

# 빅데이터 분석을 통한 과실류 가격예측 시스템 연구

임우재\*, 유정원\*, 김대호\*\*, 박상범\*\*, 서용준\*\*, 정종화\*\*

\*상명대학교 컴퓨터학과

\*\*홍익대학교 컴퓨터공학과

## A Study on Fruit Price Prediction System by Big Data Analysis

Woo-Jae Lim\*, Jung-Won Yu\*, Dae-Ho Kim\*\*, Sang-Beom Park\*\*, Yong-Jun Seo\*\*, Jong-Hwa Jeong\*\*

\*Dept of Computer Science, Sang-Myung University

\*\*Dept of Computer Engineering, Hong-Ik University

### 요 약

본 연구는 국내 각 과일별 재배 면적 및 해당 지역의 날씨 등의 데이터를 이용하여 실 생산량을 예측하고 각 과일별 경합관계를 적용하여 실제 수입 되는 과일의 수입량과의 상관관계를 분석하여 실제 가격을 예측하고자 한다. 본 결과를 통해 다변화 되는 국내 과일 시장에서 농업, 수입업에 대한 방향성 제시를 그 목적으로 둔다.

### 1. 다변화적인 국내 과일시장

국내에서 과일에 대한 관심은 해마다 증가하고 있다. 가족 단위의 소비에서 1인 가구 단위 소비로 변화하면서 가벼운 식사대용으로 과일을 찾거나 이제는 트렌드가 되어버린 웰빙에 대한 관심이 그 증가속도에 박차를 가하고 있는 것이다. 또한 소비량뿐 아니라 소비자가 선택할 수 있는 과일의 품목 또한 과거와 달리 매우 선택 폭이 넓어졌는데, FTA 발효로 인해 국내 과일 시장이 개방됨에 따라 해마다 다양한 해외 과일이 자유롭게 유통되고 있기 때문이다. 이러한 이유로 과일 소비 패턴이 더욱 다양해지고 각 과일이 다른 대체 과일로 바뀌기도 하면서 국내 과일 시장은 유래 없는 변화를 맞고 있다.

과일은 공산품과 달리 부패할 수 있는 생물(生物)이다. 때문에 그 보관 기간이 짧아 별도의 가공 처리를 거치지 않고서는 빠른 시간 내에 소비 되어야 한다. 실제로 수입 과일류는 세관에 들어온 후 일주일 이내에 소매상에게 유통되어야 하며, 국내 과일의 경우 수확 2주일 이내에 소비되어야 하는 특징을 갖는다. 이러한 특징 때문에 국내에서 유통되고 있는 과일의 가격은 매우 급하게 변화한다.

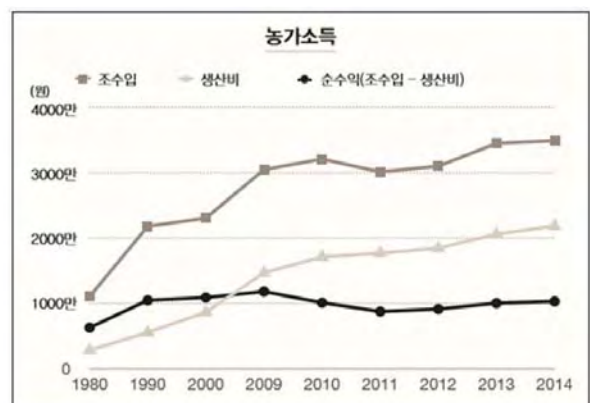
예를 들어 한 과일 품목에 대해 수입량이 급증 했을 경우 소비 시장에서 그 과일의 가격이 안정화 될 때까지 기다릴 수가 없고 손해를 감수하고라도 시장에 넘겨야 하는 것이다. 국산 과일 농가도 마찬가지로 한 과일에 대해 국내 생산량이 급증 했을 경우 실 생산비마저 감내해야할 상황에 놓이기도 한다.

