

앙상블 기법을 활용한 2019년 품목, 나라 별 한국 수입액 예측

Prediction of Korean Imports by Item and Country in 2019 Via Ensemble Learning

CUAI 4기 경영경제 A팀

강영훈(경영학부), 김민주(경영학부), 정옥준(경영학부)

[요약] 본 연구의 목적은 공공데이터를 활용해 품목, 나라 별 한국 수입액을 예측하는 데 있다. 본 연구에서는 물품의 HS code와 국가명에 따른 한국으로부터 수입액 관련 22개의 파생변수를 통해 차년도 한국에서 해당 국가에게 해당 품목을 수출한 금액을 예측한다. Ensemble Learning을 활용한 예측 결과 Random Forest Model에서 0.6921의 LMSE값을 보였다.

1. 서 론

2019년 대·중견·중소 규모별 국내 수출입 기업 숫자는 모두 2018년 대비 증가했지만, 이들 가운데 중소기업만 무역수지 적자를 면하지 못한 것으로 나타났다. 중소기업 수출 기업체 수는 9만4529개로 전년 대비 1.1% 증가했고 수출규모는 998억 달러(109조3109억 원)로 전년 같은 기간과 비교해 3.5% 줄어 들었다. 수입 기업 수는 18만9114개로 전년 대비 4.2% 증가했으며 수입규모는 1169억 달러(128조405억 원)로 8.7% 감소하였다. 이는, 기업 규모별로 상이한 정보력의 차이와 이에 따른 대응 전략의 부족이라고 예측할 수 있다.

또한 미중 무역분쟁, 동남아시아에 대한 무역 비중의 증가, 대중국 수출의 감소세, 인도 등 잠재적 신흥국의 수출 비중 증가 등 급변하는 국제 무역 현황에서 무역 데이터를 기반으로 한 수입액 예측은, 국내 유망 품목 선정 및 무역 정책에 데이터 기반 의사결정에 핵심적인 근거가 되어가고 있다.

국내 무역 품목의 수출 실적에 어떠한 변수가 영향을 미치는지 분석함으로써 국내 무역 수출입 구조의 체질 개선과 개선에 도움을 줄 수 있다. 이러한 체질 진단은 코로나 상황과 같은 특수 상황에 있어서, 국내 무역 상황에 대한 대응책 제시와 관련 정부 정책 마련에 핵심 기반이 될 수 있다.

본 연구는 기존에 진행되던 수입액 예측 분석과는 달리 국가별, 품목별을 모두 고려한 분석 결과로 유망 품목 선정뿐만 아니라 어떤 나라가 어떤 품목 수출에 유망한지, 혹은 어떤 품목은 어떤 나라에서

유망한지 등의 2 track의 수입액 예측 구조를 갖고 있다는 데에 그 의의가 있다

2. 본 론

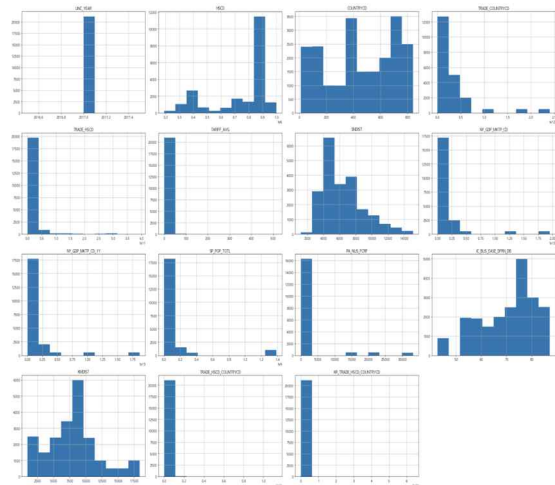
*내용 왼쪽 정렬

*자료 삽입 시 그림은 아래 캡션, 표는 위 캡션 이용, 캡션은 8p, 왼쪽 정렬로 설정

*필요에 따라 아래와 같이 소제목 사용 (구분없이 작성하여도 무관)

1) Exploratory Data Analysis

주어진 데이터의 특성을 우선적으로 확인하기 위해 탐색적 데이터 분석을 통해 데이터의 분포와 변수 간의 관계 등 데이터의 특성을 이해하였다.

