



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

빅데이터 분석을 활용한 온라인 판매
수요 예측

Prediction of Online Sales Demand Using
Big Data Analysis

2019년 8월

한신대학교 대학원

정보통신학과

오 지 연

빅데이터 분석을 활용한 온라인 판매
수요 예측

Prediction of Online Sales Demand Using
Big Data Analysis

지도교수 홍 성 찬

이 논문을 석사학위 논문으로 제출함

2019년 8월

한신대학교 대학원

정보통신학과

오 지 연

오지연의 석사학위 논문을 인준함

주심 여 협 구 인

부심 손 승 일 인

부심 홍 성 찬 인

한신대학교 대학원

2019년 6월

목 차

그림 목차	iii
표 목차	v
국문 요약	vi
I. 서 론	1
1.1 연구의 배경 및 필요성	1
1.2 연구의 방법	3
II. 빅데이터 및 온라인 판매 이론적 고찰	10
2.1 빅데이터의 개요	10
2.2 빅데이터 기술과 분석 기법	13
2.2.1 빅데이터 기술	13
2.2.2 빅데이터 분석 기법	14
2.2.3 빅데이터 시각화	15
2.3 빅데이터 분석 활용 사례	18
2.3.1 정치, 사회, 문화 활용 사례	18
2.3.2 경영, 마케팅 활용 사례	19
2.4 온라인 판매 기업의 빅데이터 분석 활용 사례	20
2.4.1 국외 패션 기업 활용 사례	20
2.4.2 국내 패션 기업 활용 사례	22

Ⅲ. 온라인 판매 시스템 설계 및 구현	26
3.1 연구 구조의 이론적 설명	26
3.1.1 시계열 예측 방법에 관한 이론적 고찰	26
3.1.2 A기업의 온라인 쇼핑몰 데이터	28
3.1.3 기상청 기후 데이터	33
3.2 판매 수요 예측 시스템 구성	34
Ⅳ. 온라인 쇼핑몰 시스템 결과 및 분석	36
4.1 효과 및 검증	36
4.1.1 평균 온도 변화에 따른 판매량 예측	36
4.1.2 평균 가격 변화에 따른 판매량 예측	45
Ⅴ. 결론	54
참고문헌	55
Abstract	58

그 립 목 차

[그림 1.1] 2017년 8월 한 달 TOP10 상품	8
[그림 1.2] 2017년 12월 한 달 TOP10 상품	8
[그림 2.1] R 프로그램으로 그린 산점도	17
[그림 2.2] R 프로그램으로 그린 꺾은선 그래프	17
[그림 3.1] 실시간으로 수집된 데이터	28
[그림 3.2] 실시간으로 수집된 상의 판매량 그래프	29
[그림 3.3] 2014년 ~ 2018년 기온 그래프	33
[그림 3.4] BAPP 구성도	35
[그림 4.1] 2014년도 온도 변화에 따른 반팔 티셔츠 판매량	36
[그림 4.2] 2015년도 온도변화에 따른 반팔 티셔츠 판매량	37
[그림 4.3] 2016년도 온도변화에 따른 반팔 티셔츠 판매량	37
[그림 4.4] 2017년도 온도변화에 따른 반팔 티셔츠 판매량	38
[그림 4.5] 2014년 ~ 2017년도 데이터로 예측한 2018년 온도 변화에 따른 반팔 티셔츠 판매량	39
[그림 4.6] 실제로 판매된 2018년도 온도 변화에 따른 반팔 판매량	39
[그림 4.7] 2014년도 온도 변화에 따른 아우터 웨어 판매량과	41
[그림 4.8] 2015년도 온도 변화에 따른 아우터 웨어 판매량	41
[그림 4.9] 2016년도 온도 변화에 따른 아우터 웨어 판매량	42
[그림 4.10] 2017년도 온도 변화에 따른 아우터 웨어 판매량	42
[그림 4.11] 2014년 ~ 2017년도 데이터로 예측한 2018년 온도변화에 따른 아우터 웨어 판매량	43
[그림 4.12] 실제로 판매된 2018년도 온도 변화에 따른 아우터 웨어 판매량	44
[그림 4.13] 2014년도 가격 변화에 따른 반팔 판매량	45
[그림 4.14] 2015년도 가격 변화에 따른 반팔 판매량	46
[그림 4.15] 2016년도 가격 변화에 따른 반팔 판매량	46

[그림 4.16] 2017년도 가격 변화에 따른 반팔 판매량	47
[그림 4.17] 2014년 ~ 2017년도 데이터로 예측한 2018년 가격 변화에 따른 반팔 판매량	48
[그림 4.18] 실제로 판매된 2018년도 가격 변화에 따른 반팔 판매량	48
[그림 4.19] 2014년도 가격 변화에 따른 아우터 웨어 판매량	49
[그림 4.20] 2015년도 가격 변화에 따른 아우터 웨어 판매량	50
[그림 4.21] 2016년도 가격 변화에 따른 아우터 웨어 판매량	50
[그림 4.22] 2017년도 가격 변화에 따른 아우터 웨어 판매량	50
[그림 4.23] 2014년 ~ 2017년도 데이터로 예측한 2018년 가격 변화에 따른 아우터 웨어 판매량	52
[그림 4.24] 실제로 판매된 2018년도 가격 변화에 따른 아우터 웨어 판매량	52

표 목 차

[표 1.1] R 프로그램에 사용한 CSV 파일	4
[표 1.2] 2017년 8월 한 달 가격대별 판매량	6
[표 2.1] 빅데이터 환경의 특징	12
[표 3.1] 실시간으로 수집된 상의 판매량 데이터	30
[표 3.2] 2018년 온도에 따른 판매 데이터	31
[표 3.3] 2017년 온도에 따른 판매 데이터	32
[표 3.4] 기상청 기후 데이터	34

