

제10회 산업통상자원부 공공데이터 활용 BI 공모전  
- 빅데이터 분석 부문 -

---

# 한국 수출금액 예측 모델 기반 기업수요 맞춤형 서비스

---

노종화, 이준희, 이현태 (한양대학교 대학원)

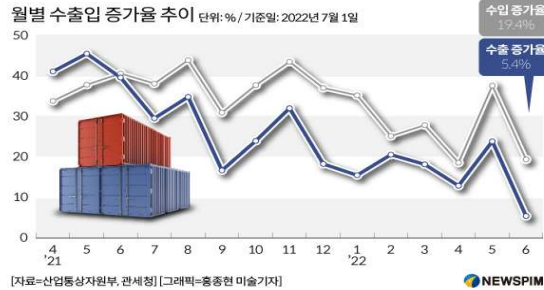
# 목 차

- |   |       |                                    |
|---|-------|------------------------------------|
| 1 | 개 요   | 분석 배경 및 필요성 / 분석 프로세스              |
| 2 | 분석 내용 | 데이터 수집 / 데이터 탐색 / 모델링 / 모델 평가 및 해석 |
| 3 | 활용 방안 | 국가별 품목별 예상 수출금액 제공 서비스             |
| 4 | 결 론   | 기대효과 / 한계점 / 참고 자료 및 분석 도구         |

# 01. 개요

분석 배경 및 필요성

## 분석 배경



### 작년에 이어 호조세 지속

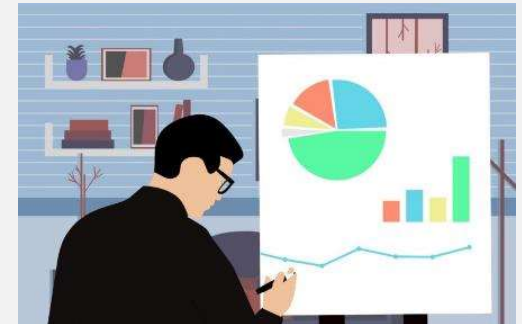
2022년 상반기 한국의 수출은 러·우 전쟁, 코로나 확산에 따른 중국 도시 봉쇄 등 어려운 대외 수출 여건에도 불구하고 역대 최고치를 갈아치우면서 호조세를 보이고 있음



### 한국 무역 전반에 불확실성 심화

수입량이 급증하면서 무역적자의 늪에 빠질 수 있으며 에너지 가격 급등에 따른 수입 급증으로 연달아 적자가 발생한 가운데, 글로벌 성장세 둔화와 공급망 불안정 심화 등 한국 무역의 전반에 불확실성이 심화되는 상황

## 필요성



### 수출 지원 서비스 제공 필요

고환율, 고유가 등에 따른 단가 상승, 지정학적 리스크에 대한 우려 등이 악재가 되고있는 현상황을 타개하기 위해 무역 데이터를 활용하여 수출 기업들에게 해외 시장 진출 및 신시장 개척을 할 수 있는 수출 지원 서비스를 제공할 필요가 있음

# 01. 개요

## 분석 프로세스

### Step 1. 데이터 수집

#### 내부 데이터 (1개)

- 한국 수출금액

#### 외부 데이터 (20개)

- 한국수입금액, 경기종합지수, 수출물가지수, 수입물가지수, 화물수송총괄, 코스피지수, 코스닥지수, 환율, 국제유가, 다우존스지수, 나스닥지수, 한국소비자물가지수, 실업률, 미국생산자물가지수, 전사업생산지수, 부산항 수출금액, 부산항 수입금액, 수출단가 지수, 반도체 수출금액, 무역경기확산지수

### Step 2. 데이터 탐색

#### 데이터 이해

- 특성별 요약 통계량
- 상관관계
- 시차 분포

### Step 3. 모델링 및 모델 평가 및 해석

#### 모델링

- 일변량 시계열 분석
  - > ARIMA, Holt Winters 지수평활, Seq2Seq
- 다변량 시계열 분석
  - > VECM, Seq2Seq

#### 모델 평가 및 해석

- 모형 성능 비교
- 최종 모형 선정 및 해석

### Step 4. 활용 방안 및 결론

#### 활용 방안

- 국가별 품목별 예상 수출금액 제공 서비스
- 서비스 활용 방법 설명

#### 결론

- 기대효과
- 한계점 및 제언
- 참고 자료 및 분석 도구

## 02. 분석내용

데이터 수집

원데이터



수집 데이터

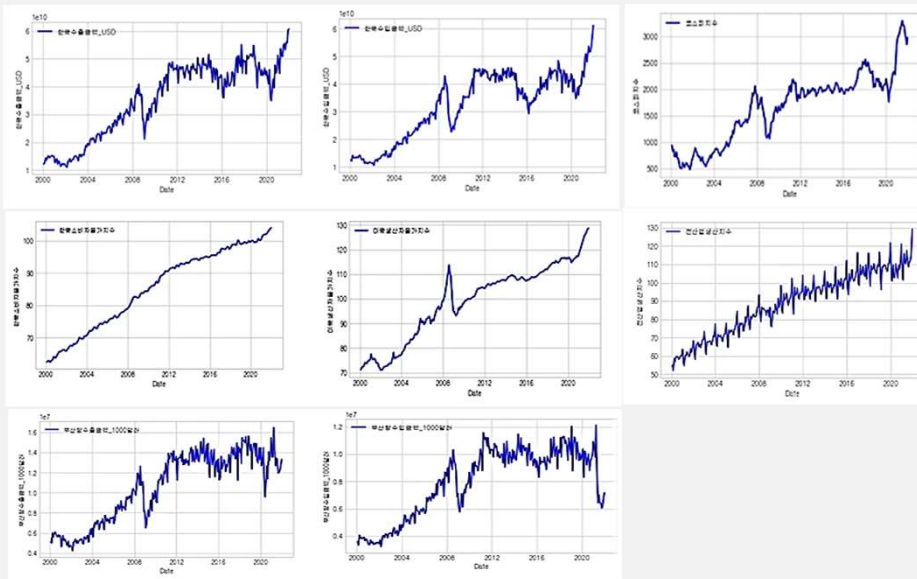
한국수입금액	미국 생산자물가지수
경기종합지수	전산업생산지수
수출물가지수	부산항수출금액
수입물가지수	부산항수입금액
화물수송총괄	반도체 수출금액
코스피지수	수출단가지수
코스닥지수	두바이유
다우존스지수	환율
나스닥지수	실업률
한국 소비자물가지수	무역경기확산지수

2000.01부터 2021.12까지의 전 기간을 학습 대상으로 지정하였고  
추가로 수집한 데이터까지 포함된 264행&21열의 형태로 구성하여 분석을 진행

