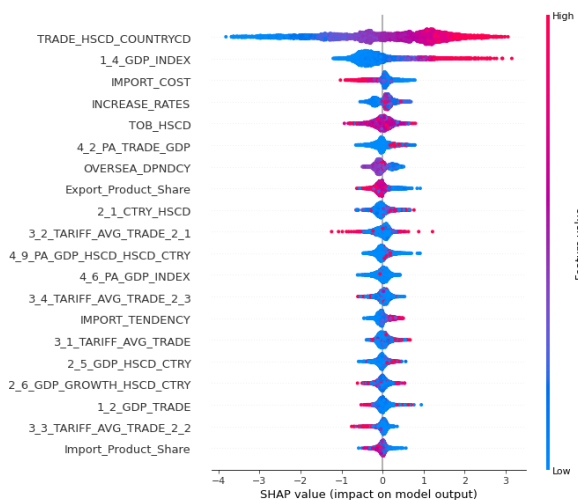
데이터 탐색 / 파생변수 정의 / 데이터 전처리 / 모델링 / 모델해석



eXplainable AI(SHAP Value)

설명가능한 인공지능, 각 변수들이 예측치에 얼마나 기여를 했는지 설명

파생 변수 영향도 확인

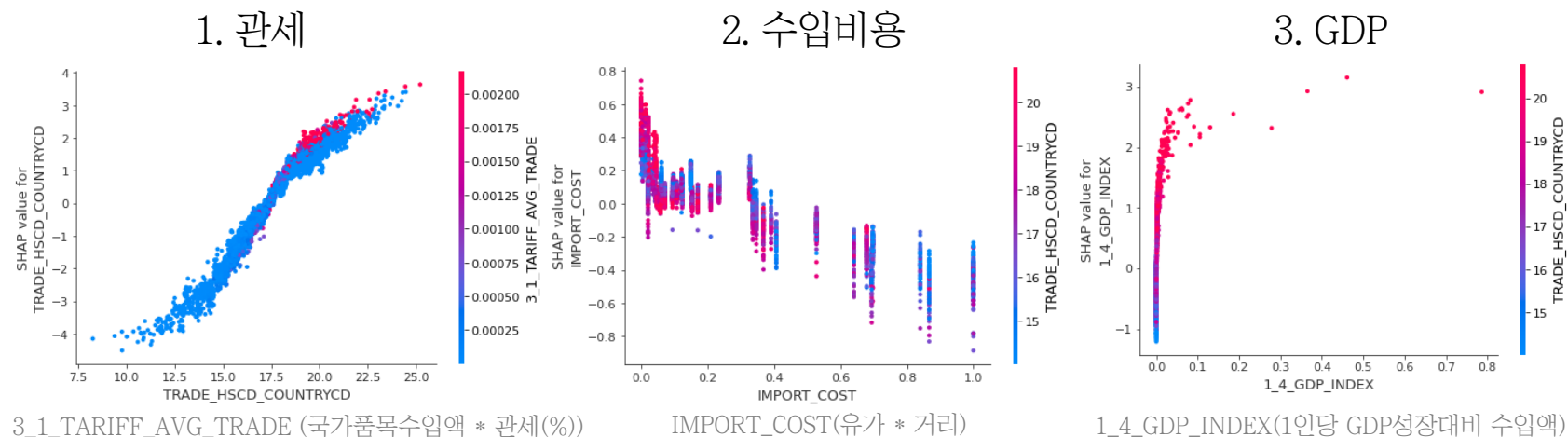


<정의한 파생변수에 대한 요약>

GDP, 수입비용, 대출금리,
무역수지, 해외의존도 등

영향도 요약 확인

특정 국가·품목의 수입 패턴 확인



국가·품목에 대한 수입액 관계를 보기 위해 SHAP dependence plot

해당 국가의 낮은 관세 품목일수록
수입하는 경향이 높음

✓ 낮은 관세 품목에 집중

해당 국가의 낮은 수입비용일수록
수입하는 경향

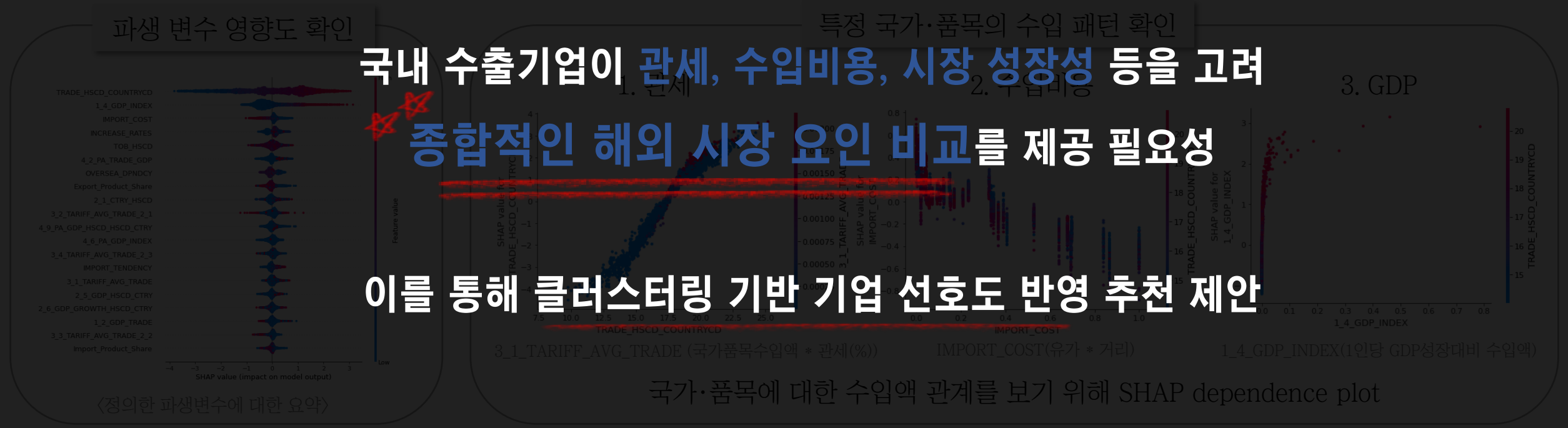
✓ 낮은 수입비용이 선호

시장 성장가능성이 클수록
국가의 수입액이 높음

✓ GDP 성장률의 영향

eXplainable AI(SHAP Value)

설명가능한 인공지능, 각 변수들이 예측치에 얼마나 기여를 했는지 설명



GDP, 수입비용, 대출금리,
무역수지, 해외의존도 등
영향도 요약 확인

해당 국가의 낮은 관세 품목일수록
수입하는 경향이 높음
✓ 낮은 관세 품목에 집중

해당 국가의 낮은 수입비용일수록
수입하는 경향
✓ 낮은 수입비용이 선호도

시장 성장가능성이 클수록
국가의 수입액이 높음
✓ GDP 성장률의 영향



서비스 접근 방식(Approach)

서비스 제안 배경