국 문 요 약

빅데이터 분석을 활용한 온라인 판매 수요 예측

한신대학교 일반대학원

정 보 통 신 학 과

오 지 연

지도교수 : 홍 성 찬

4차 산업혁명 시대에 발맞추어 기업들은 과거 어느 때 보다 다양하고 많은 양의 데이터를 확보하고 이용할 수 있는 비즈니스 환경에 놓여있다. 다양한 비즈니스 환경에서 제각각 실시간으로 해야될 수 없을 정도의 데이터 축적이 이루어지고 있으며, 이를 빅데이터(Big Data)라고 부른다. 최근 데이터 추출 기술과 분석 프로그램의 발전에 따라 빅데이터를 전략적 의사결정에 활용하려는 기업이 점차 증가하고 있다. 빅데이터 분석은 재고 관리, 수요 예측, 생산, 가격 결정, 시장 흐름 파악, 소비자 인식 및 트렌드 분석 등에 있어서 기존의 전형적인 데이터 수집 방식으로는 얻을 수 없던 새로운 정보를 창출할 수 있다. 본 논문에서 취급하는 패션 업계에서도 제품 기획, 고객 관리, 판매 등에서 웹(Web) 기반 빅데이터 분석 및 추출을 활용하려는 기업들의 시도가 늘어나고 있다. 그러나 이와 관련된 연구는 지금까지 소수에 불과하다.

이러한 점에 착안하여 본 연구에서는 지난 5년간 온라인 쇼핑몰 ‘A’ 업체에서 실시간으로 누적된 데이터를 바탕으로 상품별 판매량을 분석하고 각 변수의 변화에 따른 판매량을 예측하여 판매량 증대와 효율적인 재고 관리

를 위한 빅데이터를 활용한 온라인 판매 수요 예측 모델 BAPP(Big data Analysis for sales Prediction Platform)를 제안한다. 특히 본 연구에서는 R 프로그램을 활용하여 2014년부터 2017년도까지의 판매량을 바탕으로 2018년도 판매량을 예측하였고 예측한 판매량과 실제 판매량을 비교 분석하였다. 비교 분석한 결과 반팔 티셔츠는 더울 때뿐만 아니라 평소에도 인 웨어(In Wear)로 많은 사람이 활용하기 때문에 꾸준한 판매량을 보이며 예측과 실제 판매량이 $\pm 1.5\%$ 의 오차만 발생했다. 아우터 웨어(Outer Ware)의 경우에는 예측과 실제 판매량에서 $\pm 8\%$ 의 오차가 발생하였다. 그 원인으로서는 2016년 10월 29일부터 2017년 3월 10일까지 촛불 집회가 진행되어 아우터 웨어 판매량이 순간적으로 급증했다. 한편 수요 예측을 진행할 때 이러한 사회적 이슈를 배제하고 진행하여야 하지만 R 프로그램은 그러지 못하였고 전년도 겨울의 높은 판매량이 그대로 반영되어 2018년도에도 아우터 웨어가 많이 판매될 것으로 예측하였다. 하지만 실제 판매량은 예측 판매량보다 저조하였다. 그러나 전년도 판매량만 반영하여 예측을 진행한 것이 아닌 4년 치의 판매량을 바탕으로 예측을 진행하였기 때문에 쇼핑몰에 큰 타격을 주는 오차는 아니었다. 또한 소비자들의 선호하는 적정 가격대를 분석하여 그 가격대에 맞춰 판매를 진행한 결과 반팔 티셔츠는 $\pm 4.5\%$ 의 오차 아우터 웨어는 $\pm 1.5\%$ 의 오차가 발생 하였다. 오차가 매우 작은 편이었기에 성공적인 수요 예측을 했다고 판단할 수 있다.

이러한 결과를 볼 때 본 연구는 상품을 판매하는 패션 업계, 요식 업계 등의 다양한 기업들과 해외 진출 기업에 긍정적인 도움을 줄 수 있을 것이라 사료된다.

키워드 : 빅데이터, 수요 예측, R 프로그램, 온라인 쇼핑몰, 패션, 의류

I. 서론

1.1 연구의 배경 및 필요성

4차 산업혁명 시대에 발맞추어 기업들은 과거 어느 때 보다 다양하고 많은 양의 데이터를 확보하고 이용할 수 있는 비즈니스 환경에 놓여있다. 검색 포털, 온라인 쇼핑몰(Online shopping Mall), 소셜 네트워크 서비스(Social Network Service) 등 다양한 비즈니스 환경에서 제각각 실시간으로 소비패턴 및 상품평가와 관련된 다양하고 많은 양의 데이터 축적과 분석이 중요하게 다루어지고 있다. 우리는 이를 빅데이터라 부르며 이는 4차 산업혁명 시대의 비즈니스 모델의 핵심 포인트이며 이를 활용하지 못하는 기업은 도태된다. 시대의 흐름에 맞춰 빅데이터를 활용하려는 기업들이 늘어나고 있으며 분석 프로그램과 데이터 추출 기술들이 발전되고 있다. 기업에서는 빅데이터 분석을 통해 재고 관리, 수요 예측, 생산, 가격 결정, 시장 흐름 파악, 소비자 인식 및 트렌드 분석 등에 있어서 기존의 전형적인 데이터 수집 방식으로는 얻을 수 없던 새로운 정보를 창출할 수 있다[21]. 특히 고객의 반응에 민감한 패션 업계에서 빅데이터를 활용한 분석이 유용할 것으로 생각되지만 관련된 연구는 지금까지 소수에 불과하다.