## 02. 분석내용

데이터 탐색

### 시계열 데이터 분포



### 데이터 상관관계 표

한국수출액_USD	1	0.98	0.88	-0.27	0.76	0.83	0.93	-0.03	-0.23	0.67	0.7	0.62	0.95	0.95	0.94	-0.19	0.97	0.91	0.81	-0.22	0.047
한국수입액_USD	0.98	1	0.84	-0.18	0.82	0.81	0.93	-0.034	-0.26	0.74	0.69	0.61	0.92	0.95	0.91	-0.21	0.95	0.9	0.78	-0.14	0.043
코스피지수	0.88	0.84	1	-0.6	0.63	0.85	0.92	0.12	-0.033	0.31	0.92	0.86	0.96	0.93	0.94	-0.014	0.85	0.75	0.93	-0.62	-0.027
한국소비자물가지수	-0.27	-0.18	-0.6	1	0.3	-0.56	-0.32	0.03	-0.51	0.36	-0.54	-0.49	-0.51	-0.36	-0.49	-0.081	-0.3	-0.23	-0.46	0.93	0.19
미국생산자물가지수	0.76	0.82	0.63	0.3	1	0.5	0.65	-0.2	-0.42	0.99	0.22	0.15	0.6	0.68	0.58	-0.3	0.73	0.76	0.36	0.38	0.1
전산업생산지수	0.83	0.81	0.85	-0.56	0.5	1	0.8	-0.25	-0.057	0.42	0.68	0.55	0.89	0.85	0.89	-0.22	0.82	0.79	0.74	-0.51	-0.052
부산항수출액_1000달러	0.93	0.93	0.92	-0.32	0.65	0.8	1	0.15	-0.25	0.55	0.81	0.78	0.93	0.95	0.92	-0.059	0.89	0.81	0.86	-0.35	0.079
부산항수입액_1000달러	-0.03	-0.034	0.12	0.03	-0.2	-0.25	0.15	1	-0.073	-0.23	0.33	0.43	-0.031	0.0093	-0.019	0.61	-0.027	-0.14	0.24	-0.09	0.062
한국수출액/수입액	-0.23	-0.26	-0.033	-0.51	-0.42	-0.057	-0.25	-0.073	1	-0.43	-0.051	0.017	-0.072	-0.19	-0.11	0.13	-0.24	-0.29	-0.086	-0.43	-0.18
Dubai	0.67	0.74	0.31	0.36	0.89	0.43	0.55	-0.23	-0.43	1	0.12	0.051	0.5	0.58	0.48	-0.34	0.65	0.69	0.26	0.46	0.12
미국소비자물가지수	0.7	0.69	0.92	-0.54	0.22	0.68	0.81	0.33	-0.051	0.12	1	0.97	0.79	0.79	0.79	0.075	0.67	0.51	0.94	-0.64	-0.043
나스닥지수	0.62	0.61	0.88	-0.49	0.15	0.55	0.78	0.43	0.017	0.051	0.97	1	0.71	0.72	0.71	0.13	0.56	0.39	0.88	-0.63	0.0062
한국소비자물가지수	0.95	0.92	0.98	-0.51	0.6	0.89	0.93	-0.031	-0.072	0.5	0.79	0.74	1	0.97	0.97	-0.093	0.93	0.86	0.86	-0.47	-0.049
미국생산자물가지수	0.95	0.95	0.93	-0.36	0.68	0.85	0.95	0.0093	-0.19	0.58	0.79	0.72	0.97	1	0.95	-0.13	0.92	0.86	0.85	-0.36	-0.035
전산업생산지수	0.94	0.91	0.94	-0.49	0.58	0.89	0.92	-0.019	-0.11	0.48	0.79	0.71	0.97	0.95	1	-0.17	0.93	0.85	0.86	-0.47	0.017
실업률	-0.19	-0.21	-0.014	-0.091	-0.3	-0.22	-0.089	0.41	0.13	-0.34	0.075	0.13	-0.093	-0.13	-0.17	1	-0.14	-0.14	0.0016	-0.16	-0.087
부산항수출액_1000달러	0.97	0.95	0.85	-0.3	0.73	0.82	0.89	-0.027	-0.24	0.65	0.67	0.58	0.93	0.92	0.93	-0.14	1	0.95	0.77	-0.22	0.031
부산항수입액_1000달러	0.91	0.9	0.75	-0.23	0.76	0.79	0.81	-0.14	-0.29	0.69	0.51	0.39	0.86	0.86	0.85	-0.14	0.95	1	0.84	-0.13	0.019
반도체수출액_1000달러	0.61	0.78	0.93	-0.46	0.36	0.74	0.86	0.24	-0.086	0.26	0.94	0.88	0.86	0.85	0.86	0.0016	0.77	0.64	1	-0.54	0.0017
수출액/수입액	-0.22	-0.14	-0.62	0.93	0.38	-0.51	-0.35	-0.09	-0.43	0.46	-0.64	-0.63	-0.47	-0.36	-0.47	-0.16	-0.22	-0.13	-0.54	1	0.13
부채/총자산	0.047	0.043	-0.027	0.19	0.1	-0.052	0.079	0.062	-0.18	0.12	-0.043	0.0062	-0.049	-0.035	0.017	-0.087	0.031	0.019	0.0017	0.13	1

한국수출금액과 유사한 추세와 변동을 가진 변수 (상관계수 0.9 이상) 는 한국수입금액, 코스피지수, 한국소비자물가지수, 미국생산자물가지수, 전산업생산지수, 부산항수출/입금액으로 파악할 수 있음

