

각 경로별 해상 운임 지수에 대한 중요 변수 분석 및 예측

EDA & 모델 학습 & 향후 계획

Contents

- EDA
- 모델 학습
- 향후 계획

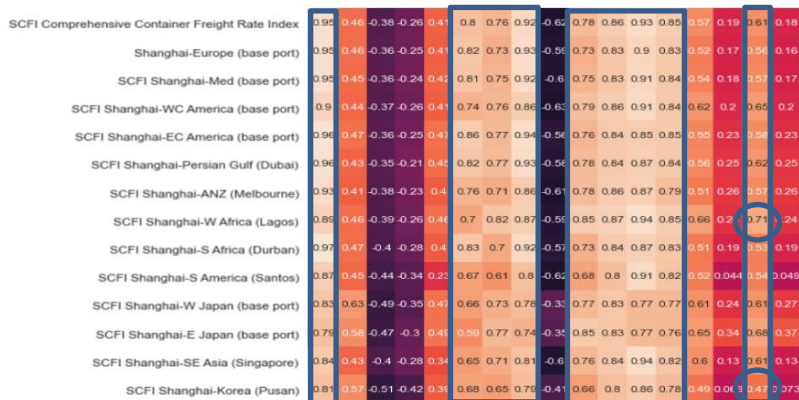
EDA

[Data]

- 경로별 SCFI
 - 외부 변수 (결측치에 대해서는 생각 x, 이유는 아래에 표기)
 - 항만 대기량
-
- 항만 대기량 데이터는 16년 1월 1일부터, SCFI 데이터는 09년 10월 16일부터, 외부 변수 데이터는 08년 1월 4일부터 보유
 - 항만 대기량 데이터를 사용하기 위해 16년 1월 1일 이후의 데이터만 추출 (16년 1월 1일 이후의 외부 변수 데이터에 결측치가 없음)
 - 전체 데이터 중에서 코로나 발생 이후 데이터만 활용(2019년 12월 31일 이후 데이터)
 - 우리가 예측하는 y는 SCFI 지수
 - SCFI : 09년 10월 16일~, 1주일 간격
 - 외부 변수 : 08년 1월 4일~, 1주일 간격
 - 항만 대기량 : 16년 1월 1일~, 하루 간격

=> 각 데이터의 추출은 16년 1월 1일부터(항만 대기량 데이터 사용 위해) 간격은 일주일로 예측(SCFI 예측 위해)
 - 데이터 전처리
 - Object(.) -> Float로 타입 변환, Datetime으로 변환
 - 분석을 위해 2016년 1월 1일 이후의 SCFI, 외부 변수, 항만 대기량 데이터 통합

전체 기간(2016.01.01~)



- 각 경로에 대한 모든 feature(외부변수&항만대기량)의 상관관계
- 같은 feature여도 SCFI에 주는 영향이 다름(동그라미)
- 특정 SCFI에 대해서 어떤 변수를 사용할지 분석
- 307개의 데이터

Clarksons Average Containership Earnings
 HSFO 380cst Bunker Prices (3.5% Sulphur), Rotterdam
 Total Containerships - % Idle/Laid Up/Scrubber Retrofit
 Total Containerships - % Idle/Laid Up/Scrubber Retrofit, 1
 Containership 1,650/1,850 TEU FCC, Glass Newbuilding Prices
 Containership 13,000/14,000 TEU Newbuilding Prices
 Containership 3,500/4,000 TEU (Wide Beam) Glass Newbuilding Prices
 Containership 13,000/13,500 TEU Glass Newbuilding Prices
 5 Year \$/Tm Finance based on Labor 1st yr
 Port Congestion Index - Containerships in Port, m TEU, 7dma
 Port Congestion Index - Containerships in Port, East Coast North America, m TEU, 7dma
 Port Congestion Index - Containerships in Port, West Coast North America, m TEU, 7dma
 Port Congestion Index - Containerships in Port, United Kingdom/Continent, m TEU, 7dma
 Port Congestion Index - Containerships in Port, Mediterranean/Black Sea, m TEU, 7dma
 Port Congestion Index - Containerships in Port, East Asia, m TEU, 7dma
 Port Congestion Index - Containerships in Port, South East Asia, m TEU, 7dma
 Port Congestion Index - Containerships in Port, China P.R., m TEU, 7dma