온라인 쇼핑의 거래액으로 대변되는 전자상거래 GDP(국내 총생산)의 약 8%를 차지할 정도로 점점 더 대중화되고 있다. 온라인 쇼핑의 쉬운 접근성과 낮은 진입장벽으로 인해 창업을 희망하는 사람의 38.2%가 온라인 쇼핑 창업을 고려할 정도로 높은 인기를 끌고 있다[11]. 그러나 한국경제신문에 따르면 창업자 10명 중 9명의 온라인 쇼핑몰이 폐쇄할 정도로 온라인 쇼핑몰 성공 및 유지 가능성이 매우 낮다[20]. 수많은 쇼핑몰이 실패하는 여러 원인중 가장 큰 원인은 판매량 수요를 잘못 예측하여 재고 관리에 실패하고 많은 어음을 떠안기 때문이다. 쇼핑몰들은 판매 상품들의 데이터를 저장하고 있지만, 데이터를 정형화하고 과학적인 분석을 통해 판매량을 예측하지는 않

는다. 노하우를 가진 직원의 견해와 감각에 의존하여 재고 관리를 결정하기 때문에 노하우를 가진 직원의 이직 및 퇴직은 쇼핑몰에 큰 타격을 주기도 한다. 예를 들어 기온이 급히 하강하여 직원의 판단 아래 많은 양의 아우터 웨어(Outer Wear) 물량을 확보하였지만 예측한 판매량보다 한참 못 미치는 판매를 하게 된다면 재고가 쌓이게 되고 이러한 일이 반복된다면 온라인 쇼핑몰은 폐쇄하게 된다.

따라서 본 연구에서는 이러한 문제점 해결을 위해 실시간으로 누적된 데이터를 바탕으로 상품별 판매량을 분석하고 각 변수의 변화에 따른 판매량을 예측하여 판매량 증대와 효율적인 재고를 위한 빅데이터 분석을 활용한 온라인 판매 수요 예측 모델 BAPP(Big data Analysis for sales Prediction Platform)를 제안하고자 한다[10]. 본 연구는 100만 이상 가입자, 2018년 기준 거래액 약 200억 원, 일 년에 500만 벌 이상을 판매하는 온라인 의류 쇼핑몰 'A' 업체에서 2014년부터 2018년까지 5년간 실시간으로 누적된 데이터를 바탕으로 연구를 진행하였다.

1.2 연구의 방법

본 연구는 과거 판매 실적 데이터를 기반으로 미래 판매량을 예측하기 위한 분석 모델이다. 데이터 수집은 파이참(PyCharm)을 사용하여 크롤링(Crawling)을 진행하였으며, 방대한 양의 판매 실적 데이터를 신속하게 정제 및 분석 처리하기 위하여 R 프로그램을 사용하였다. SPSS, SAS 등의 통계 프로그램이 아닌 R 프로그램을 사용한 이유는 첫 번째, 통계 분석 기능 제공 및 분석 자동화와 새로운 기능을 사용자가 직접 만들어낼 수 있다는 장점이 있다. 두 번째, 오픈 소스 소프트웨어이며 사용자가 쉽게 사용 가능한 환경을 가지고 있다. 세 번째, 실시간으로 분석할 수 있는 데이터양이 방대하다. 네 번째, C, C++, 자바(Java), 파이썬(Python) 등 여러 프로그램 언어와도 쉽게 연동 가능하며 윈도우, 맥(Mac)OS, 리눅스·유닉스 등 다양한 개발 환경을 지원한다. 마지막으로 하둡(Hadoop) 분산처리 환경을 지원하는 라이브러리(Library)가 제공되기 때문에 구글, 페이스북, 아마존 등에서도 빅데이터 분석에 R 프로그램을 활용하고 있다. R 프로그램으로 통계 분석을 진행하기 위해서는 CSV 파일 형태로 진행해야 한다. CSV 파일은 , (콤마)로 데이터를 분류하며 엑셀에서 저장 형식을 (.xlsx)이 아닌 (.csv)로 저장하면 된다. CSV 파일은 [표 1.1]과 같이 테이블 형태로 나타나며 [표 1.1]은 분석을 위해 실제 사용한 데이터이다. 데이터는 Date:날짜, avg Temp:평균 온도, min Temp:최소 온도, max Temp:최고 온도, rain Fall:강수량, avg Price:평균 가격, Sales:판매량의 정보를 담고 있다. 평균 온도와 평균 가격의 산출 방법은 2014년 1월 1일 ~ 2018년 12월 31일까지의 최근 5년간의 데이터를 20개씩 묶어 평균값 계산을 하는 방법으로 진행하였다. 본 연구에 사용된 2014년 1월 1일부터 2018년 12월 31일까지의 기후 데이터는 기상 자료 개방 포털을 이용하였다[2].

[표 1.1] R 프로그램에 사용한 CSV 파일

	A	B	C	D	E	F	G
1	date	avgTemp	minTemp	maxTemp	rainFall	avgPrice	Sales
2	20140101	4	0.3	7.9	0	9574	180
3	20140102	1.5	-1.3	6.2	0	9357	269
4	20140103	2.3	-1.1	7.3	0	9110	177
5	20140104	-0.1	-3.7	5.6	0	9648	137
6	20140105	-0.8	-4.3	3.8	0	9007	227
7	20140106	0.8	-3.7	6.4	0	9022	248
8	20140107	3	0.3	5.5	0	9000	243
9	20140108	0.5	-5.7	4	0	9577	234
10	20140109	-7.8	-10.4	-4.3	0	9407	213
1819	20181223	2.4	-3.4	7.5	0	5279	89
1820	20181224	-2.3	-5.9	3.1	0	9293	302
1821	20181225	0.6	-4	5.2	0	5282	61
1822	20181226	-1.5	-7.1	3	0	7613	188
1823	20181227	-10.1	-12.5	-6.5	0	6453	126
1824	20181228	-10.5	-14.4	-6.6	0	6846	107
1825	20181229	-7.6	-10.5	-3	0	4816	51
1826	20181230	-6.8	-11.3	-0.9	0	4079	78
1827	20181231	-5.4	-10.5	-0.5	0	7095	231