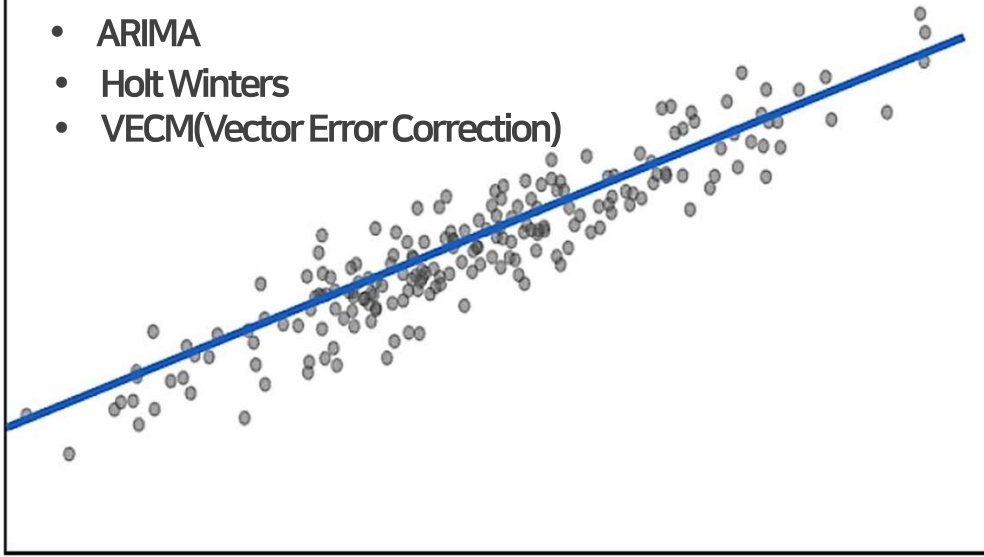
## 02. 분석내용

모델링

### Linear & Nonlinear

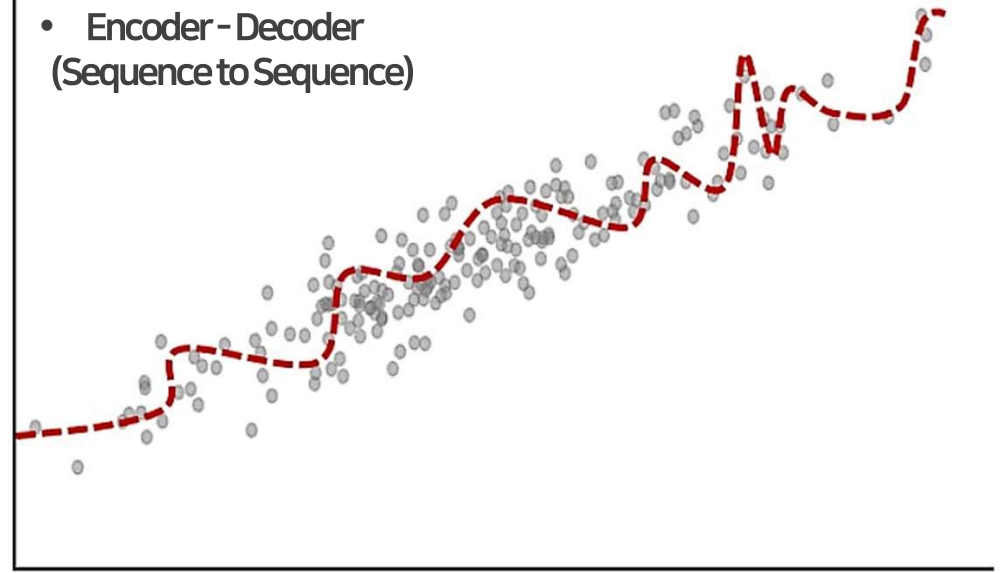
Linear

- ARIMA
- Holt Winters
- VECM(Vector Error Correction)



Nonlinear

- Encoder - Decoder  
(Sequence to Sequence)

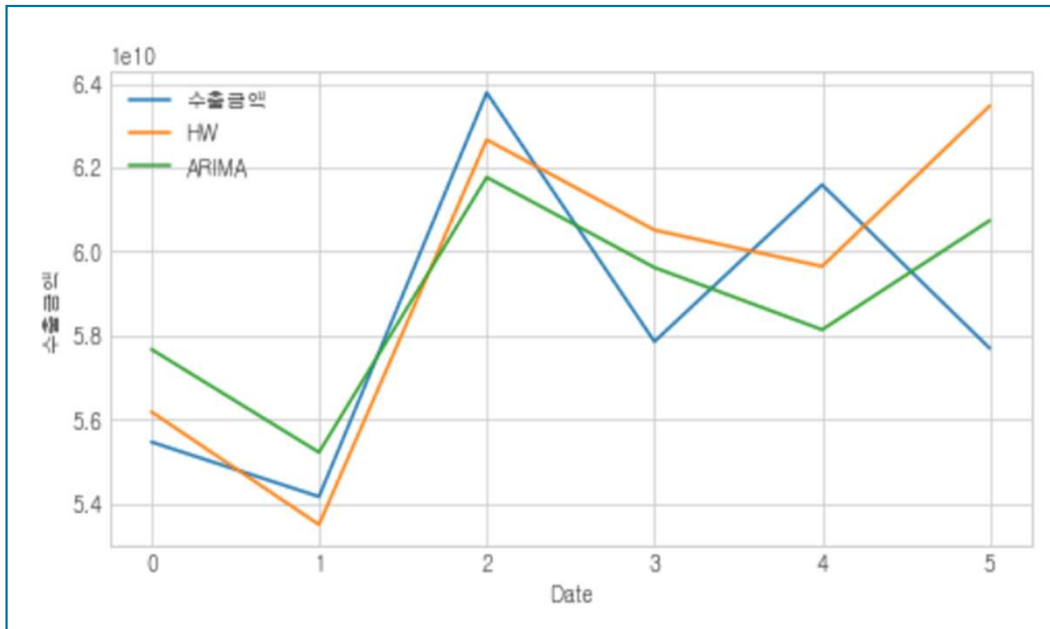


전통적 시계열은 **정상성, 오차의 정규성 & 독립성** 조건을 충족시켜야 되고 일정한 분포를 따르는 자료에서 좋은 성과를 보임  
딥러닝 시계열은 **통계적 제약이 따르지 않으며** 데이터의 **비선형 특징**을 학습하는 데에 수월하고 변동성이 높은 자료에 보다 적합

## 02. 분석내용

모델링

### 일변량시계열분석 (ARIMA & Holt Winters)



성능평가	RMSE	MAPE
ARIMA	2396668859	3.84
HoltWinters	2785746599	3.68

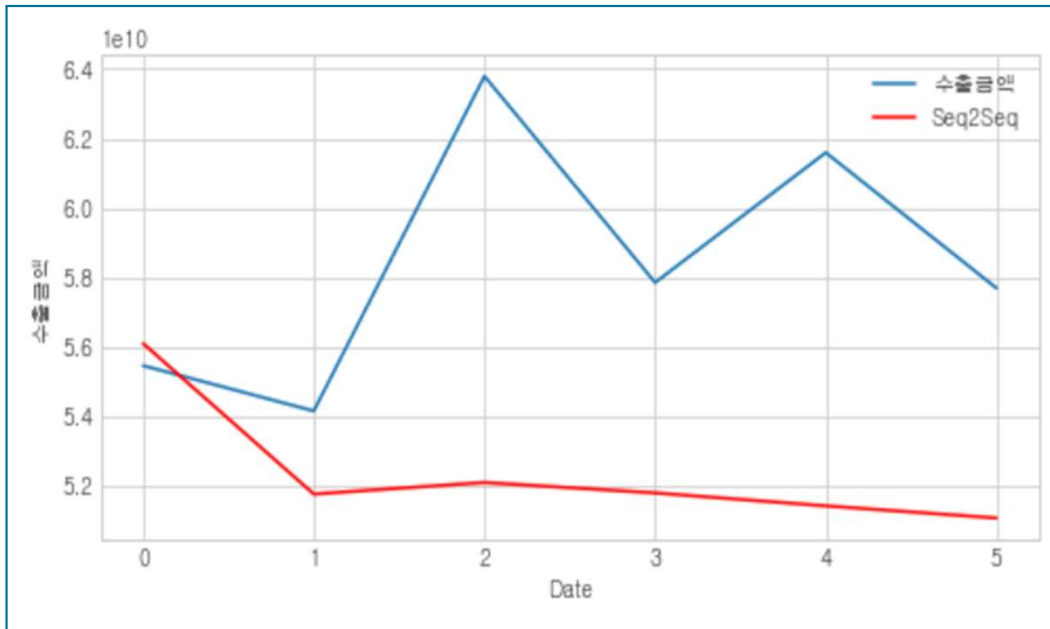
1월부터 3월 대비 4월부터 6월까지의 예측력은 높지 않으며  
전통적 단변량 시계열 분석의 통계적 제약으로 인한 결과로 추정 가능