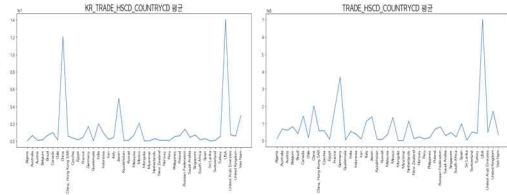


<figure 1. 각 변수 별 분포 시각화>

전체 데이터의 분포를 확인한 결과 상대적으로 한쪽으로 치우친 경향의 데이터 존재를 확인하였다.

<figure 1>을 통한 경향성 파악을 바탕으로 각 국가들 기준으로 변수 특성에 초점을 맞추어 시각화를 진행하였다.

제공된 공공데이터 각 변수들을 국가별로 시각화하여, 기존 변수를 바탕으로 한 파생 변수 설정의 중요성을 파악하였다.



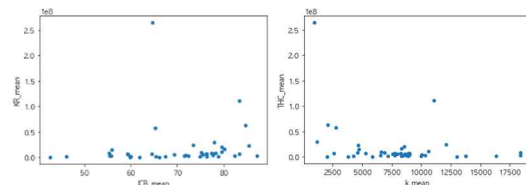
<figure 2. 국가 별 수입 금액 평균 (왼쪽부터 한국으로부터의 수입액, 전세계로부터의 수입액)>

특히 <figure 2>의 두 그래프 비교를 통해 USA, China, Japan, Mexico, India, Germany, viet Nam, Russian Federation, Singapore의 한국으로부터 수입한 평균 단가 금액이 큰 국가라는 것을 확인하였다.

분석 이후 주요 변수에 대한 가설을 설정한 후 이에 대한 분포와 상관계수 지표표를 확인하여 가설을 검증하였다.

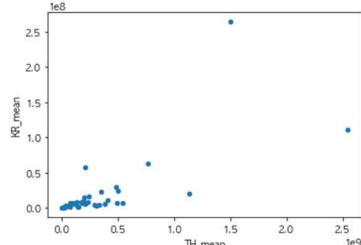
가설 1. 해당 국가 해당 품목 수입 금액의 평균이 높을수록 비즈니스 용의성 지표도 높을 것이다
 가설 2. 한국에 가까운 지역일수록 해당 국가 수입금액이 클 것이다.

가설 3. 해당 연도 해당 국가의 해당 품목 수입금액이 클수록 내년 해당 국가가 해당 품목을 한국으로부터 수입한 금액 값이 클 것이다.



<figure 4. (왼쪽부터) 가설1과 가설 2의 시각화 자료>

검증 결과, 가설 1과 가설 2의 상관계수는 각 0.011366과 -0.30406으로 매우 낮았다.

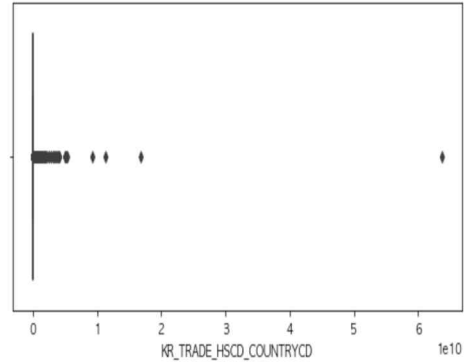


<figure 5. 가설 3의 시각화 자료>

하지만 가설 3의 상관 계수는 0.71635로, 상대적으로 높은 양의 상관을 띄었다. 해당 결과를 바탕으로 해당 국가 해당 품목 수입금액이 증가할수록 내년 해당 국가가 해당 품목을 한국으로부터 수입

한금액이 증가하는 경향을 띄을 확인할 수 있었다.

주어진 2017년 데이터의 경우, KR_TRADE_HSCD_COUNTRYCD값의 이상치가 존재하였다



< figure 6. KR_TRADE_HSCD_COUNTRYCD의 분포>

하지만 2018년 데이터에는 정답 데이터가 없으므로 KR_TRADE_HSCD_COUNTRYCD의 이상치를 제거하고 모델링하는 방법은 부적합하다고 판단하였다.