[표 1.1]은 R 프로그램에 사용한 데이터이며 표시한 avg Temp(평균 온도)와 avg Price(평균 가격)만을 기준으로 분석하였다. 판매량에 영향을 미치는 요인은 온도와 가격 외에도 후기 수, 홈페이지의 노출 수 등 다양한 요인들이 있다. 따라서 본 연구에서는 다양한 요인들은 모두 배제 시킨 후 평균 온도와 평균 가격을 기준으로만 연구를 진행하였다. 온도를 기준으로 한 이유는 우리나라는 4계절이 뚜렷하여 온도에 따라 판매되는 상품이 다르고 온도에 따라 하락세를 보이는 상품과 상승세를 보이는 상품들이 있기 때문이다. 이러한 장점으로서는 상품의 다양성이 있고 단점은 많은 양의 재고가 생긴다는 점이다. 가격을 기준으로 한 이유는 반팔 티셔츠와 같이 다른 상품 대비 가격이 저렴한 상품은 소비자가 직관적인 소비 성향을 띈다. 즉 조회수가 증가하는 만큼 판매량도 증가한다. 하지만 아우터 웨어(Outer Wear)와

같이 고가의 상품의 경우 구매자들이 신중하게 구매하기 때문에 조회 수가 높더라도 판매량은 저조한 경향이 있다. 이렇듯 가격에 민감한 제품들이 있기 때문에 가격을 기준으로 분석하였다. 아래 [표 1.2]와 [표 1.3]은 가격대별 판매량을 나타낸다. 온도에 따라 판매되는 상품을 직관적으로 보여주기 위해 여름 상품이 판매되는 8월과 겨울 상품이 판매되는 12월의 데이터를 표기하였다.

[표 1.2] 2017년 8월 한 달 가격대별 판매량

날짜	요일	0 ~ 5,000	5,001 ~ 10,000	10,001 ~ 15,000	15,001 ~ 20,000	20,001 ~ 25,000	25,001 ~ 30,000	30,000 ~
2017-08-01	(화)	1305	4346	769	36	10	3	0
2017-08-02	(수)	1192	3720	647	36	2	0	1
2017-08-03	(목)	1044	3287	559	43	6	2	1
2017-08-04	(금)	799	2658	475	27	1	0	2
2017-08-05	(토)	662	1984	314	29	1	1	0
2017-08-06	(일)	906	2535	433	23	3	0	0
2017-08-07	(월)	1042	2804	489	32	2	1	0
2017-08-08	(화)	903	2388	411	28	7	3	0
2017-08-09	(수)	807	1988	344	17	2	0	0
2017-08-10	(목)	679	1751	342	16	4	0	0
2017-08-11	(금)	600	1524	298	26	3	2	1
2017-08-12	(토)	489	1189	257	18	1	0	3
2017-08-13	(일)	702	1551	320	19	1	2	1
2017-08-14	(월)	981	1728	406	38	0	0	3
2017-08-15	(화)	886	1862	416	35	1	1	1
2017-08-16	(수)	723	1622	372	37	0	1	0
2017-08-17	(목)	682	1303	289	29	2	0	2
2017-08-18	(금)	531	996	252	33	1	0	1
2017-08-19	(토)	563	1131	287	27	1	0	0
2017-08-20	(일)	818	1618	420	38	1	0	1
2017-08-21	(월)	886	1740	498	56	3	0	1
2017-08-22	(화)	835	1416	407	35	2	1	3
2017-08-23	(수)	759	1415	439	43	2	0	0
2017-08-24	(목)	730	1505	471	33	1	0	2
2017-08-25	(금)	605	1170	397	45	2	0	1
2017-08-26	(토)	634	1534	561	19	3	0	0
2017-08-27	(일)	962	2391	886	49	3	1	1
2017-08-28	(월)	994	2123	848	77	3	1	1
2017-08-29	(화)	1061	2163	1010	88	1	0	1
2017-08-30	(수)	1041	2611	1263	103	8	0	4
2017-08-31	(목)	885	2366	1176	87	6	0	2

[표 1.2]를 보면 8월에는 10,000원 이하의 상품이 많이 판매되었다는 것을 알 수 있다.

[표 1.3] 2017년 12월 한 달 가격대별 판매량

날짜	요일	0 ~ 5,000	5,001 ~ 10,000	10,001 ~ 15,000	15,001 ~ 20,000	20,001 ~ 25,000	25,001 ~ 30,000	30,000 ~
2017-12-01	(금)	258	804	877	103	60	25	49
2017-12-02	(토)	324	901	1020	102	46	28	44
2017-12-03	(일)	448	1445	1573	154	86	47	49
2017-12-04	(월)	429	1232	1360	141	77	41	62
2017-12-05	(화)	394	1143	1161	131	62	43	63
2017-12-06	(수)	364	1082	1170	114	60	27	56
2017-12-07	(목)	311	899	1054	107	70	30	58
2017-12-08	(금)	314	736	836	80	42	27	44
2017-12-09	(토)	383	902	1085	104	45	44	37
2017-12-10	(일)	554	1414	1683	163	85	53	67
2017-12-11	(월)	485	1316	1543	157	78	40	55
2017-12-12	(화)	413	1094	1228	109	69	35	43
2017-12-13	(수)	405	1105	1215	112	71	41	54
2017-12-14	(목)	404	1033	1045	121	46	40	50
2017-12-15	(금)	264	798	854	91	53	24	48
2017-12-16	(토)	310	1044	1150	91	58	20	59
2017-12-17	(일)	503	1563	1594	156	97	39	80
2017-12-18	(월)	447	1187	1254	134	97	38	72
2017-12-19	(화)	325	1231	1185	130	90	31	58
2017-12-20	(수)	384	1090	1067	114	63	25	55
2017-12-21	(목)	312	931	859	99	48	15	55
2017-12-22	(금)	226	685	549	65	45	13	33
2017-12-23	(토)	252	747	654	83	19	10	35
2017-12-24	(일)	330	852	832	95	34	16	48
2017-12-25	(월)	355	1082	1145	112	32	22	51
2017-12-26	(화)	385	1018	988	101	56	20	57
2017-12-27	(수)	316	891	811	107	38	20	56
2017-12-28	(목)	226	822	729	75	48	17	35
2017-12-29	(금)	240	701	633	95	41	14	30
2017-12-30	(토)	220	791	700	88	25	11	32
2017-12-31	(일)	259	858	845	106	30	11	34