## 02. 분석내용

모델링

### 일변량시계열분석(Seq2Seq\_Encoder-Decoder)



Scaler	Input Sequence	Output Sequence
MinMax(0,1)	5	6

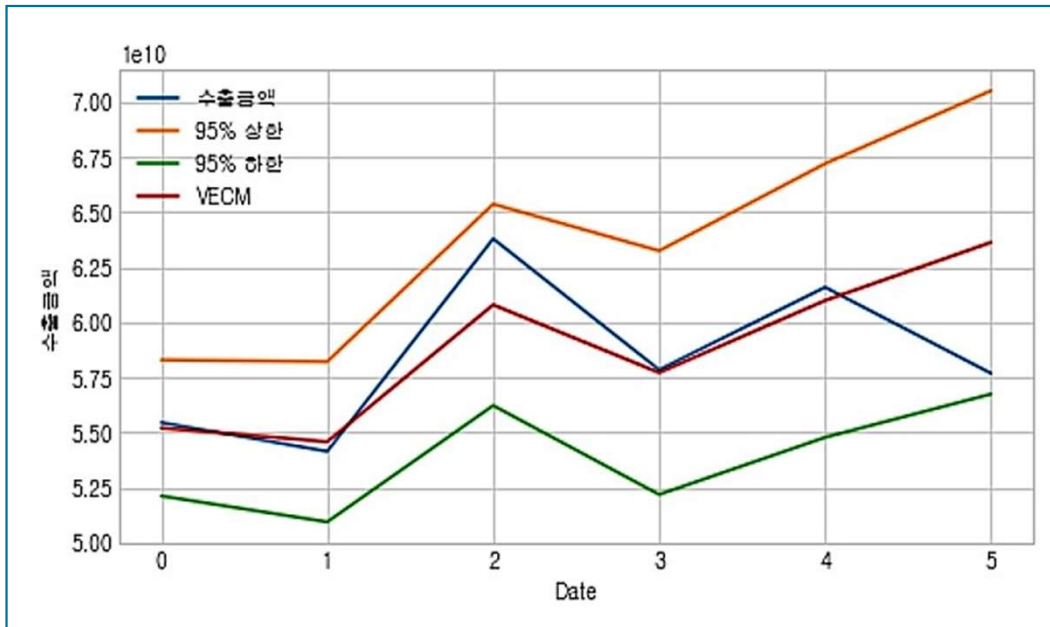
성능평가	RMSE	MAPE
Seq2Seq(단)	7379971350	10.39

원활한 딥러닝 학습을 위해서 **1970년 1월까지의 자료를 추가**하여 분석을 수행  
정확도는 떨어지나 단변량 통계 시계열과 다르게 **추세의 하락**을 보이는 점에서 **추가적 분석(다변량) 필요**

## 02. 분석내용

모델링

### 다변량시계열분석(VECM\_벡터오차수정모형)



ADF Test (유의수준 1% 기준)	공적분 개 수 (Trace & Max Eigen)	시차 (AIC기준)	Deterministic Term	계절 차수
17개변수 사용	5	3	Constant Inside('ci')	4

❖ 수출물가지수, 환율, 실업률, 무역경기확산지수 제거 (정상시계열 판정)

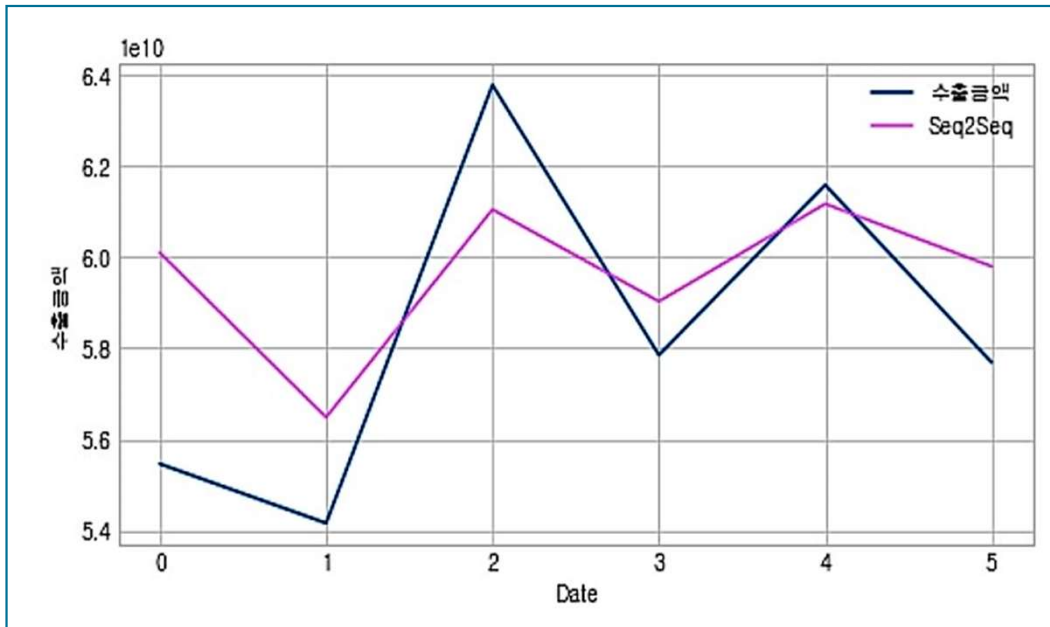
성능평가	RMSE	MAPE
VECM	2733172705	2.90

단변량 통계 시계열 대비 **개선된 추세와 변동을 가진 예측값**을 나타냄 (**6월 금액은 여전히 전월대비 상승으로 추정**)  
**예측값에 대한 신뢰구간** 또한 안정적인 흐름과 범위를 나타내기 때문에 이를 토대로 **VECM을 최종 예측 모델로 선정**

## 02. 분석내용

모델링

### 다변량시계열분석(Seq2Seq\_Encoder-Decoder)



- ❖ Input/Output & Scaling 단변량 Seq2Seq과 동일
- ❖ 한국수입금액, 경기종합지수, 한국소비자물가지수, 다우존스지수 수집 (1970.01~)
- ❖ 파생변수(무역수지)&불규칙요소(수출금액 분해요소 중1)추가
- ❖ 입력변수 7 & 예측변수 1개 구조
- ❖ 몬테카를로드롭아웃 기반 베이지안 딥러닝 추론
- ❖ 다변량 Seq2Seq 점추정값은 ARIMA, HW, VECM의 평균값을 벤치마킹