각 경로끼리의 상관관계

SCFI Comprehensive Container Freight Rate Index	1	0.99	0.99	0.96	0.97	0.98	0.95	0.96	0.98	0.96	0.84	0.81	0.95	0.86	0.95
Shanghai-Europe (base port)	0.99	1	1	0.93	0.96	0.98	0.93	0.95	0.98	0.96	0.82	0.79	0.93	0.85	0.95
SCFI Shanghai-Med (base port)	0.99	1	1	0.93	0.96	0.98	0.93	0.96	0.97	0.96	0.82	0.8	0.94	0.85	0.95
SCFI Shanghai-WC America (base port)	0.96	0.93	0.93	1	0.96	0.92	0.93	0.92	0.92	0.89	0.81	0.79	0.9	0.82	0.9
SCFI Shanghai-EC America (base port)	0.97	0.96	0.96	0.96	1	0.97	0.93	0.92	0.96	0.89	0.83	0.81	0.87	0.8	0.96
SCFI Shanghai-Persian Gulf (Dubai)	0.98	0.98	0.98	0.92	0.97	1	0.92	0.95	0.97	0.93	0.84	0.83	0.9	0.81	0.96
SCFI Shanghai-ANZ (Melbourne)	0.95	0.93	0.93	0.93	0.93	0.92	1	0.89	0.96	0.9	0.8	0.78	0.9	0.82	0.93
SCFI Shanghai-W Africa (Lagos)	0.96	0.95	0.96	0.92	0.92	0.95	0.89	1	0.92	0.93	0.84	0.86	0.94	0.84	0.89
SCFI Shanghai-S Africa (Durban)	0.98	0.98	0.97	0.92	0.96	0.97	0.96	0.92	1	0.95	0.84	0.8	0.9	0.85	0.97
SCFI Shanghai-S America (Santos)	0.96	0.96	0.96	0.89	0.89	0.93	0.9	0.93	0.95	1	0.79	0.77	0.94	0.84	0.87
SCFI Shanghai-W Japan (base port)	0.84	0.82	0.82	0.81	0.83	0.84	0.8	0.84	0.84	0.79	1	0.98	0.77	0.77	0.83
SCFI Shanghai-E Japan (base port)	0.81	0.79	0.8	0.79	0.81	0.83	0.78	0.86	0.8	0.77	0.98	1	0.76	0.73	0.79
SCFI Shanghai-SE Asia (Singapore)	0.95	0.93	0.94	0.9	0.87	0.9	0.9	0.94	0.9	0.94	0.77	0.76	1	0.9	0.84
SCFI Shanghai-Korea (Pusan)	0.86	0.85	0.85	0.82	0.8	0.81	0.82	0.84	0.85	0.84	0.77	0.73	0.9	1	0.81