우리나라의 4계절 특성상 12월에는 겨울 상품이 판매된다. [표 1.3]을 보면 12월에는 10,000원 이상의 상품이 8월보다 더 많이 판매되었다는 것을 알 수 있다. 즉 겨울은 여름보다 상대적으로 고가의 상품이 더 많이 판매되었다는 것을 알 수 있으며 위 [표 1.2]와 [표 1.3]을 통해 소비자들이 선호하는 적정 가격대를 알 수 있다.

No.	이미지	상품명	판매량	조회수	구매전환(%)	No.	이미지	상품명	판매량	조회수	구매전환(%)
1		♥ 365DAYS 베이식가방1 / 30주(1벌)베이식티 공급사: 골든어페일 원) 3,000 원 판) 3,900 원 등록일: 2017-08-04 [재고] ★윤선별 보기 ◆장바구니	6375 (227.7)	106365	3.42	2		♥ 365DAYS 캐주얼가방1 / 베이식(1벌)티 공급사: 골든어페일 원) 3,000 원 판) 3,900 원 등록일: 2017-03-03 [재고] ★윤선별 보기 ◆장바구니	1722 (55.5)	51022	3.32
3		♥ 러브벌룬 티셔츠 / 볼룬T 공급사: 퀸스메이도 원) 3,500 원 판) 5,500 원 등록일: 2017-08-03 [재고] ★윤선별 보기 ◆장바구니	1597 (55.1)	90243	1.77	4		♥ 문라이트 자수티셔츠 / 문라이트 자수티 공급사: 골든어페일 원) 4,000 원 판) 6,900 원 등록일: 2017-06-01 [재고] ★윤선별 보기 ◆장바구니	1462 (47.2)	61145	2.39
5		♥ 카와이 양말 / 카와이패턴 공급사: 라온 원) 700 원 판) 1,000 원 등록일: 2017-03-24 [재고] ★윤선별 보기 ◆장바구니	1458 (47)	10390	14.03	6		♥ 조인트레비 단가라티♥ / 뉴트레비 공급사: 골든어페일 원) 5,700 원 판) 8,900 원 등록일: 2017-03-11 [재고] ★윤선별 보기 ◆장바구니	1277 (60.3)	91024	1.40
7		♥ (뚜대)문라이트 자수티셔츠 / (톡)문라이트 자수티 공급사: 골든어페일 원) 7,800 원 판) 10,900 원 등록일: 2017-08-22 [재고] ★윤선별 보기 ◆장바구니	1255 (125.5)	160817	0.78	8		♥ 리쿠이 디셔츠 / 리쿠이디 공급사: DISSONA(디제이소나) 원) 3,500 원 판) 5,000 원 등록일: 2017-07-12 [재고] ★윤선별 보기 ◆장바구니	1199 (38.7)	81266	1.48
9		♥ 1벌비 블랙스카니셔 / 13082 공급사: IT(아이티) 원) 7,000 원 판) 9,900 원 등록일: 2017-01-19 [재고] ★윤선별 보기 ◆장바구니	1187 (38.3)	101068	1.17	10		♥ 주류테스 사수양생티셔츠 / &월렛티셔츠 공급사: 골든어페일 원) 4,500 원 판) 6,900 원 등록일: 2017-04-14 [재고] ★윤선별 보기 ◆장바구니	1153 (37.2)	34365	3.36

[그림 1.1] 2017년 8월 한 달 TOP10 상품

No.	이미지	상품명	판매량	조회수	구매전환(%)	No.	이미지	상품명	판매량	조회수	구매전환(%)
1		♥ [저울HIT] 핫팩 / 핫팩는 핫팩 공급사: 도매국 원) 136 원 판) 300 원 등록일: 2016-12-02 [재고] ★윤선별 보기 ◆장바구니	2919 (94.2)	7479	39.03	2		♥ [기모]나인턴 후드티 / [기모]나인턴후드티 공급사: 골든어페일 원) 9,500 원 판) 13,900 원 등록일: 2017-12-01 [재고] ★윤선별 보기 ◆장바구니	2175 (70.2)	213547	1.02
3		♥ [골렛] 기모슬림자켓 / 7017-1(록 상) 7017-2(진상) 공급사: KOG 원) 7,000 원 판) 10,900 원 등록일: 2017-11-02 [재고] ★윤선별 보기 ◆장바구니	1997 (64.4)	73800	2.71	4		♥ [기모뽀뽀] 후드티 / [톡기모]무지후드 공급사: 골든어페일 원) 7,800 원 판) 10,300 원 등록일: 2017-09-29 [재고] ★윤선별 보기 ◆장바구니	1953 (63)	77874	2.51
5		♥ [기모]9리자넬 후드 / 9리자넬후드 공급사: 골든어페일 원) 9,500 원 판) 13,900 원 등록일: 2017-11-30 [재고] ★윤선별 보기 ◆장바구니	1577 (50.9)	146372	1.08	6		♥ 버엑스 양말 / 9버엑스8 공급사: 라온 원) 700 원 판) 1,000 원 등록일: 2017-12-01 [재고] ★윤선별 보기 ◆장바구니	1513 (48.8)	30573	4.95
7		♥ [기모]레옹 배색후드 / 레옹후드 공급사: 골든어페일 원) 9,200 원 판) 12,900 원 등록일: 2017-11-17 [재고] ★윤선별 보기 ◆장바구니	1318 (42.5)	103002	1.28	8		♥ [골렛] 블랙스카니셔 / 13082 공급사: IT(아이티) 원) 7,000 원 판) 9,900 원 등록일: 2017-01-19 [재고] ★윤선별 보기 ◆장바구니	1270 (41)	77173	1.65
9		♥ 10벌터 골지목롤라 / 골지목롤라 공급사: 다니모 원) 3,500 원 판) 5,000 원 등록일: 2016-12-22 [재고] ★윤선별 보기 ◆장바구니	1227 (39.6)	41293	2.97	10		♥ 캐스터 기모롤라티 / 캐스터T 공급사: 빈(Bean) 원) 7,000 원 판) 9,900 원 등록일: 2017-11-13 [재고] ★윤선별 보기 ◆장바구니	1158 (37.4)	56421	2.05