무역투자동계

- 한국 수출 현황, 글로벌 수출 현황, 외국인 직접투자동계
- 각종 데이터의 시각화 및 현황 정보 제공 (대시보드)

국가별 시장정보

- 국가별 시장개요, 국가주요자료(WorldBank), KOTRA 수출 품목 정보
- 국내외 무역정보, 무역사기서세, 수입과세 정보, 수출 분석 보고서

품목별 유통시장

- 품목별 해당시장의 수입 규모 예측
- 관련 무역 통계의 다차원 제공, 수출 분석 보고서, 수입과세 정보

기업별 맞춤정보

- 기업개요, 주원사원/사업장 정보제공
- 기업별 맞춤정보 보고서 제공

경제패널서 정보

- KOTRA가 보유한 경제기업 정보
- 기업별 경제패널서 추천

시장 정보	시장 정보	시장 정보	시장 정보	시장 정보	시장 정보	시장 정보
86	0.5	4.55	44,253.79	3.88	3,444,356	3,208

〈코트라 빅데이터 플랫폼의 현황〉

시장 정보, 유망 시장, 분석 보고서, 기업 맞춤형 등 제공

특정 품목(HSCD)에 대한 과거 정보에 집중
전반적인 시장 정보 획득의 어려움

수출희망 기업의 특성과 선호를 고려한
맞춤형 정보의 부족

개선 방안

드릴 다운(Drill Down)



직관적인 분류 단위부터
품목 코드까지 정보 세분화

사용자의 정보 검색 편리성 제고
종합적인 정보를 통한 효율적 선택

기업 선호도 반영



수출기업의 선호도를 조사하여
맞춤형 추천 정보 제공

수입액 예측과 선호도를 반영한
수출 유망 클러스터 도출

기업의 비용적 효율화 기대



03. 클러스터링

제안 배경 / 모델링



모델링(Modeling)