이와 같이 국내 과일 시장의 가격은 여러 조건이 개입되어 형성 되는데, 본 연구에서는 기존 과일 시장에 수입과일의 양이 크게 증가함에 따라 변하게 되는 가격 추이를

빅 데이터 적으로 분석 해보고자 한다.

### 2. 국내 실 사정에 맞는 지표의 필요성

1990년대 바나나, 파인애플에 이어 오렌지에 대한 수입이 개방되었고, 2000년대에는 포도, 키위 등이 추가 되었으며 최근에는 체리, 석류, 망고, 블루베리 등의 수입이 크게 증가하고 있다. 국내 과일 시장에서 수입 과일의 점유율은 2000년대 초 10% 수준에서 2014년에는 20% 수준으로 뛰어 올랐으며 최근에는 약 30% 수준을 웃돌고 있다. 실제로 KBS News 기사(박지은, 2017)를 보면 현재 수입 과일의 과일 시장 점유율은 약 30%에 육박 했으며, 국내 과일 시장을 빠르게 잠식 하고 있다고 하면서 2012년 1조 5억 원이던 과일 수입 규모가 2016년에는 1조 2천7백억 원으로 크게 증가 했다고 말하고 있다. 이 같은 추세에서 국내 과일 농가는 기존의 시장에서의 위치를 유지하기 위해 재배하던 과일 대신 외국 품종을 국내에서 재배하는 등의 다양한 변화를 시도하고 있으나 조수입에서 생산비를 제외한 순수익 측면은 매년 큰 차이가 없었다.



출처 - 농림축산식품부, 통계청

< 그림 1 > 농가 소득

또한 매해 재배할 과일을 선택함에 있어 참고할 마땅한 지표가 존재하지 않았다. 더해서 매년 다양해지는 수입과일의 품목으로 각 과일품목간의 경합관계가 복잡해지면서 그 문제가 더욱 가중 되고 있다.

이와 같은 상황에서 실 국내 재배면적과 기후 등을 반영한 과일의 생산량과 때에 맞춰 들어오는 외국 과일의 수입량에 의한 가격변화율을 빅 데이터 적으로 분석하고 예측해보는 것은 중요한 의미를 갖는 지표가 될 것이다.

### 3. 본 연구의 목적

본 연구는 국내 재배면적과 기후에 따른 생산량을 예측하고 그 시기에 들어오는 수입 과일의 양에 따라 국내 유통되는 과일의 가격이 어떻게 변화 했는지를 빅 데이터 분석 기법을 이용하여 해당 변화에 대해 농업과 수입업 등에서 대응할 수 있는 시사점을 제시하는 데에 있다. 그 세부적인 연구 목적으로는 국내 생산량 파악을 위한 지표로 국내 기상 정보와 재배면적을 이용한 1. 국내 재배면적과 기온에 의해 변화하는 실 생산량 예측, 다양화된 과일 품목간의 가중치를 이용한 2. 각 과일 품목간의 경합관계를 분석, 때마다 맞춰 들어오는 실 수입과일의 양을 통한 3. 국내 실 생산량과 수입 과일의 양이 가격 변화에 미치는 영향 관계 분석, 그리고 4. 분석 결과를 바탕으로 국내 농업 및 수입업에 대한 방향 제시를 하는 데에 있다.

①	국내 재배면적과 기온에 의해 변화하는 실 생산량 예측
②	각 과일 품목간의 경합관계 분석
③	국내 실 생산량과 수입과일의 양이 가격 변화에 미치는 영향 관계 분석
④	분석 결과를 바탕으로 한 국내 농업 및 수입업에 대한 방향제시