[그림 1.2] 2017년 12월 한 달 TOP10 상품

[표 1.2]와 [표 1.3]만 보면 8월과 12월 모두 판매량이 높은 것처럼 보인다. 정확한 판단은 제품명, 가격, 판매량, 조회 수, 구매 전환율의 데이터를 담고 있는 [그림 1.1]과 [그림 1.2]를 통해 알 수 있다. [표 1.2]와 [표 1.3]에서 파악할 수 없었던 조회 수 대비 구매율을 파악할 수 있으며 가격이 저렴한 상품들의 경우 조회가 판매로 직결되기 때문에 구매 전환율이 높다는 것을 알 수 있다. 반대로 고가의 상품들의 경우 구매자가 심사숙고하여 구매하기 때문에 판매 전환율이 상대적으로 낮다는 걸 알 수 있다.

II. 빅데이터 및 온라인 판매 이론적 고찰

2.1 빅데이터의 개요

빅데이터 개념은 로벌 컨설팅 기업인 맥킨지(McKinsey)에 의해 최초로 제안되었다[30]. 빅데이터는 일반적으로 데이터의 수집, 처리, 관리에 있어 소프트웨어의 수용 한계를 넘어서는 방대한 크기의 데이터를 말한다[35]. 그러나 오늘날의 빅데이터는 단순히 방대한 크기만으로 평가되지 않는다. 현재 빅데이터는 일반적으로 사용하는 도구나 방법으로는 자료를 검색, 수집, 분석, 저장, 시각화하기 어려운 상태의 정형 데이터 또는 비정형 데이터를 의미한다. 데이터베이스(Data Base)의 규모에 초점을 맞춘 맥킨지는 일반적인 데이터베이스 소프트웨어(Database Software)가 분석, 관리, 저장하는 범위를 초과하는 방대한 규모의 데이터를 빅데이터(Big Data)로 정의하였다[30]. IT 트렌드 조사와 컨설팅 전문 업체인 Forrester Research는 빅데이터는 빅밸류(Big Value)를 의미한다고 언급하고 있다[15]. 또한 IDC(International Data Corporation)는 데이터베이스가 아닌 업무수행에 중점을 두어 다양한 종류의 방대한 Data로부터 저가의 비용으로 의미 있는 데이터를 추출하고 초고속으로 수집, 분석, 발굴 할 수 있도록 고안 된 차세대 기술과 아키텍처(Architecture)라고 정의하고 있다. 결론적으로 다양한 분야에 빅데이터를 도입하고자 하는 목적은 분석된 데이터를 근거로 한 판단에 있으며[16], 즉각적이고 개별적이며 여러 방면으로 검토를 거친 자료로 더 많은 가치를 얻고자 함에 있다[7]. 그러므로 빅데이터는 방대한 데이터를 저장 및 분석하며, 빠른 처리 속도로 즉시성 있게 도출하고, 다양하고 복잡한 데이터를 효과적으로 분석하여 문제를 해결하는 데 중점을 두고 있다.

빅데이터의 대표적인 특징은 Volume(데이터의 양), Velocity(속도), Variety(다양성) 등 총 3가지로 정의 할 수 있다[29]. 첫 번째, Volume(데이

터의 양)은 데이터의 크기를 의미한다. 오늘날의 인터넷상의 데이터의 양은 과거의 인터넷상에서 발생한 모든 데이터를 합친 것보다 많다. 두 번째, 빅데이터의 특징은 Velocity(속도)이다. 경쟁자보다 빠른 분석을 할 수 있다는 건 그 자체로 경쟁력이 될 수 있다. 그렇기 때문에 현대의 모든 기업은 속도 경쟁을 이어가고 있다. 온라인 쇼핑을 하다 보면 소비자에 관한 행동 정보와 특성이 실시간으로 분석되어 해당 소비자에게 최적의 상품이 제시되는 경우가 있다. 이런 분석은 해당 기업의 고객 가치를 증가시키며 결과적으로 경쟁력 제고에 도움이 된다. 빅데이터의 마지막 특징은 Variety(다양성)이다. 빅데이터의 다양성이란 비정형 데이터부터 정형 데이터까지의 모든 데이터를 포함하는 것을 의미한다. 빅데이터는 센서가 읽어낸 정보일 수도 있고 이미지일 수도 있고 메시지일 수도 있다. 스마트 폰에서 발생한 여러 정보일 수도 있다. 이런 점에서 정형화된 데이터와 차이를 보인다. 빅데이터의 주요 원천인 이러한 다양한 데이터들은 최근에 등장한 것들이 많다. 페이스북은 2004년도, 트위터는 2006년도, 스마트폰의 조상이라 불리는 아이폰은 2007년에 출시되었다. 즉, 몇 년 사이에 수많은 비정형화된 데이터들이 발생되기 시작한 것이다. 이러한 비정형화된 데이터들은 기업들이 사용하던 기존의 정형화된 데이터베이스에서 저장 및 처리할 수 없는 유형들이 대부분이었다. 다행히 장치의 구매 비용이 지속적으로 저렴해지면서 기업들은 기존의 데이터 저장 및 처리방식을 빅데이터에 맞춰 처리 할 수 있게 되었다. 이에 따라 일상적인 비즈니스 활동에서 발생하는 빅데이터는 점차 의미 있는 자료로 변화 할 수 있게 되었다. 결과적으로 기업들이 기존에 자료를 처리 및 분석하는 방식이 주로 정형화된 자료 중심이었다면, 빅데이터의 등장에 따라 비정형화된 자료까지 포함하여 분석 및 처리가 가능하게 되었다. 가트너(Gartner)는 초기의 3V 모델에 2V를 추가하여 5V 모델을 제시하였다. 3V 모델이란 위에서 언급한 데이터의 양, 속도, 다양성을 의미한다. 추가된 2V는 Veracity(진실성)과 Value(가치) 이다[25]. 데이터의 Veracity(진실성)이란 데이터의 관리와 품질 그리고 법적 문제의 고려와 함께 데이터의 정확성을