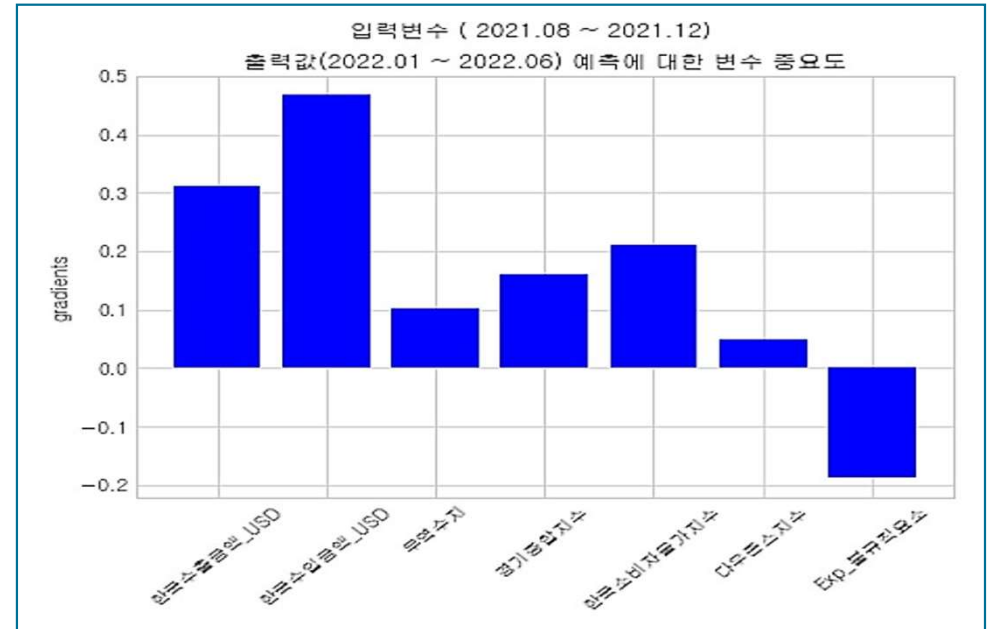
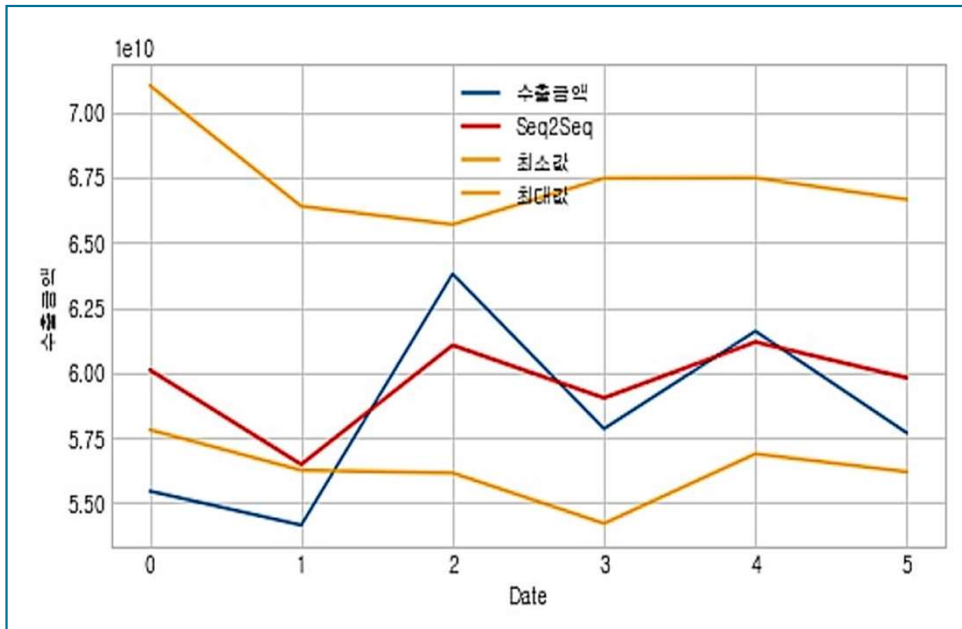
성능평가	RMSE	MAPE
Seq2Seq(다)	2598603277	3.89

다변량 Seq2Seq의 점추정 값의 변동성은 실제 수출금액과 전 구간에서 일치하고 유일하게  
전월대비 6월 금액의 하락을 예측한 결과를 나타냈으며 효율적으로 예측 및 위험관리에 대한 가능성을 보여줌

## 02. 분석내용

모델링

### 다변량시계열분석(딥러닝시계열발전방향)

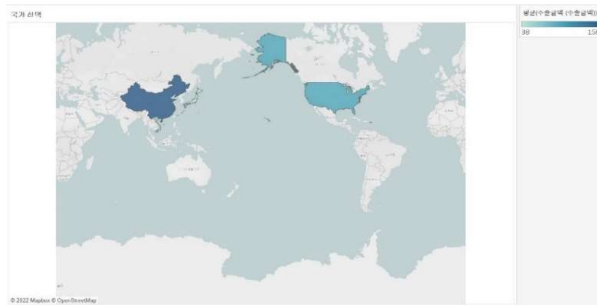


몬테카를로 드롭아웃 기반 베이지안 딥러닝 추론으로 VECM과 마찬가지로 예측값의 범위를 제공 가능  
변수 중요도 출력으로 설명 가능한 딥러닝 모델을 구축하여 VECM과의 상호보완적 활용을 기대할 수 있음

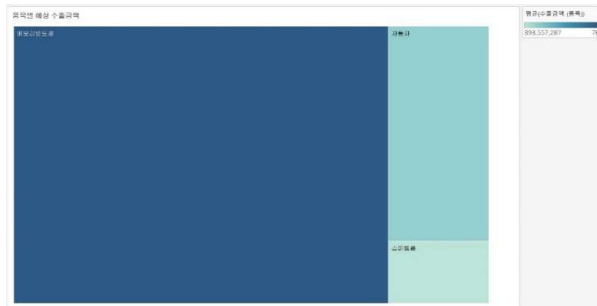
# 03. 활용 방안

서비스 화면

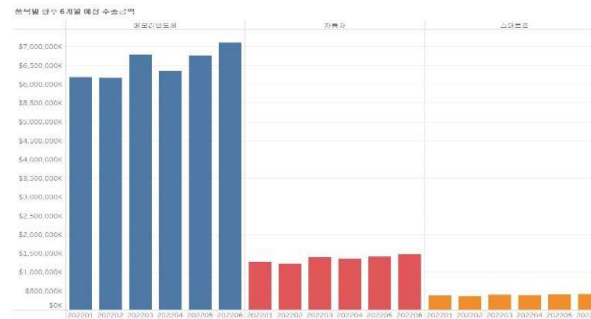
## 서비스화면구성



분석대상국가



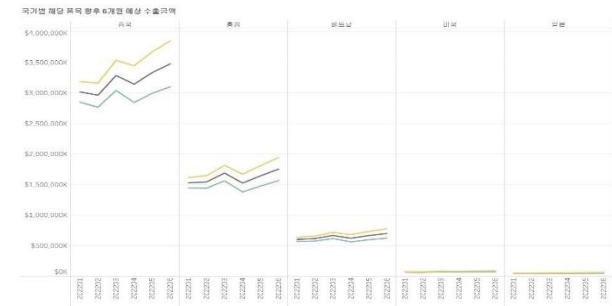
분석대상품목



품목별 향후 6개월 수출금액



국가별 예상 수출금액



국가별 해당 품목 예상 수출금액

## 03. 활용 방안

서비스 화면

### 국가/품목선택

#### 국가 & 품목 조회

<input type="checkbox"/>	국가코드	국가명
<input type="checkbox"/>	US	미국
<input type="checkbox"/>	VN	베트남
<input type="checkbox"/>	JP	일본
<input type="checkbox"/>	CN	중국
<input type="checkbox"/>	HK	홍콩