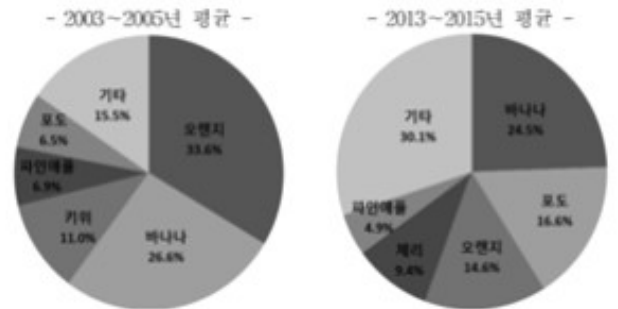
< 표 1 > 연구 목표

### 4. 관련 선행 연구 분석과 그 차별성

수입 과일에 대한 선행 연구는 대개 소비자의 성향을 예측하거나 기존 과일에 대한 대체 과일의 영향을 연구 하였는데, 노수정 외(2012)는 로테르담 수요 모형을 이용하여 수입 오렌지의 양으로 국내의 과일 수요 탄력성을 측정 하였고 조덕래·조재환 (1992)과 조덕래(1995)의 경우 연도별 자료를 이용하여 단일방정식 형태의 과일 수요 함수를 각각 추정하였다. 강지용·고성보(1998) 역시 연도별 자료를 이용하여 재배면적 함수와 단수 함수 등을 추정하여 오렌지 수입자유화 시나리오에 따른 국내 감귤 산업 피해를 계측 하고자 하였다.

위의 선행 연구들에서는 국내산 과일 품목 간 대체 관계

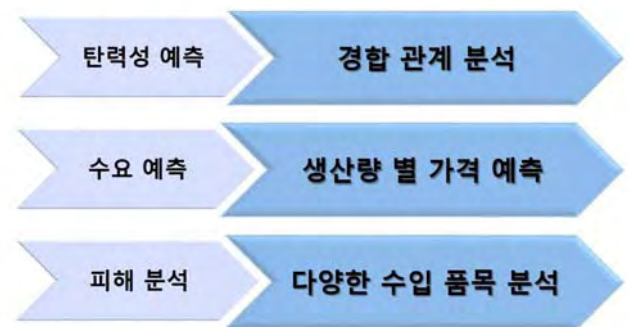
를 중심으로 수행 되어 수입 과일과의 관계에 대해서는 불분명하다. 또한 국내 과일에 대한 수입 과일의 영향을 분석한 연구는 주로 오렌지에 집중되었다. 연구 당시 오렌지가 사회적 관심 대상임과 동시에 주 소비 과일로서의 역할을 했으나, 연합뉴스 권순완(2017) 기사를 보면 바나나 매출액은 2015년 577억 원, 지난해 633억 원을 달성하면서 올해는 700억 원을 넘어서며 과일 소비에서 가장 높은 순위를 차지하고 있음을 알 수 있다. 이처럼 과일의 종류가 다양화 되고 소비 패턴에 변화가 생겼음에도 그 이후에 연구는 거의 수행되지 않았다.



출처 - 한국 무역 협회

< 그림 2 > 수입 과일의 수입 비율

최근 소비가 급증하는 체리, 블루베리, 망고, 파인에플 등의 수입 과일에 대해서도 연구가 진행되어야 한다. 본 연구의 차별성은 1. 과일 수입의 증가로 다변화된 과일 시장에서 분석 대상 품목을 다양화 하며, 2. 다양한 과일에 대한 분석을 국산 과일의 생산량에 대조해 그 가격 추이를 예측해 보고자 함에 있다.



< 그림 3 > 연구 분석 목표

### 5. 연구의 방향

실 연구범위는 정부에서 공공데이터 개방으로 제공하는 년도를 기준으로 전망 시기는 1년 단위로 하며, 국내 생산 과일 품목으로는 사과, 감귤, 포도, 배, 단감, 복숭아, 수박, 참외로 하며 수입 과일 품목으로는 바나나, 오렌지, 수입포도, 체리, 키위, 파인에플로 한다. 또한 각 시기, 계절 별로 수입량이 증가하는 상 하위 20%의 특정

과일을 제외하여 추후 추가시킴으로 한다.