입증된 가설 3을 바탕으로 정답 값과 상관성이 높은 TRADE_HSCD_COUNTRYCD 변수의 이상치를 제거한 후 모델링 하였다.

2) 데이터 전처리

이상치 제거 외에도 데이터의 누수를 막기 위해 KR_TRADE_HSCD_COUNTRYCD의 값이 '0'인 경우를 제외하여 학습을 진행하였다.

탐색적 데이터 분석(EDA)의 결과를 통해 기존 데이터를 이용한 다양한 파생 변수와 KR_TRADE_HSCD_COUNTRYCD의 값과 유의미한 관계를 나타낼 외부 데이터를 선정하였다.

선정된 외부데이터는 <table 1.>과 같다.

<table 1. data description>

Variable	Description
gdp_diff	연간 GDP 변화율: (NY.GDP.M.KT.CD - NY.GDP.M.KT.CD_1Y) / NY.GDP.M.KT.CD_1Y
1_gdp	1인당 GDP : NY.GDP.M.KT.CD / GDP_SF_POP_TOTL
Item_depend	품목 별 무역 의존도: 해당 연도 해당 국가의 해당 품목 수입 금액 / 해당 연도 해당 국가의 전체 품목 수입 금액
Country_depend	나라 별 무역 의존도: 해당 연도 해당 국가의 해당 품목 수입 금액 / 해당 연도 해당 품목의 전체 수입 금액
TARIFF_AVG	해당 국가에서 해당 품목에 적용되는 평균 관세율
Sndist_mean	품목 별 평균 거리: 품목 별 거리 Sndist의 합 / 품목 별 국가 개수
Item_md_rv	품목 별 평균 거리 당 수입: TRADE_HSCD_COUNTRYCD / Sndist
FTA	FTA 가입국 = 1, FTA 미가입국 = 0 으로 one-hot 인코딩
OECD	OECD 가입국 = 1, OECD 미가입국 = 0 으로 one-hot 인코딩
WTO	WTO 가입국 = 1, WTO 미가입국 = 0 으로 one-hot 인코딩
ASEM	ASEM 가입국 = 1, ASEM 미가입국 = 0 으로 one-hot 인코딩
ASEAN	ASEAN 가입국 = 1, ASEAN 미가입국 = 0 으로 one-hot 인코딩
FEALAC	FEALAC 가입국 = 1, FEALAC 미가입국

	= 0 으로 one-hot 인코딩
X_FCL(diff)	X년도 FCL 컨테이너 수출실적 ■ _diff: X년도 대비 X+1년도 FCL 컨테이너 수출실적 증감율
X_LCL(diff)	X년도 LCL 컨테이너 수출실적 ■ _diff: X년도 대비 X+1년도 LCL 컨테이너 수출실적 증감율
X_EMPTY(diff)	X년도 EMPTY 컨테이너 수출실적 ■ _diff: X년도 대비 X+1년도 EMPTY 컨테이너 수출실적 증감율
X_TOTAL(diff)	X년도 총 컨테이너 수출실적 ■ _diff: X년도 대비 X+1년도 총 컨테이너 수출실적 증감율
X_hang(diff)	X년도 항만 컨테이너 물동량 ■ _diff: X-1년도 대비 X년도 항만 컨테이너 수출실적 증감율
X_lpi	X년도 물류 성과 지표 ■ 2년마다 측정되는 지표이기에 2017년의 경우 2016년도 데이터를 사용
X_fi(diff)	X년도 항공 화물 운송량 ■ _diff: X-1년도 대비 X년도 항공 화물 증감율
HSCD	HSCD의 앞 두 숫자를 추출하여 수출 품목 상위 5개를 선정 ■ 상위 5개 수출 품목: 84, 85, 90, 39, 73 상위 5개 수출 품목 = 1, 이외 품목 = 0으로 one-hot 인코딩
TARIFF_AVG_sum	TARIFF_AVG 수정한 변수 명
TRADE_HSCD_COUNTRYCD_sum	TRADE_HSCD_COUNTRYCD의 수정한 변수 명
TARIFF_AVG_mean	TARIFF_AVG의 총 평균
TRADE_HSCD_COUNTRYCD_mean	TRADE_HSCD_COUNTRYCD의 총 평균
TRADE_HSCD_COUNTRYCD_ratio	품목별 TRADE_HSCD_COUNTRYCD 비율