각 경로끼리도 높은 상관관계를 확인할 수 있음
=> 추가 분석을 할 때 활용될 가치가 높음

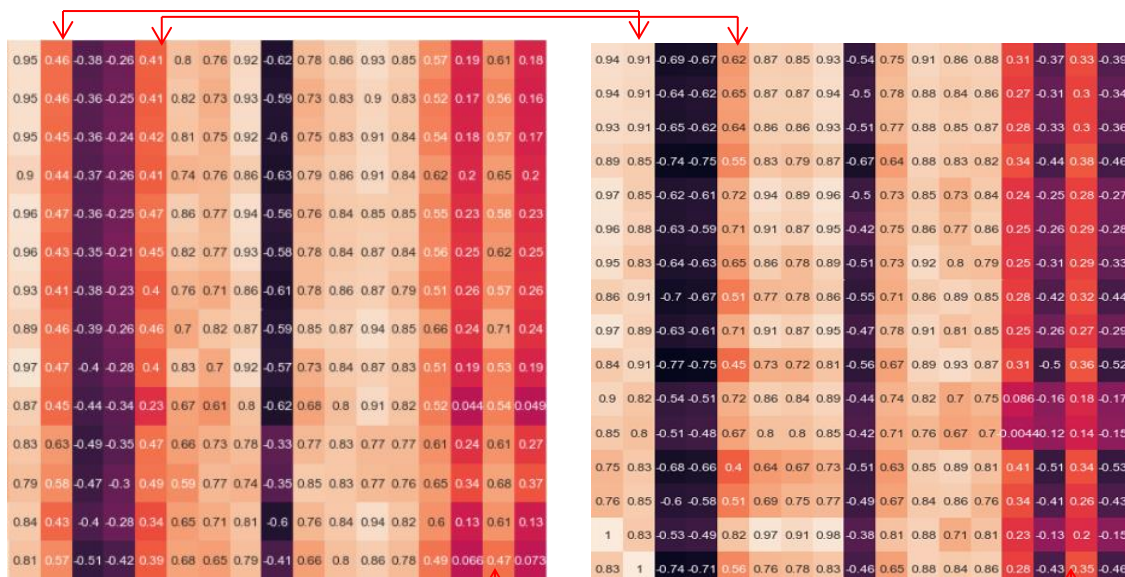
코로나 발생 이후(2019.12.31~

SCFI Comprehensive Container Freight Rate Index	0.94	0.91	1.69	-0.6	0.62	0.87	0.85	0.5	-0.5	0.75	0.91	0.86	0.88	0.31	-0.37	0.33	-0.39
Shanghai-Europe (base port)	0.94	0.91	1.64	-0.6	0.85	0.87	0.87	0.5	-0.5	0.78	0.88	0.84	0.86	0.27	-0.31	0.3	-0.34
SCFI Shanghai-Med (base port)	0.93	0.91	1.65	-0.6	0.84	0.86	0.86	0.5	-0.5	0.77	0.88	0.85	0.87	0.28	-0.33	0.3	-0.36
SCFI Shanghai-WC America (base port)	0.89	0.85	1.74	-0.7	0.55	0.83	0.79	0.5	-0.4	0.64	0.88	0.83	0.82	0.34	-0.44	0.36	-0.46
SCFI Shanghai-EC America (base port)	0.97	0.85	1.62	-0.6	0.72	0.94	0.89	0.5	-0.5	0.73	0.85	0.73	0.84	0.34	-0.25	0.28	-0.27
SCFI Shanghai-Persian Gulf (Dubai)	0.96	0.88	1.63	-0.5	0.71	0.91	0.87	0.5	-0.4	0.75	0.86	0.77	0.86	0.25	-0.26	0.29	-0.28
SCFI Shanghai-ANZ (Melbourne)	0.95	0.83	1.64	-0.6	0.85	0.86	0.78	0.5	-0.5	0.73	0.92	0.8	0.79	0.25	-0.31	0.29	-0.33
SCFI Shanghai-W Africa (Lagos)	0.86	0.91	1.7	-0.6	0.51	0.77	0.78	0.5	-0.5	0.71	0.86	0.89	0.85	0.26	-0.42	0.32	-0.44
SCFI Shanghai-S Africa (Durban)	0.97	0.89	1.63	-0.6	0.71	0.91	0.87	0.5	-0.4	0.78	0.91	0.81	0.85	0.25	-0.26	0.27	-0.29
SCFI Shanghai-S America (Santos)	0.84	0.91	1.77	-0.7	0.45	0.73	0.72	0.5	-0.5	0.67	0.89	0.93	0.87	0.31	-0.5	0.36	-0.52
SCFI Shanghai-W Japan (base port)	0.9	0.82	1.54	-0.5	0.72	0.86	0.84	0.5	-0.4	0.74	0.82	0.7	0.75	0.08	-0.16	0.18	-0.17
SCFI Shanghai-E Japan (base port)	0.85	0.8	1.51	-0.4	0.67	0.8	0.8	0.5	-0.4	0.71	0.76	0.67	0.7	0.04	-0.12	0.14	-0.15
SCFI Shanghai-SE Asia (Singapore)	0.75	0.83	1.68	-0.6	0.4	0.14	0.67	0.7	-0.5	0.63	0.85	0.89	0.81	-0.41	-0.51	0.34	-0.53
SCFI Shanghai-Korea (Pusan)	0.76	0.85	1.6	-0.5	0.71	0.69	0.75	0.7	-0.4	0.67	0.84	0.86	0.76	0.34	-0.41	0.26	-0.43
	1	0.83	1.53	-0.4	0.82	0.37	0.91	0.5	-0.3	0.81	0.88	0.71	0.81	0.23	-0.13	0.2	-0.15
	0.83	1	1.74	-0.7	0.56	0.76	0.78	0.5	-0.4	0.65	0.88	0.84	0.86	0.28	-0.43	0.35	-0.46