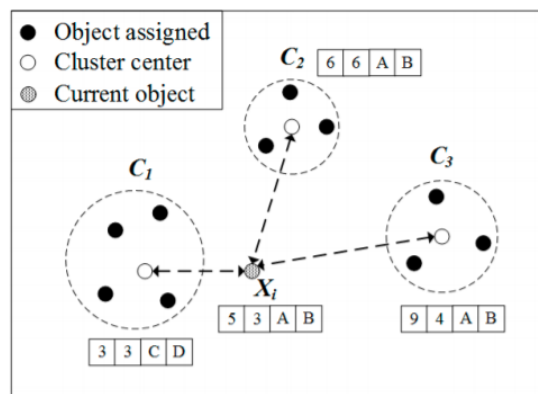
데이터 전처리



〈21,189건의 데이터와 41개의 변수〉

상관변수 제거
범주형 변수 정의
정규화
종속변수 제거
파생변수 추가

클러스터링 알고리즘



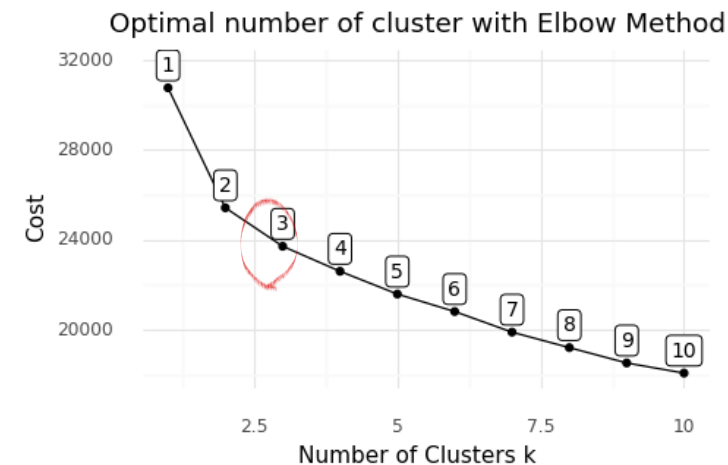
〈K-Prototypes 알고리즘〉

500개의 품목과 43개 국가의 범주형과
수치형 변수로 구성된 **혼합 데이터 특성** 고려



K-means(수치형)와 K-mode(범주형)를 혼합한
K-Prototypes 알고리즘 활용

최적 K 탐색



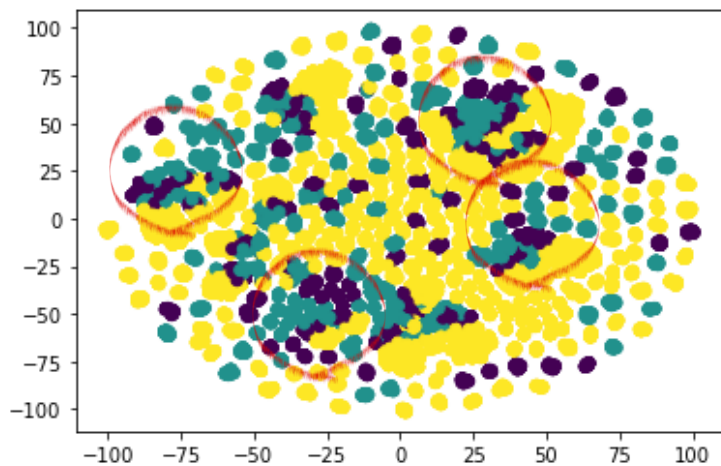
〈Elbow Method를 이용한 최적 K 탐색〉

클러스터의 최적 개수는
비용(Cost)이 가장 완만하게 감소하는
3구간에서 최적 K 선정



모델링(Modeling)

클러스터 시각화



〈t-SNE를 이용한 차원 축소 시각화〉

높은 차원의 비선형 데이터 시각화에
유용한 차원축소(t-SNE)
클러스터 별 시각화 확인

세그먼트 간 비교

기업의 선호를 반영한 변수

	COUNTRYNM	Category_M	Category_S	Detail_Category	TARIFF_AVG	VISA_DAYS	IC_BUS_EASE_DFRN_DB	INCREASE_RATES	KR_TRADE_HSCD_COUNTRYCD	
cluster_id										
0	11525	Mexico	IT부품	반도체	집적회로반도체	0.00000	0.50000	0.69141	0.19542	1007357483.73663
	11005	Chile	내구소비재	자동차	가솔린 승용차(신차)	0.23427	0.50000	0.65784	0.09533	870729910.54894
	16828	Mexico	IT제품	정보기기	컴퓨터 주변기기	0.00000	0.50000	0.69141	0.19542	811607926.18132
1	11448	Philippines	IT부품	반도체	집적회로반도체	0.00000	0.16667	0.37135	0.14562	1153803068.14066
	11537	Viet Nam	IT부품	반도체	집적회로반도체	0.14061	0.08333	0.54498	0.17810	1116114091.27686
	12427	Viet Nam	IT부품	기타 IT부품	기타 IT부품	0.14061	0.08333	0.54498	0.17810	930318743.96188
2	11471	China	IT부품	반도체	집적회로반도체	0.26167	0.00000	0.50469	0.09978	6115797429.04325
	11436	China, Hong Kong SAR	IT부품	반도체	집적회로반도체	0.00000	0.50000	0.95519	0.11771	5413303494.70538
	11476	China, Hong Kong SAR	IT부품	반도체	집적회로반도체	0.00000	0.50000	0.95519	0.11771	4701454018.55636

〈시장 성장성과 수출 규제를 고려한 3개 집단〉

기업 맞춤형 서비스 제공을 위해
기업의 선호도를 반영한 변수를 추출
이를 통해 세그먼트 간 비교 가능



04. 활용 방안

페르소나 정의 / 기업 선호도 조사 / 추천 결과 시각화 / 결과 해석



페르소나 정의(Persona Define)

수출액 예측에 기반한 기업 선호도 반영 클러스터링 프로토타입

기업 맞춤형 수출 유망 국가·품목 추천 서비스



(주) 델타기계테크

(주) 델타기계테크

기업 형태 : 중소기업

업종 : 기계 전기 변환장치 제조업

사원 수 : 58 명

설립일 : 2013년 9월 30일

매출액 : 111억 2,612만원



기업 현재 상황 :

- ① 최근 기계 전기 관련 해외 특허를 취득하였음.
- ② 내년 사업 방향으로 해외 진출을 시도하고자 함
- ③ 엔지니어 중심 회사로 무역 정보 획득이 어려움
- ④ 업종 상 유지보수 관련 출장이 잦을 것으로 예상



검색

기업 A
델타기계테크

Home

Profile

Settings

기업 선호도 조사

☒ Q1.☒ Q2.☒ Q3.