3) 모델 구축 및 검증

데이터 전처리 과정에서 만든 변수를 바탕으로 2017년 최종 데이터셋을 생성하고 이를 통해 최종 모델을 구축하였다. 모델의 학습을 위한 학습 데이터와 학습된 모델을 검증하기 위한 테스트 데이터로 나누었으며, 학습 데이터의 경우 이상치를 제거한 데이터를, 테스트의 경우 전체 데이터를 사용하였다. 모델링을 진행시에 교차검증을 진행하여 성능 향상을 꾀했다.

독립변수들은 Standard-scaling을 진행하여 앞선 치우친 데이터분포의 문제를 개선 한 후 모델링을 진행하였다.

종속변수의 경우 일반적으로는 스케일링 하지 않으나 절대적인 값의 차이로 인해 학습에 문제를 확인 후 로그 스케일링(Log Scaling)을 진행하였다.

주어진 공공데이터는 시계열(time series) 데이터가 아니었으며, 여러 데이터를 종합한 단적인 데이터의 형태를 보였다.

따라서 랜덤포레스트, XGBoost, LightGBM 과 같은 앙상블 모델을 데이터의 특성에 맞는 최적의 알고리즘으로 선택했다. 생각했다.

앙상블이란 어떤 데이터의 값을 예측한다고 할

때, 여러 개의 모델을 조화롭게 학습시켜 그 모델들의 예측 결과들을 이용하여 더 정확한 예측값을 구하는 것이다. 여러 개의 결정 트리(Decision Tree)를 결합하여 하나의 결정트리보다 더 좋은 성능을 낼 수 있다.

Hyper parameter tuning의 경우 grid search와 임의 탐색 등을 통해 진행하였다. 또한 실제 변수들의 제거 방법으로 A/B 테스트를 통해 RMSE에 부정적인 영향을 주지 않는 변수들을 전진선택법(Forward Selection)을 통해 선택하였다.

<table 2>는 test set에 대한 모델링 결과이다.

<table 2. 각 모델 별 cross_val_score.mean값>

모델명	Cross_val_score.mean
Random Forest	2.009
Xgboost	2.062
Lightgbm	1.975

Forward selection 기반의 변수 선택 과정을 거쳐 최종적으로 선택한 변수는 <table 3> 와 같다.

<table 3. 최종 모델링 변수>

HSCD	TRADE_COUNTRYCD	NY_GDP_MKTP_CD
NY_GDP_MKTP_CD_1Y	SP_POP_TOTL	PA_NUS_FCRF
IC_BUS_EASE_DFRN_DB	KMDIST	gdp_diff
l_gdp	item_depend	country_depend
item_md_rv	sndist_mean	TRADE_HSCD_COUNTRYCD
FTA	OECD	WTO
ASEAN	FEALAC	2017_FCL
2018_FCL	2018_Total	17EMPTY_diff
2016_lpi	TARIFF_AVG_sum	TRADE_HSCD_COUNTRYCD_sum
TARIFF_AVG_mean	TRADE_HSCD_COUNTRYCD_mean	TRADE_HSCD_COUNTRYCD_ratio

하이퍼 파라미터 튜닝 결과값은 table 4와 같다.

<table 4. 하이퍼 파라미터 튜닝 결과 값 요약>

튜닝 방법	max_features	min_samples_leaf	n_estimators	RMSE
그리드 탐색	8	5	300	0.6657
임의 탐색	4	4	438	0.6657

하이퍼 파라미터 튜닝 후 최종적으로 랜덤 포레스트를 주 모델로 선택하였다. 주어진 데이터의 연결성의 문제와 주기가 존재하지 않기 때문이다. 무엇보다 2019년을 예측하는 것이 이번 공모전의 목표이기에 일반화 성능에 주목하였다.

일반적으로 부스팅 계열의 모델이 랜덤포레스트 모델보다 성능이 높지만, 부스팅계열의 모델은 오답에 가중을 주는 방식으로 훈련하므로 학습 데이터의 기간이 짧다면 과적합의 가능성이 높다. 따라서 부스팅 계열 모델의 사용을 지양하였다.