Clarksons Average Containership Earnings
 HSFO 380cst Bunker Prices (3.5% Suptur), Rotterdam
 Total Containerships - % Idle/Laid Up/Scrubber Retrofit
 Total Containerships - % Idle/Laid Up/Scrubber Retrofit.1
 Containership 1,650/1,850 TEU FCC, Glass Newbuilding Prices
 Containership 13,000/14,000 TEU Newbuilding Prices
 Containership 3,500/4,000 TEU (Wide Beam) Glass Newbuilding Prices
 Containership 13,000/13,500 TEU Glass Newbuilding Prices
 5 Year \$/ton Finance based on Labor 1st yr
 Port Congestion Index - Containerships in Port, m TEU / 7dms
 Port Congestion Index - Containerships in Port, East Coast North America, m TEU / 7dms
 Port Congestion Index - Containerships in Port, West Coast North America, m TEU / 7dms
 Port Congestion Index - Containerships in Port, United Kingdom Continent, m TEU / 7dms
 Port Congestion Index - Containerships in Port, Mediterranean/Black Sea, m TEU / 7dms
 Port Congestion Index - Containerships in Port, East Asia, m TEU / 7dms
 Port Congestion Index - Containerships in Port, South East Asia, m TEU / 7dms
 Port Congestion Index - Containerships in Port, China P.R., m TEU / 7dms

- 각 경로에 대한 모든 feature(외부변수&항만대기량)의 상관관계
- => 전체 기간을 다룰 때와 상관 관계가 높은 변수가 다름
- 같은 feature여도 SCFI 주는 영향이 다름(동그라미)
- 특정 SCFI에 대해서 어떤 변수를 사용할지 분석
- 104개의 데이터 (과소적합 가능성)
- 코로나 기간 동안 각 SCFI의 추세를 예측하는 것이 목표
기 때문에 이 데이터를 활용하여 분석