의미한다. 데이터의 Value(가치)란 방대한 데이터에서 의미 있는 정보의 추출을 의미한다. 가치는 유용한 비즈니스 정보를 생성할 수 있기 때문에 중요한 특성으로 여겨진다. 빅데이터에 대한 정의는 보는 관점에 따라 다양하다. 하지만 결론적으로 현재 시스템의 처리 범위를 넘어서는 방대한 데이터를 말한다. 현재 빅데이터는 소비자의 선택, 개인의 행동, 교통 패턴, 질병 관리, 검색 행동 등 여러 분야에서 패턴을 분석하며 미래를 예측하는 도구로 성장하고 있다[26][27].

[표 2.1] 빅데이터 환경의 특징[14]

구분	기존 환경	빅데이터 환경
데이터	<ul style="list-style-type: none"> - 수치 자료 중심 (정형) 	<ul style="list-style-type: none"> - 다양하고 방대한 데이터(비정형)
하드웨어	<ul style="list-style-type: none"> - 데이터 베이스 - 데이터 웨어 하우스 - 고가의 저장 장치 	<ul style="list-style-type: none"> - 클라우드 컴퓨팅 등 효율적인 장비 활용 가능
소프트웨어/분석 방법	<ul style="list-style-type: none"> - RDBMS (관계형 데이터 베이스) - SAS, SPSS (통계 패키지) - Data Mining (데이터 마이닝) - Machine Learning, Knowledge Discovery 	<ul style="list-style-type: none"> - 오픈 소스 형태의 무료 소프트웨어 - 오픈 소스 통계 솔루션 R - Hadoop, NoSQL - 텍스트 마이닝 - 온라인 버즈분석 - 감성 분석

2.2 빅데이터 기술과 분석 기법

2.2.1 빅데이터 기술

빅데이터 기술은 시스템적으로 복잡한 구성과 기능을 가지고 있기 때문에 분야별로 전문화된 기술 및 제품의 통합이 필요하다. 빅데이터 기술은 크게 데이터 수집, 저장, 처리, 분석, 표현/활용, 관리 분야로 나눌 수 있다. 각 분야에서 요구되는 문제점 및 기술은 다음과 같다[3].

- 수집 기술: 데이터 적재 시간 단축 및 실시간 대용량 데이터 수집에 관련된 기술이 필요하다. 문제점은 대용량 데이터의 적재 시간이 전체 시간의 대부분을 차지하고 데이터가 계속해서 증가한다는 점이다.

- 저장 기술: 대용량 데이터의 저장과 효율적 관리, 낮은 TCO(Total Cost of Ownership) 실현 기술이 필요하다. 문제점은 데이터 저장 및 관리에 높은 비용이 든다는 점이다.

- 처리 기술: 다차원 데이터의 다중 노드와 고속 연산을 수행 할 수 있는 분산/병렬처리 기술이 필요하다. 그러나 장시간의 처리로 인해 데이터의 적시성이 부족하며 처리 및 연산에 높은 비용이 든다.

- 분석 기술: 통계적 기법 기반의 실시간 분석과 고급 분석, 탐색적 데이터 분석 기술 등이 필요로 한다. 문제점은 데이터분석에 높은 비용, 너무 많은 연산 시간이 소요된다는 점과 통계 분석 기법의 검증 문제와 프로그램 구현, 분석 전문가의 부재 등이다.

- 표현/활용 기술: 방대한 데이터의 요약 기술, 실시간 인지 및 대응 기술, 직관적 표현 기술, 분석 표현 기술 및 다차원 데이터의 비교, 분석된 빅데이터의 해석 및 의미 파악을 위한 지원 도구 등이 필요하다.

- 관리 기술: 오픈소스 관리 도구 제공 및 설치, 모니터링, 관제 등의 통합 인프라 관리 체계 기술, 수집부터 활용까지 각 분야를 통합하는 비즈니스 로직 관리 기술 등이 필요하다. 단점은 오픈소스 기반 관리 도구의 부족, 다