최종 데이터셋에 적합한 모델 두 가지(랜덤포레스트, LightGBM)를 선택한 후, 하이퍼파라미터 튜닝을 통해 최종적으로 5개의 모델을 구축하였다.

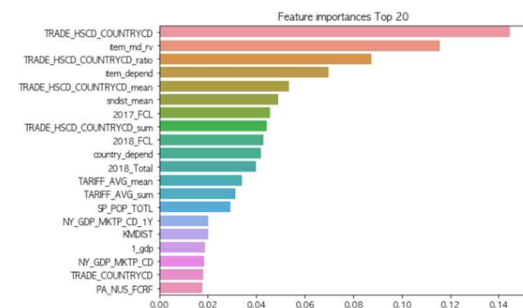
<table 5. 최종 모델 구축 및 결과 값 >

모델명	사용된 데이터 변수	하이퍼 파라미터 튜닝	RMSE
LightGBM	Input_var1	1000, 0.05, dart	1.5409
Random Forest	Input_var1	81,500	0.6921
Random Forest	Input_var1	3,2500	1.1016
Random Forest	Input_var2	4,2500	1.2275
Random Forest	Input_var3	7,4500	1.2964

데이터 변수와 모델링의 조합을 비교하여 Random Forest 모델링을 사용한 <table 5>의 두 번째 조합이 가장 낮은 RMSE값(0.6921)을 뽑을 확인하였다.

2017년 분석을 토대로 동일한 모델링 기법으로 2018년도 데이터 분석을 진행하였고, 예측된 KR_TRADE_HSCD_COUNTRYCD 값을 얻을 수 있었다.

3. 결론



<figure 7. Feature importance>

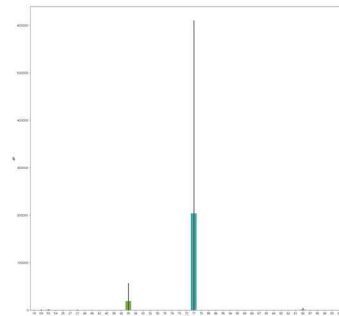
<figure 7>은 최종 모델링의 피쳐 인포턴스를 시각화한 결과이다.

모델링에 가장 큰 영향을 미친 변수 상위 다섯개는 각각 품목 별 수입금액, 품목 별 평균 거리당 수익, 품목 별 TRADE_HSCD_COUNTRYCD 비율, 품목별 무역 의존도 변수였다. 이를 통해 전반적으로 전 세계 거래량이 높은 품목일 수록, 우리나라와의 거래량도 높음을 확인하였다. 또한 FCL컨테이너 물동량, 무역 의존도 등 물류 성과 지표 및 수출 실적이 높은 국가일수록 결과값에 유의미한 영향 미침을 파악하였다.

거시 경제변수는 상기 기술한 산업, 국가 변수보다 큰 영향을 미치지 않았지만 유의한 영향이 있었다.

해당 제안과 결과를 한국의 수출 유망 품목 선정의 근거로 선정할 수 있다. 실제로 우리는 단순 한

국으로부터의 각 국가 수입액이 2018년도 대비 2019년에 큰 폭으로 증가 될 것이라고 예상되는 품목을 시각화 할 수 있었다. 그 결과로 가장 큰 폭으로 증가하는 품목코드는 711790(기타 모조신변 장식용품), 711719(모조신변장식용품 기타), 350699(조제글루(glue)와그밖의조제접착제 기타), 350691(고무 또는 플라스틱을 기제로한 접착제), 350610(글루(glue)나접착제로 사용하기에 적합한 물품), 600632(그밖의메리야스편물이나뜨개질편물 염색한 것), 600410(그밖의메리야스편물이나뜨개질편물 탄성사의 함유중량이 전 중량의 100분의 5 이상이며, 고무실은 함유하지 않은 것) 등으로 파악되었다.