전체 기간을 다룰 때와 코로나 이후에 영향을 끼치는 변수가 다른 것을 확인 가능



<전체 기간>

<코로나 발생 이후>

모델 학습

코로나 발생 이후 학습 및 예측

- 2019년 1월 1일 이후의 104개의 데이터 활용
- 전체 SCFI(**SCFI Comprehensive Container Freight Rate Index**)를 y로 두고 학습 및 예측
 - 각 경로에 대한 SCFI를 분석하기 전에 전체 SCFI를 활용하여 학습과 분석이 잘 되는지 확인
- 활용한 feature(=코로나 발생 이후 기간과 상관 관계가 높았던 변수) => 총 10개
 - 'Clarksons Average Containership Earnings' (0.94)
 - 'HSFO 380cst Bunker Prices (3.5% Sulphur), Rotterdam' (0.91)
 - 'Containership 1,650/1,850 TEU FCC, G'less Newbuilding Prices' (0.62) => 약간 낮은 수치이지만 일단 포함
 - 'Containership 13,000/14,000 TEU Newbuilding Prices' (0.87)
 - 'Containership 3,500/4,000 TEU (Wide Beam) G'less Newbuilding Prices' (0.85)
 - 'Containership 13,000/13,500 TEU G'less Newbuilding Prices' (0.93)
 - 'Port Congestion Index - Containerships In Port, m.TEU, 7dma' (0.75)
 - 'Port Congestion Index - Containerships In Port, East Coast North America, m.TEU, 7dma' (0.91)
 - 'Port Congestion Index - Containerships In Port, West Coast North America, m.TEU, 7dma' (0.86)
 - 'Port Congestion Index - Containerships In Port, United Kingdom/Continent, m.TEU, 7dma' (0.88)
- MinMaxScaler를 활용하여 전처리

모델 학습

코로나 발생 이후 학습 및 예측

Window_size : 내가 얼마동안(기간)의 다음날 데이터를 예측할 것인가를 정하는 parameter

- 데이터 분리
 - 총 104개의 데이터에서 80개를 Train, 나머지 24개를 Test 데이터셋으로 활용
 - Window_size를 12로 하여 학습 데이터 생성
 - **Window_size = 12 : 데이터 하나가 1주일이니 이것이 12개 모여 3달 => 3달씩 학습**
- (간단한)모델 생성

```
model = Sequential()  
model.add(LSTM(16,  
               input_shape=(train_feature.shape[1], train_feature.shape[2]),  
               activation='relu',  
               return_sequences=False)  
)  
model.add(Dense(1))  
model.summary()
```

Model: "sequential_6"

Layer (type)	Output Shape	Param #
lstm_6 (LSTM)	(None, 16)	1728
dense_6 (Dense)	(None, 1)	17

Total params: 1,745

Trainable params: 1,745

Non-trainable params: 0

모델 학습

코로나 발생 이후 학습 및 예측

- 모델 학습
 - 에폭 100, 배치사이즈 8로 학습
 - Loss : 0.3187, val_loss : 0.2939 => loss : 0.0019, val_loss : 0.0025까지 학습

```
# 모델의 학습
model.compile(loss='mean_squared_error', optimizer='adam')
early_stop = EarlyStopping(monitor='val_loss', patience=5)
filename = os.path.join(model_path, 'tmp_checkpoint.h5')
checkpoint = ModelCheckpoint(filename, monitor='val_loss', verbose=1, save_best_only=True, mode='auto')

history = model.fit(X_train, y_train,
                    epochs=100,
                    batch_size=8,
                    validation_data=(X_test, y_test),
                    callbacks=[early_stop, checkpoint])
```

Epoch 1/100

8/8 [=====] - 3s 102ms/step - loss: 0.3187 - val_loss: 0.2939

Epoch 00001: val_loss improved from inf to 0.29389, saving model to C:/Users/hyunj/KM1/tmp_checkpoint.h5