+ Add Menu

다음의 질문에 답변해주세요.

답변하신 내용은 적합한 수출 유망 국가 · 품목 추천에 반영됩니다.

Q1. 귀사 수출 사업의 특징은 무엇입니까?

- ☒ 특정 품목을 수출하고자 한다.
- ☐ 특정 국가 기반으로 수출하고자 한다.
- ☐ 특정 국가 대상으로 특정 품목을 수출하고자 한다.
- ☐ 정해진 품목과 국가가 아직 없다.

 검색기업 A
델타기계테크 Home  Profile  Settings 

기업 선호도 조사

☒ Q1.☒ Q2.☒ Q3.

+ Add Menu

Q2. 귀사 수출 품목의 분류는 무엇입니까?

분류

기계류

x v

소분류

전기기계

x v

세분류

전기변환장치

x v

☒ 전기기계

기계요소 및 공구

일반기계

정밀기계

금속 구조물

☒ 전기변환장치



변압기

발전기

조명장치

일차전지 및 축전지

⋮

 검색기업 A
델타기계테크 Home  Profile  Settings 

기업 선호도 조사

☒ Q1.☒ Q2.☒ Q3.

+ Add Menu

Q3. 귀사 수출 사업의 희망 방향은 무엇입니까? (최대 2개)

- ☐ 수출 비용 부담이 적었으면 좋겠다.
- ☐ 시장 규모가 컸으면 좋겠다.
- ☒ 시장의 성장 가능성이 컸으면 좋겠다.
- ☒ 수출에 제약이 적었으면 좋겠다.
- ☐ 시장 진입 장벽이 낮았으면 좋겠다.