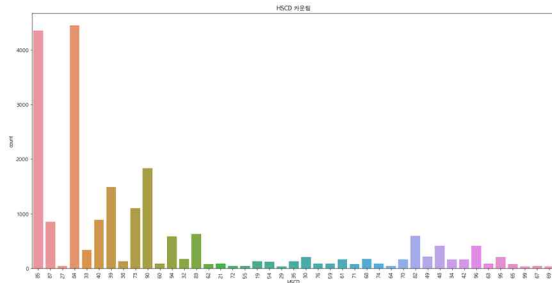


<figure 8. 가장 큰 폭으로 증가하는 품목 코드 시각화>

이러한 결과와 같이 2018년 대비 19년도 한국의 무역 동향이 개선될 것으로 예상되는 품목을 선정하고 해당 정보와, 해당 품목이 주로 수출되는 국가 정보, 해당 국가와 우리나라의 우호 정도, 정치적 안정성 등의 정성적 지표 등을 활용하여 단기적인 한국 유망 품목 선정의 지표로서 위 모델링이 활용될 수 있다.

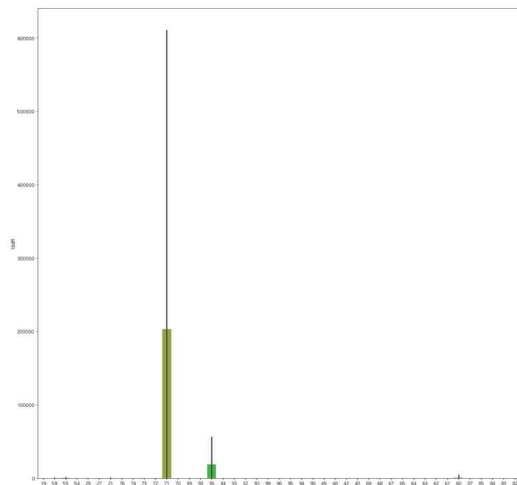
해당 제안에 사용된 모델링을 통해서 중소기업 및 신생 무역업 회사에서 단기 예측을 진행한다면, 어느 국가에서 얼마만큼의 규모와 시장경쟁력이 있을지 손쉽게 결정할 수 있다. 중소기업의 무역업 진출 의사결정 판단의 기반으로 활용되어 한국 중소기업의 경쟁우위가 될 수 있다.

또한, 국내 품목에 대한 수입액 예측과, 이에 영향을 미치는 변수를 파악하는 과정은 무역 관련 정책 혹은 전략 등의 기본 틀을 마련하는 것에 중요한 근거가 될 것이다. 정부 정책 제정의 근거를 사실적인 데이터 분석에 기초로 한 자료로 사용함으로써, 미중 무역 분쟁, 아세안 지역의 성장, 선진국의 수출 비중 하락과 같은 급변하는 세계 무역 현황에 대해서 더욱 합리적인 의사 결정이 가능할 것이다.



<figure 9. 2019년 KR_TRADE_HSCD_COUNTRYCD 값의 분포>

해당 과제를 통해서 짧은 미래에 우리나라의 산업 구조에 대한 예측이 가능할 것이다. 실제로 2019년 예측 결과에 따르면, 85(전자기기 및 그 부품), 84(원자로 보일러와 기계류 및 이들의 부분품) 90(광학기기, 사진용 기기, 영화용기기, 측정기기, 검사기기, 정밀기기 및 의료용기기 및 이들의 부분품과 부속품), 39(플라스틱 및 그 제품) 등의 순서로 산업 구조를 이룰 것이라는 것을 확인 할 수 있다.



<figure 10. 18년도 대비 19년도 국내에서 성장세를 보일 품목 군>

해당 분석 과제의 결과를 바탕으로 18년도 대비 19년도 국내에서 성장세를 보일 품목 군을 바 그래프로 표현한 결과, 71류 (천연진주·양식진주·귀석·반귀석·귀금속·귀금속을 입힌 금속과 이들의 제품, 모조 신변장식용품, 주화, 해설서. 15, 비금속(卑金屬)과 그 제품)와 35류 (제35류 단백질계 물질, 변성전분, 글루(glue), 효소, 해설서. 36, 제36류 화약류, 화공품, 성냥, 발화성 합금, 특정 가연성 조제품)의 성장 가능성이 두드러지게 나타났다. 이러한 결과를 토대로 해당 무역 산업 구조의 변화를 미리 예견하고 관세를 조정, 해당 품목 관련 국제협약 체결 등의 선제 대응 조치가 가능할 것이다.