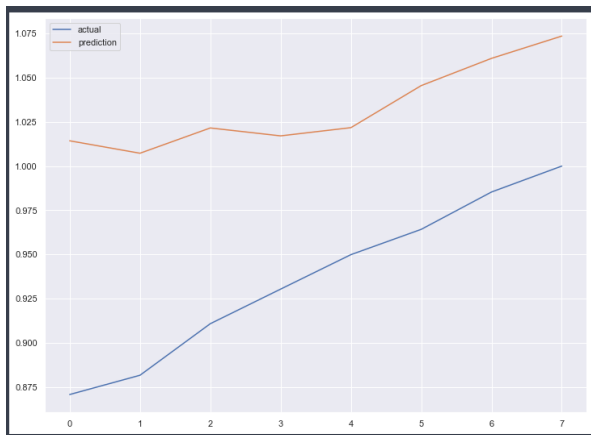
Epoch 2/100

8/8 [=====] - 0s 7ms/step - loss: 0.1932 - val_loss: 0.1934

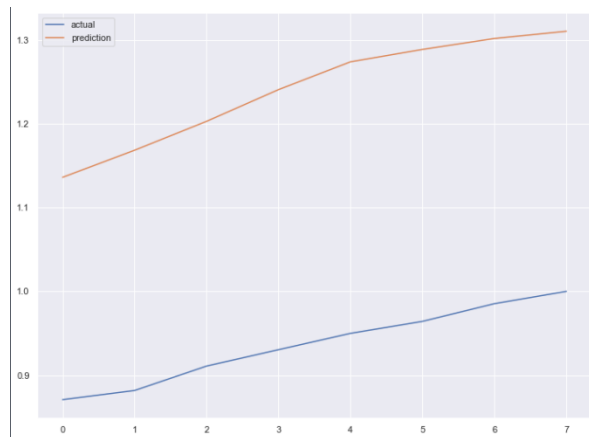
모델 학습

코로나 발생 이후 학습 및 예측

- 모델 예측



<16 hidden layer, 100 epoch, batch size 8 >



<16 hidden layer, 200 epoch, batch size 16 >

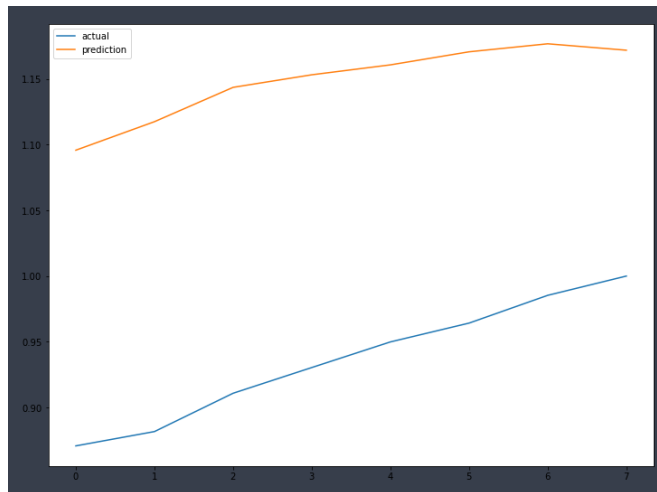
모델 학습

코로나 발생 이후 학습 및 예측

실장님께서 port congestion 변수 중에서 가장 중요한 것은 **LA**라고 설명

LA = West Coast이기 때문에 Port congestion 변수 중에서 '**Port Congestion Index - Containerships In Port, West Coast North America, m.TEU, 7dma**' 이 변수만 사용한 후 학습

- 모델 예측



<16 hidden layer, 200 epoch, batch size 4 >

향후 계획

- ‘외부 변수 결측치를 개선하여 2016년 1월 1일 이후의 SCFI 데이터 분석’ 혹은 ‘2019년 12월 31일 코로나 발생 이후 SCFI 데이터 분석’ 중 확실하게 방향 결정하기
 - 전자면 결측치에 대한 도메인 지식 도움 부탁드립니다, 후자면 부족한 데이터 보완 방안 마련
- 시계열 분석 모델(CNN, LSTM, GRU, ATTENTION 등) 공부하여 예측이 더 잘되는 모델 생성
 - => 각 경로에 따른 모델 각각 생성
- 각 경로에 대한 SCFI를 예측하기 위해 영향력 있는 변수를 뽑아내어 사회적 요인과 관련하여 분석하는 방향도 좋지만 **차별화**된 방안 생각해보기
- 한국 무역 협회에서 발표한 2021년 4분기의 SCFI 지수 급등 원인 중에서 물류현장에서 직원들의 코로나 확진도 있었음 => 크롤링 등 방법을 강구해서 활용