제출

🔍 검색

기업 A
델타기계테크

Home

Profile

Settings

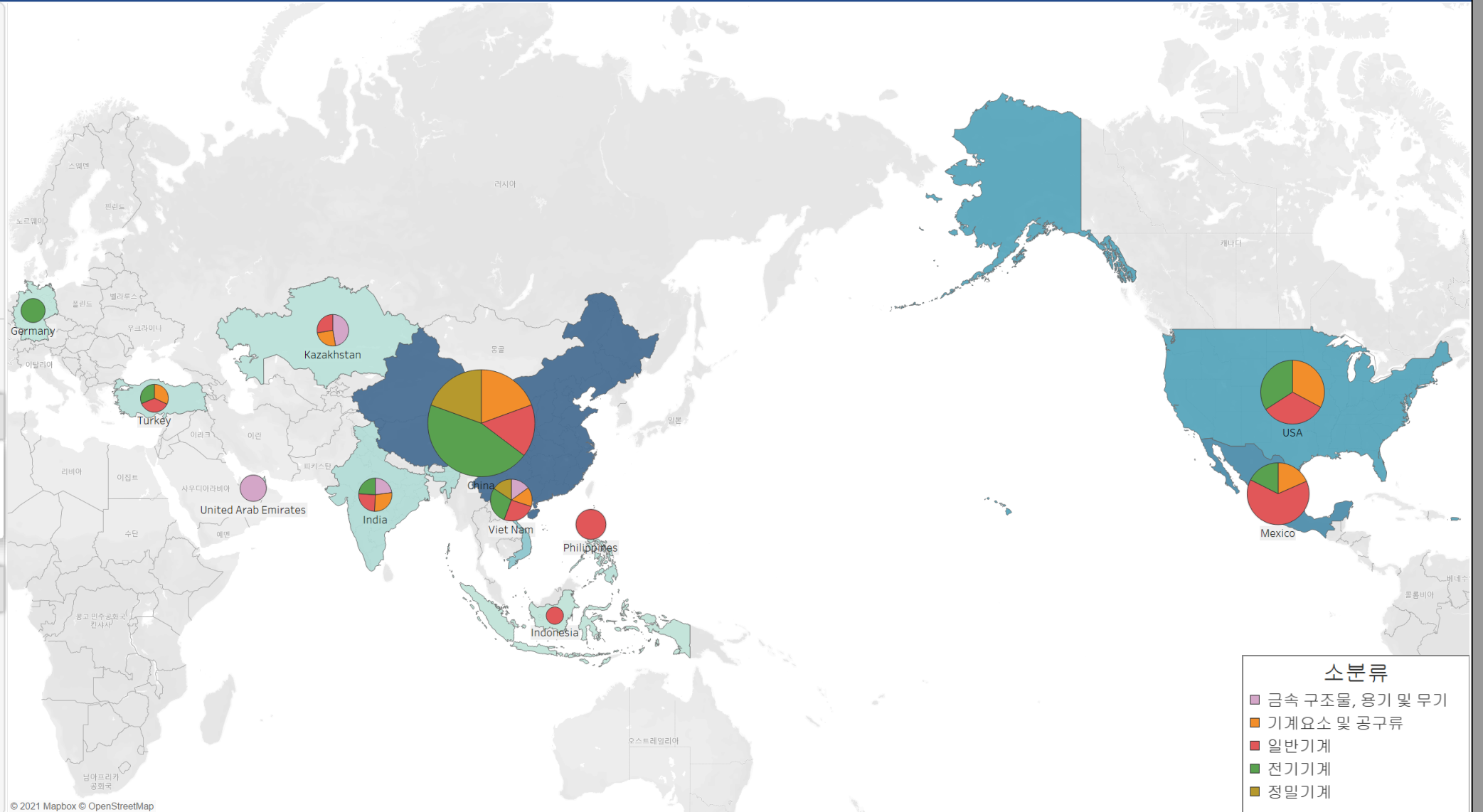
추천 결과

수출 지도

✓ 유망 시장 탐색

시장 상세

품목 클러스터



🔍 검색

기업 A
델타기계테크

Home

Profile

Settings

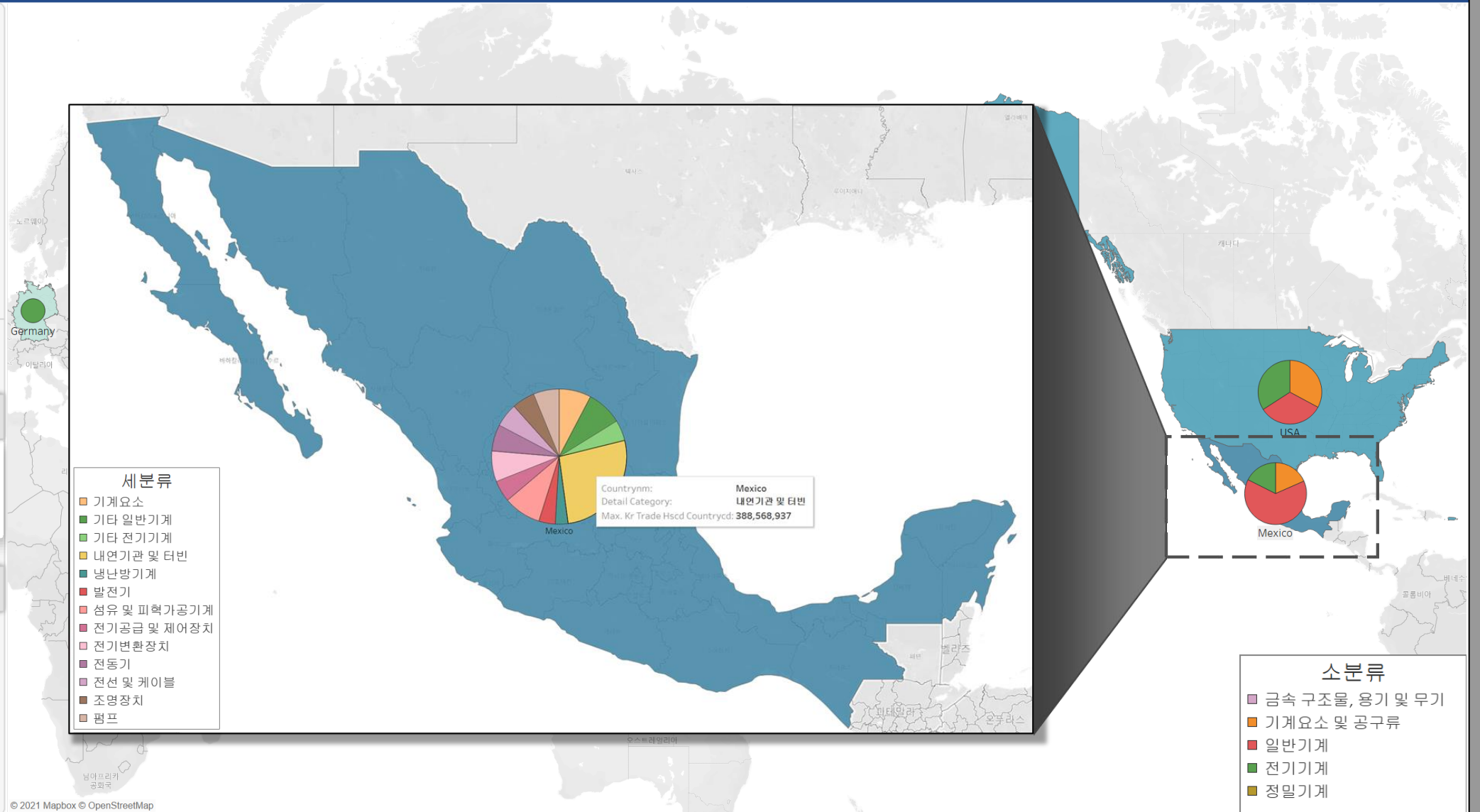
추천 결과

수출 지도

유망 시장 탐색

✓ 시장 상세

품목 클러스터



검색


기업 A
델타기계테크

Home

Profile

Settings

추천 결과

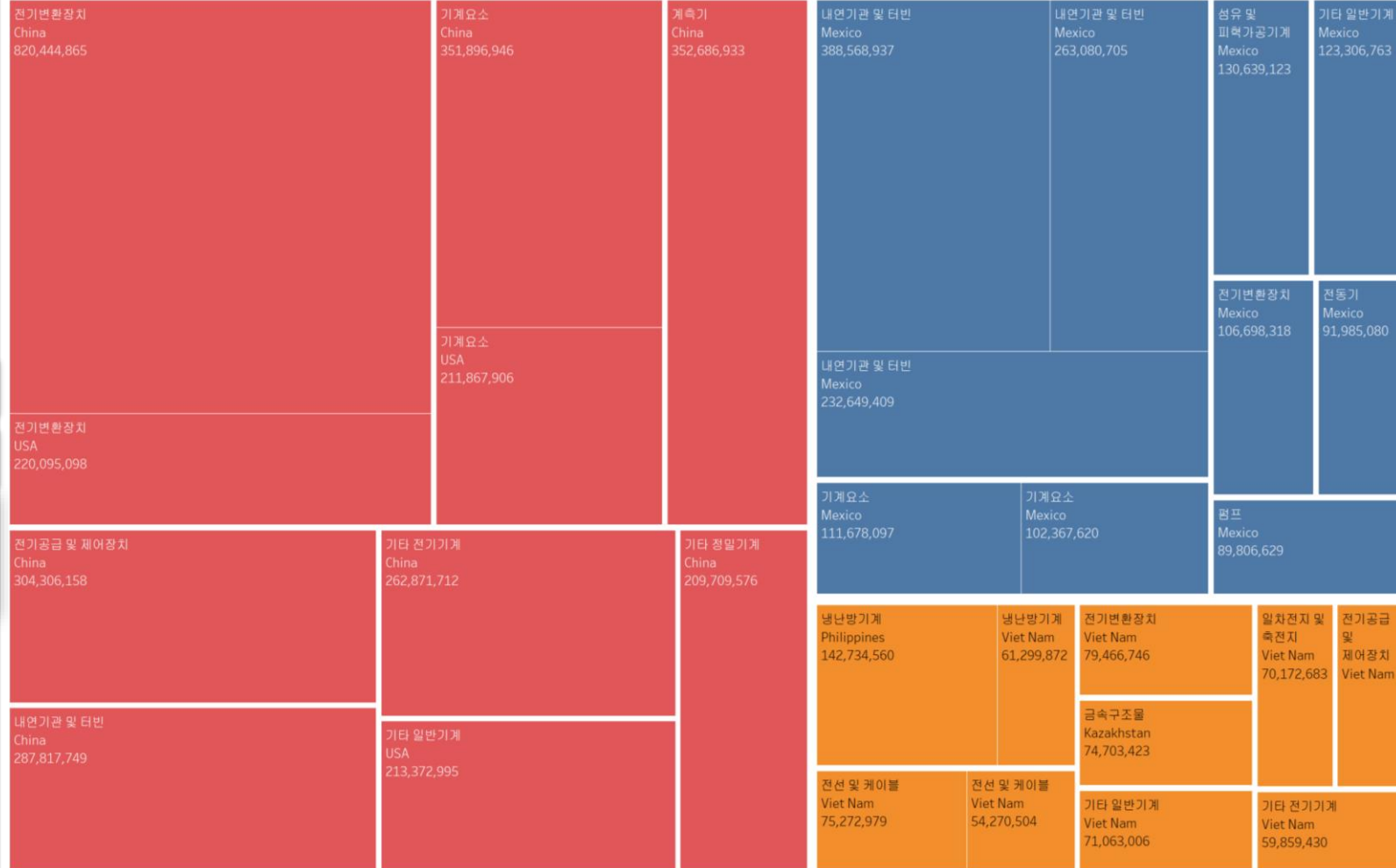
수출 지도

품목 클러스터

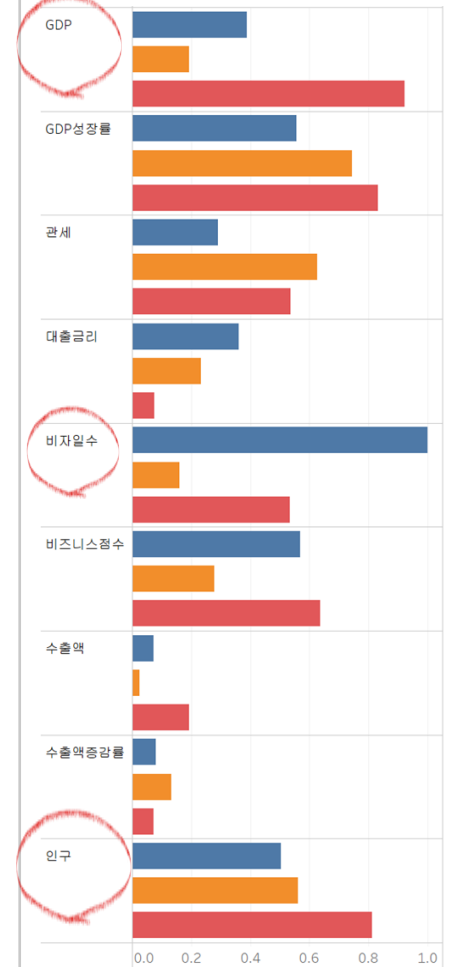
✓ 유망 품목 트리맵

품목 상세

Filter : 기계류 × 전기기계 × 전기변환장치 ×



기업 선호도 반영 클러스터



🔍 검색

기업 A
델타기계테크

Home



Profile



Settings

추천 결과

수출 지도

품목 클러스터

유망 품목 트리맵

✓ 품목 상세

Filter :

기계류 x

전기기계 x

전기변환장치 x

전기변환장치
China
820,444,865기계요소
China
351,896,946계측기
China
352,686,933내연기관 및 터빈
Mexico
880,568,937전기변환장치
USA
220,095,098기계요소
USA
211,867,906내연기관 및 터빈
Mexico
32,649,409전기공급 및 제어장치
China
304,306,158기타 전기기계