참고 문헌

- J. M. Yoon, "Effectiveness Analysis of Credit Card Default Risk with Deep Learning Neural Network," Journal of Money & Finance, vol. 33, no. 1, pp. 151-183, Mar. 2019.
- Kaggle. UCI Credit Card Dataset [Internet]. Available: <https://www.kaggle.com/uciml/default-of-credit-card-client-s-dataset>.
- A. Shen, R. Tong, and Y. Deng, "Application of Classification Models on Credit Card Fraud Detection," in 2007 International Conference on Service Systems and Service Management, pp. 1-4, Jul. 2007.
- B. M. Pavlyshenko, "Machine-Learning Models for Sales Time Series Forecasting," Data, vol. 4, no. 1, Apr. 2019.
- B. Scholkopf, C. J. C. Burges, A. J. Smola, 'Advances in kernel methods', The MIT Press 1999
- C.L. Wilson and M.D.Garris, Handprinted character database 3, February 1992, <http://www.nist.gov/srd/nistsd19.htm>, National Institute for Standards and Technology, Advanced Systems Division pg.43
- D.M.J. Tax, 'One-class classification', PhD Thesis, Delft University of Technology, <http://www.ph.tn.tudelft.nl/~davidt/thesis.pdf> ISBN: 90-75691-05-x, 2001
- D.M.J. Tax, R.P.W. Duin, 'Support Vector Data Description', Pattern Recognition Letters, December 1999, vol. 20(11-13), pg. 1191-1199
- Dacon. Korea data competition platform [Internet]. Available: <https://dacon.io/>. [13] R. J. A. Little and D. B. Rubin, Statistical Analysis with Missing Data, 2nd ed. Hambrug, NJ: John Wiley & Sons Inc., 2014.
- Dacon. Korea data competition platform. Card Sales Prediction contest [Internet]. Available: <https://dacon.io/competitions/official/140472/overview/>.
- Documents for Catboost [Internet]. Available: <https://catboost.ai/>.
- Documents for Lightgbm [Internet]. Available: <https://lightgbm.readthedocs.io/en/latest/index.html>.
- Documents for Xgboost [Internet]. Available: <https://xgboost.readthedocs.io/en/latest/#>.

ETSI, Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Iterative Service, News, Gathering the other Broadband Satellite Applications (DVB-S2), EN 302 307 v1.2.1, 2009.

G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach: CA, pp. 3149-3157, 2017.

H. Kim, "The Prediction of PM2.5 in Seoul through XGBoost Ensemble," Journal of the Korean Data Analysis Society, vol. 22, no. 4, pp. 1661-1671, Aug. 2020.

J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," The Annals of Statistics, 2nd ed. Cambridge, MA: The MIT Press., vol. 29, no. 5,

pp.1189-1194, 2001.

J. Lee and D. Yoon, "Soft-decision demapping algorithm with low computational complexity for coded 4 + 12 APSK," *Int. J. Satellite Commun. Networking*, vol. 31, no. 3, pp. 103-109, May/June 2013.
<https://doi.org/10.1002/sat.1022>

J. Lee, Y. Jang, and D. Yoon, "Approximated soft-decision demapping algorithm for coded 4+ 12+ 16 APSK," *J. Korean Inst. Commun. Inform. Sci. (KICS)*, vol. 37, no. 9, pp. 738-745, Sep. 2012.
<https://doi.org/10.7840/kics.2012.37A.9.738>

K. Cho, K. Hyun, S. Park, and S. Cho, "An approximated soft decoding algorithm of 16-APSK signal for DVB-S2," in *Proc. IEEE Int. Conf. Consumer Electron. 2007 (IEEE ICCE 2007)*, pp. 1-2, Las Vegas, U.S.A., Jan. 2007.

Kaggle. Rossmann Store Sales [Internet]. Available: <https://www.kaggle.com/c/rossmann-store-sales>. [6]
J. H. Lee, "Stock price prediction model using deep learning," M. S. Thesis, Soongsil University, Seoul, 2016.