단위 : 천ha

<b>국내 과일</b>	사과, 포도, 배, 감귤, 복숭아, 수박, 참외, 감, 자두, 멜론, 딸기, 토마토
<b>수입 과일</b>	오렌지, 바나나, 포도, 석류, 레몬, 망고, 파인애플, 자몽, 아보카도, 블루베리, 키위, 토마토

< 표2 > 연구 선정 과일 품목 목록

먼저 국내 과일별 실 재배 면적과 재배 지역의 연도별 기상 자료를 이용하여 과일별 국내 생산량 추이를 모형 계측한다. 모형 추정에 사용되는 자료로는 공공데이터로 주어지는 통계청과 기상청의 월별, 년별 데이터를 이용한다. 그 후 수입 과일별 수입량 추이를 이용하여 실제 과일 도매단의 가격을 예측 모델을 구축한다. 이때 사용되는 자료로는 관세청에서 신고된 수입 과일에 대한 데이터를 이용하며 도매단의 가격은 각 과일 경매장의 일일 과일별 경매 데이터를 축적해 지표로 사용한다.

연도	사과	배	복숭아	포도	감귤	감	자두	합계
2000	29.1	26.2	13.9	29.1	26.8	31.2	4.7	161.1
2005	26.9	21.7	15.0	22.1	21.5	26.8	6.7	140.7
2010	31.0	16.2	13.9	17.6	21.1	31.8	5.9	137.5
2011	31.2	15.1	13.8	17.4	21.4	31.3	5.7	136.0
2012	30.7	14.4	14.3	17.2	21.4	30.3	5.7	133.9
2013	30.4	13.7	14.6	16.9	21.3	29.2	5.7	131.9
2014	30.7	13.1	15.5	16.3	21.3	28.0	5.7	130.7
2015	31.6	12.7	16.7	15.4	21.3	27.0	5.9	130.6

출처 - 국가통계포털(KOSIS)

< 표3 > 과일 별 재배 면적 추이

단위 : 천ha

구분	세부내용
<b>이론적 연구</b>	가) 수집 대상 - 정부 기관의 각종 통계 자료 나) 수집 방법 - Open API를 이용한 JSON 데이터 사용 - 웹크롤링 기술을 이용한 데이터 추출
<b>실증적 연구</b>	가) 통계 분석 (R 프로그램) - 과일 품목 사이의 상관관계 가중치 분석 - 날씨에 따른 과일 생산량 분석 나) 예측 분석 (머신 러닝) - 선형 모델과 TensorFlow를 활용하여 과일별 생산량 및 가격 예측

< 표3 > 연구 방향

시점	수박	참외	딸기	토마토	멜론	합계
2000	30.5	10.2	7.1	4.9	0.7	53.3
2005	23.2	7.1	7.0	6.7	1.2	45.2
2010	16.4	6.2	7.0	5.3	1.7	36.6
2011	15.7	5.9	5.8	5.9	1.4	34.7
2012	15.2	5.8	6.4	6.3	1.5	35.3
2013	14.9	5.5	6.9	6.1	1.5	34.8
2014	16.9	5.5	6.9	7.1	1.5	37.8
2015	15.2	5.4	6.4	7.0	1.6	35.6

출처 - 국가통계포털(KOSIS)

< 표 4 > 과채 별 재배 면적 추이

위에서 언급 한 국내 주요 과일류의 개별 생산량 추이 파악을 위해 과거의 과일 별 재배지와 재배면적 데이터 그리고 기온, 풍속, 강수량, 일조량의 축적된 날씨 데이터를 사용한다. 데이터의 주요 출처는 국내 공공기관의 오픈 데이터 셋을 이용하며 2000년 이후부터 제공되고 있는 가장 최근 데이터까지 포함한다. 국내 과일 재배면적은 1년 단위로 측정이 되어 제공되므로 한번 측정된 이후 해당 년도에는 변수가 아닌 상수 값으로 치부 하여 사용할 수 있으며 해당 재배지역의 현재 1년 동안의 일별 날씨데이터와 누적된 그 지역의 기상 데이터를 모 데이터로 삼아 Training Data를 구축하고 Supervised Learning을 통해 그해 생산량을 유추하여 재배면적과 지역별 기상 변화로 생산량을 미루어 예측하는 모델을 설계 하고자 한다.