China
262,871,712기타 정밀기계
China
209,709,576냉난방기계
Philippines
142,734,560내연기관 및 터빈
China
287,817,749기타 일반기계
USA
213,372,995전선 및 케이블
Viet Nam
15,272,979전선 및 케이블
Viet Nam
54,270,500

세분류	HSCD	국가	예상 수출액
전기변환장치	850440	China	820,444,865
		USA	220,095,098
전기공급 및 제어장치	853710	China	304,306,158
		USA	160,309,195
기타 전기기계	854370	China	262,871,712
		USA	150,263,100
전선 및 케이블	854442	China	200,465,957
	854430	USA	93,812,061
발전기	850300	USA	97,570,039
조명장치	851220	Germany	91,838,615
일차전지 및 축전지	850760	China	91,053,706
계측기	903149	China	352,686,933
기타 정밀기계	902780	China	209,709,576
기계요소	848180	China	351,896,946
		USA	211,867,906
	848120	China	169,304,851
	848210	China	99,029,591
	731815	USA	93,950,456
내연기관 및 터빈	840991	China	287,817,749
		USA	91,672,945
	840734	USA	149,513,897
기타 일반기계	392690	USA	213,372,995
냉난방기계	732111	USA	146,605,657
	847990	USA	102,456,292
운반하역기계	842890	China	99,329,684
		USA	90,444,262
건설기계	843149	China	95,428,420
펌프	841430	USA	93,361,410
	841480	China	90,949,584
금속구조물	730890	United Arab Emi..	109,921,087

클러스터



클러스터링 해석(Clustering Result)

기업의 선호에 따른 맞춤형 선택 가능

국가	유망 세부 품목		
멕시코	전동기, 펌프, 전선 및 케이블		
터키	기계요소, 전기 변환장치		
필리핀		냉난방기계, 건설 기계	
베트남		전기변환장치, 전선 및 케이블	
카자흐스탄		금속구조물, 기계요소	
인도		기계요소, 금속구조물	
인도네시아		건설기계	
중국			전기변환장치, 계측기
미국			전기변환장치, 일반기계
UAE			금속 구조물
독일			조명장치



① 해외 진출에 용이한 그룹

- 관세, 비자일수, 비즈니스 용이성 점수 ↑

비즈니스 용이성 점수가 높고, 무비자 · 저관세
혜택으로 해외 진출이 수월하다고 판단

② 시장 성장성이 높은 그룹

- GDP 성장률, 수출액 증가율 ↑

GDP 성장률이 높고, 예상 수출액 증가율이 높아
해당 시장의 성장성이 높다고 판단

③ 시장 안정성이 있는 그룹

- 예상 수출액, 인구, GDP ↑

인구와 국내총생산(GDP)이 높고, 예상 수출액이
높아 시장의 규모가 크다고 판단

관세, GDP, 예상 수출액 등 다양한 설명 변수를 종합하여 기업 선호도를 반영한 클러스터링 결과의 예시임.