<input type="checkbox"/>	품목코드		품목명
<input type="checkbox"/>	8542	31	프로세서와 컨트롤러
<input type="checkbox"/>	8542	32	메모리
<input type="checkbox"/>	8542	33	증폭기
<input type="checkbox"/>	8542	39	기타
<input type="checkbox"/>	8543	10	임자가속기
<input type="checkbox"/>	8543	20	신호발생기
<input type="checkbox"/>	8543	30	전기도금융 기기
<input type="checkbox"/>	8543	40	개인용 기화 장치
<input type="checkbox"/>	8543	70	그 밖의 기기
<input type="checkbox"/>	8543	90	부분품

국가/품목선택화면

### 03. 활용 방안

서비스 화면(국가 선택)

#### ①수출 희망국가선택

국가 선택

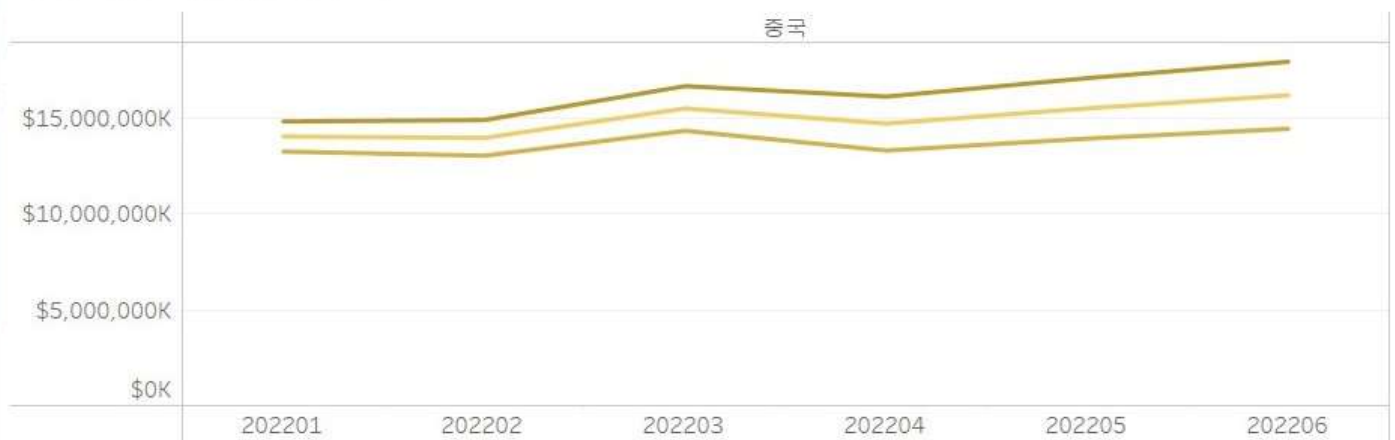


평균 수출금액 (수출금액)

2,710,277,736

14,994,280,284

국가별 향후 6개월 예상 수출금액

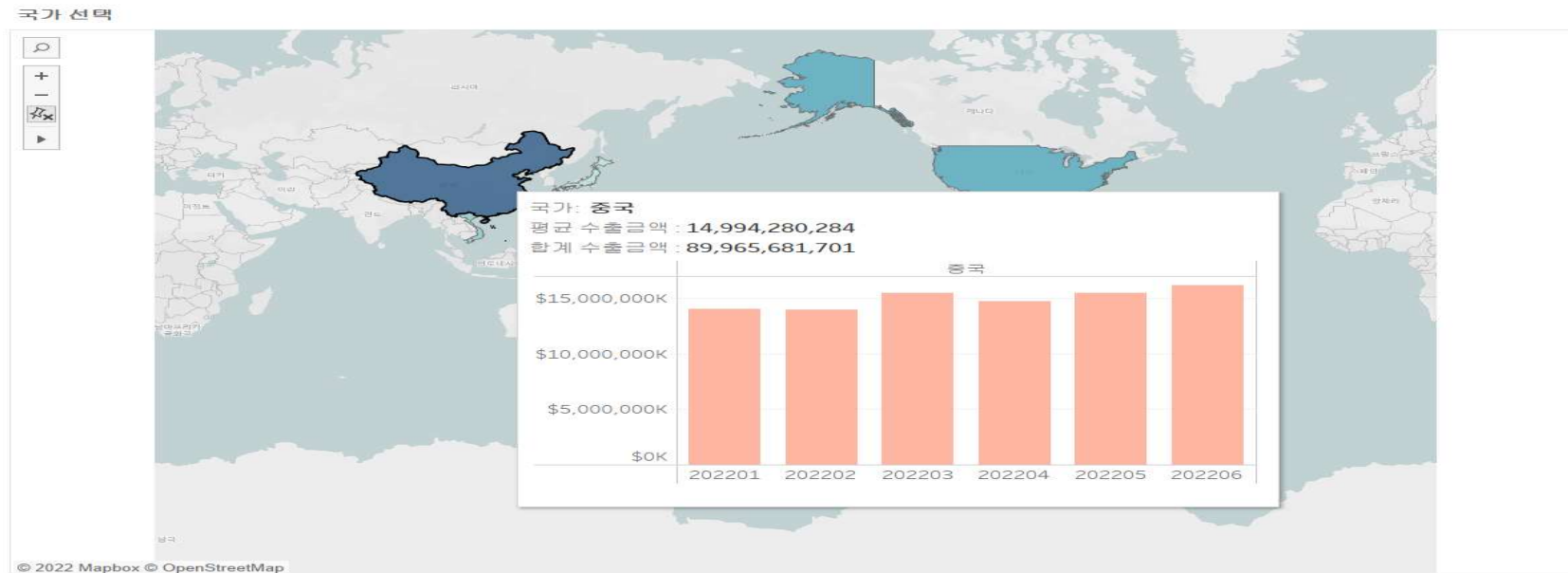


특정 국가만 지도에 클릭할 경우, 국가별 예상 수출금액의 **점추정액과 95% 상/하한 값**을 월별로 볼 수 있음

### 03. 활용 방안

서비스 화면(국가 선택)