또한 과일의 경우 기상 상황 등에 의해 공급량 변화 추이가 크게 작용하여 이상적인 수요 조건 성립이 어려울 것으로 보고 시장에 공급되는 시점에서의 물량과 그해의 수입과일양에 따라 가격이 형성 된다고 판단하고 국내 각 도매시장의 일별(Daily) 도매가격을 과일 별로 축적하여 수요 수입 과일류의 수입량이 국내 과일의 생산량 대비 가격형성에 미치는 과일 별 경합 관계를 준이상역수요 체계(IAIDS)를 이용하여 파악하고자 한다.

나아가 평균 가격 예측을 하기 위하여 전국 도매시장의 도매가격을 일별(Daily)로 파악하여 해당 시기의 평균치를 계산하고 과일 별 경합 관계를 적용하여 산출된 가중치를 곱한 국내 과일 별 생산량과 수입 과일 별 수입량을 Training Set으로 구성하고 누적된 평균 도매가격을 통해 그해 형성될 과일 별 평균 가격을 예측 해보고자 한다.

위 과정에서 실 시장지표로 활용하기 위한 신뢰성 확보

를 위해 중간마다 도출되는 중간 산출물 데이터를 실제 수치화 비교하고 수입업과 과일 경매에 종사하는 담당자와 면접 조사 및 자문을 진행하고자 한다.

## 6. 연구의 기대효과

먼저 국내 과일 생산 농가의 재배 작물에 대한 지표를 제공 해줌으로써 더욱 경쟁력 있는 과실의 재배를 가능하게 해줄 수 있다. 국내에서 생산되는 과일과 수입 되는 과일 간의 경합 관계를 파악하여 실제 유통 되고 있는 과일과 경합 관계 안에 있는 과일의 생산량추이를 예측하여 작물을 선택 재배 할 수 있을 것이다.

또한 형성 되는 가격 추이와 예측되는 국내 생산량으로 경합 관계에 있는 수입 과일의 수입량을 예측하는데 지표로 삼을 수치정보를 얻을 수 있다.

실제로 국내에 들어오는 수입 과일의 경우 국제 물류 이송 체계를 따르기 때문에 국내 과일시장에 유통되기까지는 최소 2주간의 시간이 필요하다. 다시 말해 현재 예측이 되는 시점에서가 아닌 예상되는 수요량으로 수입량을 결정하는 것인데, 이에 있어 참조 할 수 있는 자료는 중요한 지표가 될 것이다.

결과적으로 국내 과일의 공급량을 보다 일정하게 유지하며 그 가격의 변동추이를 좁혀 국내 과일 시장에서 각 과일별 합리적인 가격을 구성하고 안정을 꾀하는 방향을 제시하는데 기여 할 수 있을 것으로 기대한다.

## 참고문헌

이용선 외, 『과일 수입 확대에 따른 소비행태 변화 분석』, 연구보고서 R716, 한국농촌경제연구원, 2014.

정준호, 『시장개방 확대에 따른 과일 수요 변화에 관한 실증적 연구』, 동국대학교 박사학위논문, 2017.

조재환, 『오렌지 수입이 국내산 과일 및 과채 가격에 미친 영향』, 한국유기농업학회지 Vol.23 No.4 pp.703-713, 2015

노수정, 『로테르담모형을 이용한 한국의 주요 과일류 수요분석』, 부산대학교 석사학위논문, 2012.

강지용, 고성보, 『경기침체와 오렌지 수입자유화에 따른 감귤의 수입전망 분석』, 한국농업경제학회 Vol.39 No.2 pp.143-170, 1998.

국가 통계 포털 (Kosis), 과일별 재배면적 추이, 과채별 재배면적추이,  
[http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT\\_1ET0292&vw\\_cd=MT\\_ZTITLE&list\\_id=F1H&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=E1](http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1ET0292&vw_cd=MT_ZTITLE&list_id=F1H&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=E1)

KBS(2017.08.14.), “수입 과일 30% 육박...국내 과일도 변신”

조선일보(2017.07.07.), “바나나, 사과 누르고 과일 매출 1위 등극”

[본 논문은 2017년 한이음 ICT멘토링 프로젝트의 결과물입니다]