L. Prokhorenkova, G. Gusev, A. Vorobev, V. A. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical feature," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montreal, Canada, pp. 6639-6649, 2018.

M. Eroz, F. Sun, and L. Lee, "DVB-S2 Low Density Parity Check codes with near Shannon limit performance," *Int. J. Satellite Commun. Networking*, vol. 22, no. 3, pp. 269-279, May/June 2004.
<https://doi.org/10.1002/sat.787>

M. Zhang and S. Kim, "A simple soft decision detection for APSK schemes for satellite systems," in *Proc. Korean Inst. Commun. Inform. Sci. (KICS) Summer Conf. 2011*, pp. 249-250, Jeju, Korea, June 2011.

M. Zhang and S. Kim, "Efficient bit-interleaved APSK scheme for LDPC codes," in *Proc. 6th ASMS and 12th SPSC Workshop 2012*, pp. 49-53, Baiona, Spain, Sep. 2012.

M. Zhang and S. Kim, "Efficient soft demapping for M-ary APSK," in *Proc. Int. Conf. ICT Conv. 2011 (ICTC 2011)*, pp. 641-644, Seoul, Korea, Sep. 2011.

M. Zhang, S. Kim, and Y. Kim, "Universal soft decision demodulator for M-ary adaptive modulation systems," in *Proc. IEEE Int. Conf. Asia-Pacific Conf. Commun. 2012 (IEEE APCC 2012)*, pp. 574-578, Jeju, Korea, Oct. 2012.

Metz, C. 'Basic principles of ROC analysis' *Seminars in Nuclear Medicine*, VIII(4) 1978

O. Afelumo, A. Awoseyila, and B. Evans, "Simplified evaluation of APSK error performance," *Electron. Lett.*, vol. 48, no. 14, pp. 886-888, July 2012.
<https://doi.org/10.1049/el.2012.1740>

O. Afelumo, A. Awoseyila, and B. Evans, "Simplified expressions for APSK error performance," in *Proc. AIAA Int. Commun. Satellite Sys. Conf. 2012 (ICSSC 2012)*, pp. 1-11, Ottawa, Canada, Sep. 2012.

S. B. Jha, R. F. Babiceanu, V. Pandey, and R. K. Jha, "Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study," *arXiv*:

2006.10092v1, Jun. 2020.

S. R. Lee, "Comparison of algorithms for the missing data imputation methods," M. S. Thesis, Hankuk University of Foreign Studies, Seoul, 2020.

T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco: CA, pp. 785-794, 2016.

V. N. Vapnik, 'The nature of statistical learning theory', 1995 Springer-Verlog, New York, Inc.

Y. G. Lee, J. Y. Oh, and G. B. Kim, "Interpretation of Load Forecasting Using Explainable Artificial Intelligence Techniques," *The Transactions of the Korean Institute of Electrical Engineers*, vol. 69, no. 3, pp. 480-485, Feb. 2020. [10] S. I. Jang and K. C. Kwak, "Comparison of Safety Driver Prediction Performance with XGBoost and LightGBM," in *Proceeding of Korea Institute of Information Technology Conference*, pp. 360-362, Jun. 2019.

Yonsei Structure & Bridge Eng Lab. Interpolation [Internet]. Available:
<http://str.yonsei.ac.kr/korean/portal.php>.

염정우, 2020.12.18, 수출기업 살펴보니... 중소기업만 무역수지 적자 기록, 조세일보
(<http://www.joseilbo.com/news/htmls/2020/12/20201218412941.html>)

외교부 홈페이지, <https://www.mofa.go.kr/www/index.do>
임은진, 2020-04-08, WTO "코로나19로 올해 세계 무역 32% 감소할 수도"
(<https://www.yna.co.kr/view/AKR20200408189300088>)

홍성욱 외 3명(2017), 국가별 경제권별 수출결정요인 분석을 통한 수출전망모형 구축, 산업연구원
환율 데이터, <https://kr.investing.com/currencies/usd-krw-historical-data>

컨테이너별 수출 실적, <https://unipass.customs.go.kr/>

남윤미 외 1(2020), 우리나라 제조업 기업의 수출중단 요인 분석