#### ①수출 희망국가선택(국가별예상수출금액)



특정 국가에 커서를 놓을 경우, 해당 국가의 전체적인 수출금액의 평균/합계 예상액과 월별 점추정액을 볼 수 있음

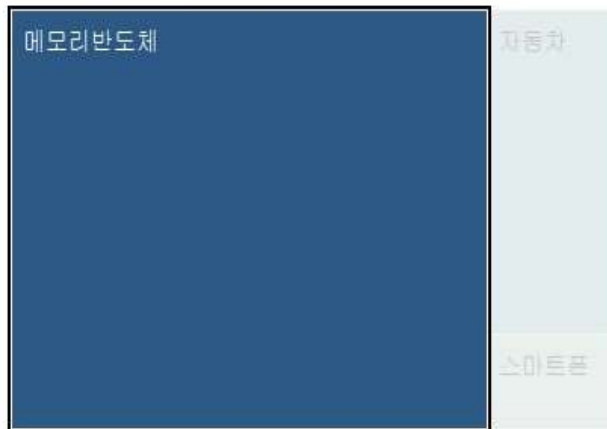


### 03. 활용 방안

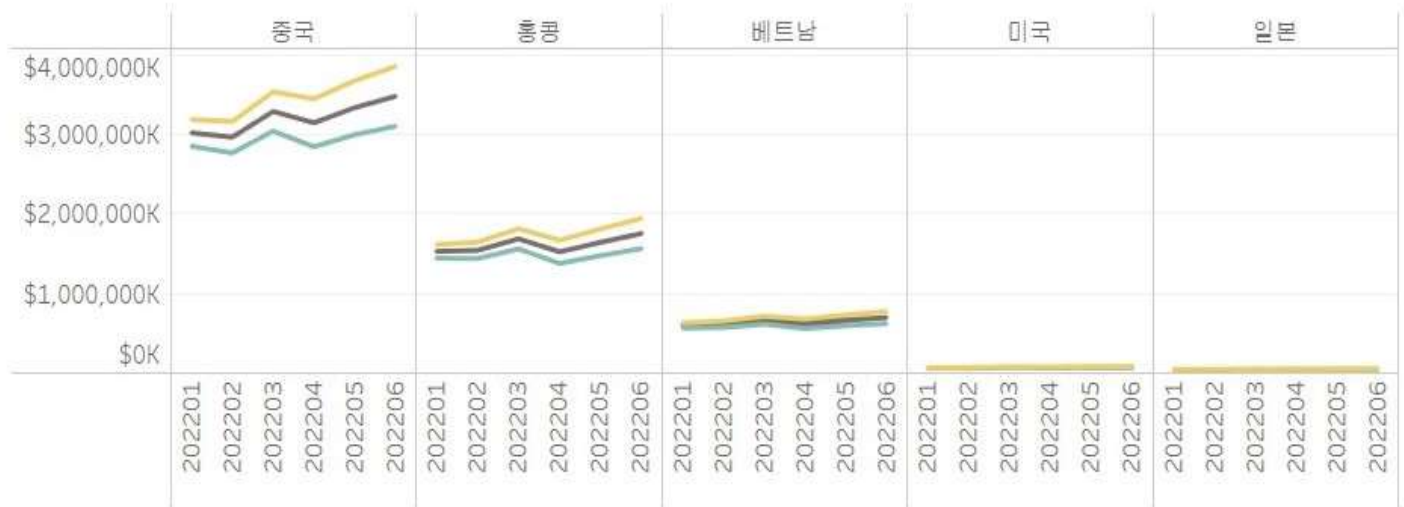
서비스 화면(품목 선택)

#### ② 수출 희망 품목 선택

품목별 예상 수출금액



국가별 해당 품목 향후 6개월 예상 수출금액



트리맵에 특정 품목을 선택할 경우, 국가별로 선택한 **품목의 6개월 수출금액의 점추정액과 95% 상/하한 값**을 월별로 볼 수 있음

### 03. 활용 방안

서비스 화면(품목 선택)

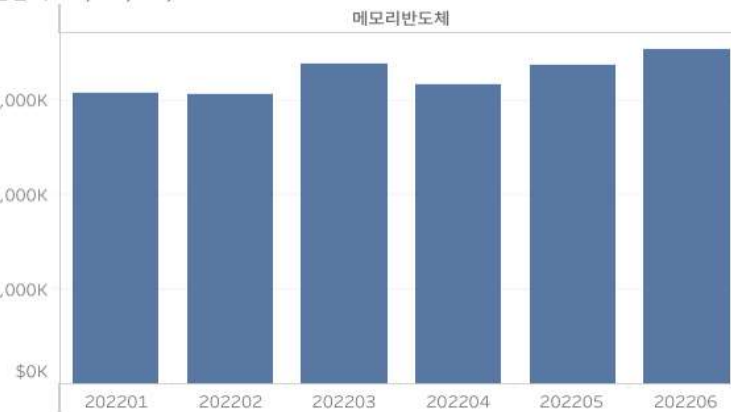
#### ② 수출 희망 품목 선택 (품목별 예상 수출금액)

품목별 예상 수출금액

국가별 해당 품목 향후 6개월 예상 수출금액

메모리반도체

품목 (품목) : 메모리반도체  
평균 수출금액 : 6,556,283,537  
합계 수출금액 : 39,337,701,222



특정 품목에 커서를 놓을 경우, 해당 품목의 **전체적인 예상 평균/합계 수출금액**과 **월별 예상 수출 점추정액**을 볼 수 있음

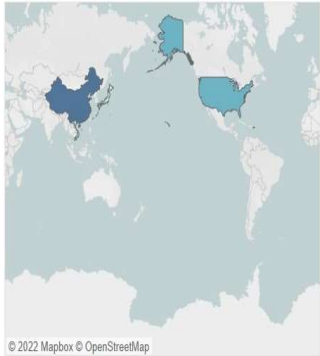
# 03. 활용 방안

서비스 화면(국가/품목 선택)

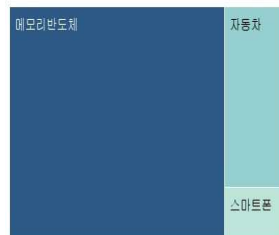
## ③ 다중 국가, 다중 품목 선택

해당 국가별 품목별 예상 수출금액

국가 선택



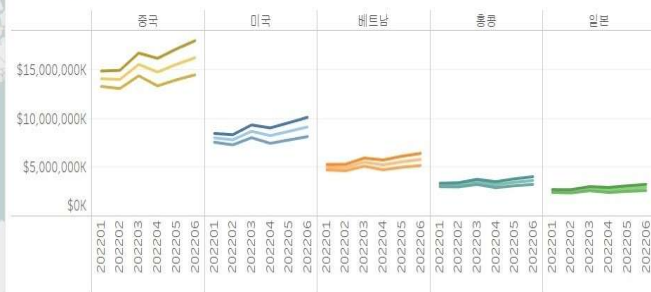
품목별 예상 수출금액



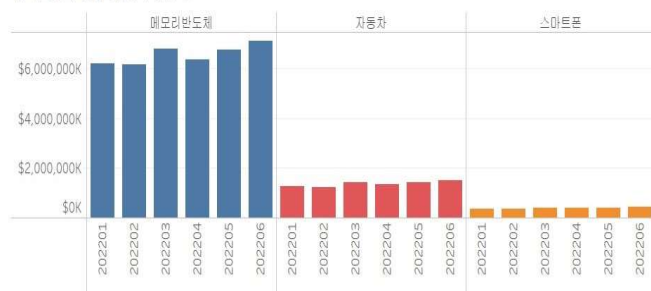
평균 수출금액 (수출금액)