04. 결론

기대효과 / 한계점 및 제언 / 참고 문헌



기대효과(Expected Results)

#1



수출입 관련
정보획득 비용 절감

정보 획득을 위한 수출 무역 전시회 부스
비용 절감 가능 (2칸 기준 약 800만원)

과거 정보와 달리 수출액 예측을 통한
기업 맞춤형 정보 획득 가능

#2

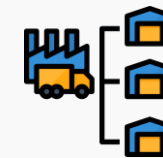


무역 빅데이터 플랫폼
이용 기업 만족도 증가

기존 5,000개 이상 세분화 품목 코드
이를 통한 접근의 어려움

분류 단위부터 시작하는 정보 제공으로
기업의 이용 접근성 증대

#3



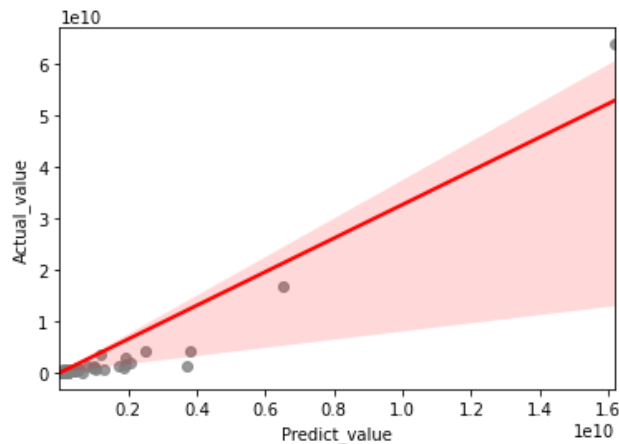
클러스터링을 통한
세그먼트 시장 추천

기업 선호도를 반영한 클러스터링
설명 가능한 수출유망 시장 추천

국내 수출 중소-중견기업의
해외진출 확대 및 경쟁력 강화 제고



한계 및 제언(Limitations and Suggestions)

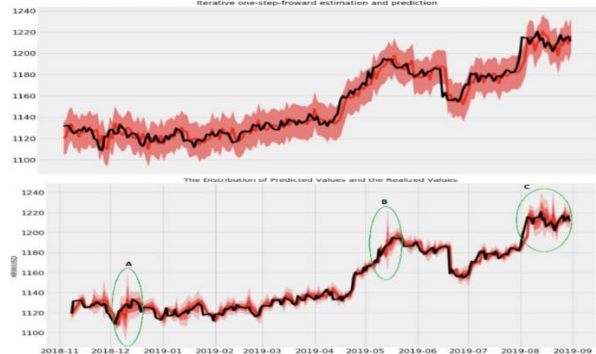


〈단적 요인으로 예측하는 것의 한계〉

한계점(Limitations)

- ✓ 한국으로부터 수입하지 않은 국가에 대한 예측의 어려움
- ✓ 시계열 정보가 아닌 단적 연도를 통한 수출액 예측 한계

Figure 11. Prediction of Korean won-US dollar exchange rates
Iterative one-step-forward estimation and prediction



〈딥러닝으로 환율·수출 예측 분석(한국은행)〉

향후 과제(Future Work)

- ✓ 무역 데이터 뿐만 아닌, 기업 내부 환경의 데이터 수집의 필요성
- ✓ 국가 간 무역 협상, 코로나 등 국제 정세와 같은 정형·비정형 데이터 적용
- ✓ 시계열 정보를 반영한 데이터와 시계열 딥러닝(RNN, LSTM, GRU) 모델 제안



참고 자료 및 활용 데이터

참고 자료

- ◆ 신현수·배상근·최호상, 단기 수출입 예측모형 개발. 한국산업경제연구원, 1999
- ◆ “한은, 딥러닝으로 환율·수출 예측분석…‘오차범위 줄어’” 한국은행, 2020
- ◆ 빅데이터로 본 내년 품목별 유망시장, 자동차는 미국, 반도체는 중국. 이투데이, 2020
- ◆ 무역보험공사 '수출 빅데이터 분석 리포트' 발간, 파이낸셜투데이, 2020
- ◆ 내년 수출 6~7% ↑ 5400~5500억 달러…코트라 전망
- ◆ 윤상혁·최윤진·이소현·김희웅, 머신러닝 기반의 뷰티커머스 고객 세그먼트 분류 및 활용방안. Information System Review, 2020
- ◆ 무역투자 빅데이터 서비스. 코트라, 2021

분석 도구



활용 데이터

활용 데이터	데이터 형태	목적	출처
Exchange Rates Domestic Currency per U.S. Dollar, Period Average	정형 데이터	결측값 처리	IMF
Yearly Time Series, Imported Value by HS Code	정형 데이터	결측값 처리	ITC TradeMap
HS 신성질별 연계표	정형 데이터	파생변수 생성	관세청 수출입무역통계
OPEC Basket Price	반정형 데이터	파생변수 생성	OPEC
World Development Indicators	정형 데이터	파생변수 생성	World Bank
Export, Import by HS Commodity Codes	정형 데이터	파생변수 생성	UN Comtrade
WITS Product Exports, Imports by Country	정형 데이터	파생변수 생성	World Integrated Trade Solution
IMF Lending Intrests Rates	정형 데이터	파생변수 생성	IMF
외교부 해외안전여행	정형 데이터	파생변수 생성	외교부