2,710,277,736 14,994,280,284

국가별 향후 6개월 예상 수출금액



품목별 향후 6개월 예상 수출금액



- 앞서 선택한 국가와 품목에 대한 전체적인 정보를 보여줌
- **수출금액 기준**으로 국가별/품목별 수출금액 표를 **오름차순으로 정렬**함
- **트리맵**에서 **선택한 특정 품목의 전체 품목 대비 수출금액 비중**을 보여줌  
(품목별로 수출금액을 비교할 수 있음)
- 해당 국가와 품목의 **최근 6개월 수출금액 비중의 평균**을 토대로 국가별 품목별 예상 수출금액을 제공함
- 잘못된 예상에 대비하기 위해 **95% 신뢰구간 제공**

## 04. 결론

### 기대효과

#### 기대효과 ①



#### 국내 수출 기업의 해외 진출 확대

수출 목표 시장에 대한 예상 수출금액 제공 뿐만 아니라 수출 희망 품목의 예상 수출 금액을 제공함으로써 국내 수출 기업들의 시장 조사 편리성과 국내 수출 기업의 해외 진출 확대

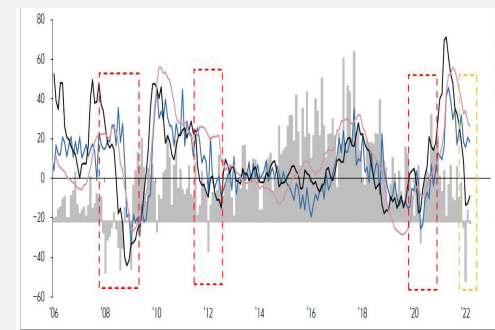
#### 기대효과 ②



#### 데이터 기반 의사결정에 도움

수출 초보 기업, 수출에 어려움을 겪고 있는 기업 등의 미래에 대한 매출 예상, 전략 도출, 수출 및 판매 계획 수립 등 데이터 기반 의사결정에 도움

#### 기대효과 ③



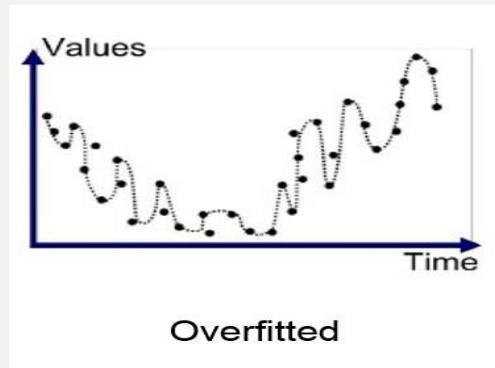
#### 수출 시장 변동성에 대비가능

예상 수출금액과 더불어 하한가와 상한가의 범위를 제공함으로써 해당 국가 또는 해당 품목의 수출 시장의 변동성에 수출기업들이 대비할 수 있음

## 04. 결론

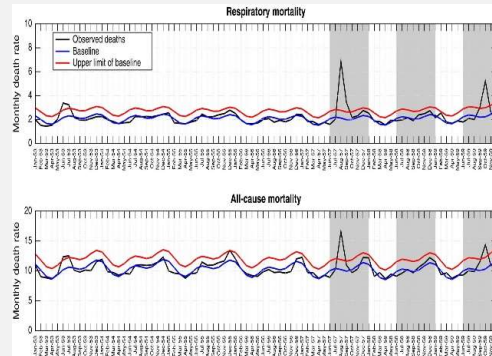
한계점 및 제언

### 한계점



- VECM은 통계적 제약이 따르므로 예측의 유연성 저하
- Seq2Seq는 비선형적 특징을 반영하지만 데이터가 부족할 경우 과적합 문제 발생

### 향후 과제



- 추가데이터 수집을 통한 다양한 변수 조합 연구 필요
- VECM과 같은 통계 기반 모델만으로 시계열의 변동성을 반영하기 어려우므로, 다변량 딥러닝 시계열(어텐션, 트랜스포머) 중심 모델 발전 필요

# 04. 결 론

## 참고 자료 및 분석 도구

### 참고 자료

- 강지원. "Encoder Decoder 알고리즘을 이용한 시계열 자료 예측", 경기대학교 석사학위 논문 (2021)
- 김두현. "수출과 경제성장간의 인과관계 분석: 오차수정모형", 연세대학교 석사학위 논문 (2006)
- 김두환, 이강배. "LSTM을 활용한 부산항 컨테이너 물동량 예측", 한국항만경제학회지 제36집 제2호 (2020)
- 김미현. "A Hybrid VECM and LSTM for Forecasting Multivariate Time Series", 이화여자대학교 석사학위 논문 (2021)
- 김준기. "Prophet 모형을 활용한 국내 중소형 컨테이너항만 물동량 예측에 관한 연구: 인천,평택,당진,울산항을 중심으로", 한국해양대학교 석사학위 논문 (2022)
- 민경찬, 하헌구. "SARIMA 모형을 이용한 우리나라 항만 컨테이너 물동량 예측", 대한교통학회지 제 32권 제6호 (2014)
- 산업통상자원부, 2022년 6월 및 상반기 수출입 동향
- 유원준. "딥 러닝을 이용한 자연어처리 입문", <https://wikidocs.net/24996>
- 유현종, 안승범. "인천내항, 인천북항, 평택항간 물동량의 인과관계 분석", 한국항만경제학회지 2014 제30집 제4호 (2014)
- 이성윤. "ARIMA와 VAR, VEC 모형에 의한 부산항 물동량 예측과 관련성 연구", 한국항해항만학회지 제 44권 제1호
- 이은주. "시계열 분해를 활용한 딥러닝 예측 모델 성능 향상에 관한 연구", 부산대학교 석사학위 논문 (2021)
- 이재득. "승법계절 ARIMA 모형에 의한 부산항 컨테이너 물동량 추정과 예측", 한국항만경제학회지 2013 제 29권 3호 (2013)
- 이종호. "베이지안 LSTM을 이용한 블랙리터만 모형의 개선 연구", 한국방송통신대학교 석사학위 논문 (2021)
- 장영진, 원종관, 이채록. "분리학습 모델을 이용한 수출액 예측 및 수출 유망국가 추천", J Intell Inform Syst 2022 March:28(1): 69-88 (2021)
- 조우성. "VECM모형을 활용한 거시경제변수가 성장에 미치는 영향분석", 통상정보연구 제14권 4호 (2012)
- 최봉호. "공적분분석을 이용한 지역 수출합수의 추정 - 부산지역을 중심으로 -", 국제통상연구 제9권 제 3호 (2004)
- 최봉호, 김상춘. "부산항, 광양항, 인천항의 물동량간 인과관계 분석", 한국항만경제학회지 제 26집 제 1호 (2010)
- 황유상. "한국의 수출금융과 수출보험이 수출에 미치는 영향에 관한 실증적 연구", 동아대학교 박사학위 논문 (2012)
- Yue Tan and Chunjin Hu. "LSTM-based Anomaly Detection for Non-linear Dynamical System", Electrical Engineering and Systems Science (2020)

### 분석 도구



TensorFlow



# Thank you!