수 장비로 구성된 인프라의 관리가 어렵다는 점과 분석 관련 비즈니스 로직이 분산되어 관리하기 어렵다는 점 등이다. 빅데이터 기술의 분류는 처리, 저장, 분석의 구현 방식에 따라 Disk(디스크) 기반 기술, In-Memory(인-메모리) 기반 기술, In-Database(인-데이터베이스) 기반 기술 등으로 나눈다. 디스크 기반 기술은 대부분 오픈소스 기술들로 저비용 디스크를 사용하여 대용량 데이터를 관리, 저장, 분산, 병렬 처리하는 기술이다. 다수의 서버를 네트워크(Network)로 분산, 연결, 저장하여 저장소를 확장하고, 병렬 처리를 이용해 연산 성능을 향상해준다. 디스크 기반 SW 솔루션은 Hadoop, DISCO, Hbase, Cassandra, MongoDB, Sybase IQ, GreenPlum, Vertica Endeca 등이 있다. In-Memory 기반 기술은 탐색적 데이터 분석이 가능하며, 메모리 내에 있는 고속 연산 기능을 이용하여 방대한 양의 데이터를 신속하게 처리할 수 있다. 디스크보다 읽기, 쓰기 속도가 빠르며, 압축 기술과 Grid computing(그리드 컴퓨팅) 형태로 단일 노드의 메모리 제약을 극복하였다. SW 솔루션으로는 R, Revolution R, Tableau, Spotfire, SAP HANA 등이 있다. In-Database 기반 기술은 RDBMS(Relational DataBase Management System) 또는 분산/병렬 DBMS(DataBase Management System) 내부에 분석 기능을 탑재하여 DB 서버 간 데이터 이동이 가능, 적재 부하 시간을 단축 등 분석의 효율성을 강조한 기술이며 SAS In-Database, Tera data Aster 솔루션 등이 있다[22].

2.2.2 빅데이터 분석 기법

빅데이터 분석 기법은 통계학 및 전산 분야에서 사용되던 기법을 대규모 데이터 처리에 맞도록 보완 및 수정하여 적용되고 있다. 최근에는 SNS, 동영상 등에서 발생하는 비정형 데이터의 증가로 인해 새로운 분석 기법들이 주목받고 있다.

- Text Mining: 텍스트 마이닝은 텍스트 형태의 데이터 중 비정형, 반

정형 데이터에서 자연어 처리 기술을 기반으로 하여 유용한 정보를 추출한 후 가공하는 것을 목적으로 하는 기술이다. 이 기술을 통하여 방대한 텍스트 데이터에서 가치 있는 정보를 추출하고, 다른 데이터와의 관련성을 파악하며 단순한 정보 검색이 아닌 그 이상의 결과를 얻어낼 수 있다.

- Opinion Mining: Text Mining: 기술과 연관된 분야로 평판 분석 (Sentiment Analysis)이라고 불리기도 한다. 소셜 미디어와 웹 사이트에 나타난 의견과 여론을 분석하여 유용한 정보로 재가공하는 기술이다. 인터넷 이용자들이 각각의 사건에 대하여 이야기하는 포스팅이나 댓글 등을 긍정, 부정, 중립 등으로 분류하여 정확하고 객관적으로 평판을 파악할 수 있다. 단순히 묻고 답하는 방식이 아닌 이용자들의 생각과 표현의 파편을 모아 일정한 법칙을 찾아내고 특정 서비스 및 상품에 대한 시장규모 예측과 소비자의 반응 등을 분석할 때 활용된다.

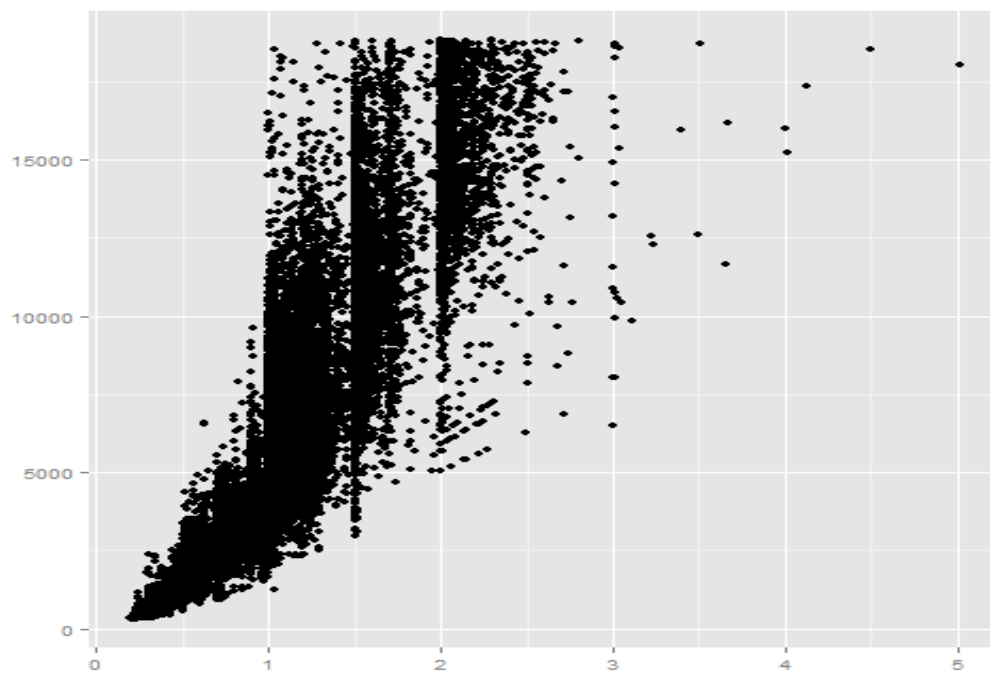
- Social Network Analytics: 수학의 그래프 이론에 기반을 두고 있다. 스마트 디바이스의 발달로 더욱 활성화된 SNS상에서 사용자의 영향력을 측정하고, 주요 사용자를 모니터링 및 관리하여 마케팅에 활용한다.

- Cluster Analytics: 클러스터 분석은 유사성을 측정한 후 유사성이 높은 집단을 분류하고, 군집에 속한 객체들의 유사성과 다른 군집에 속한 객체 간의 서로 다른 성질을 밝히는 통계 분석 방법이다. 이 기술을 통해 소셜 미디어 등에서 취미나 서로의 관심이 유사한 사용자군을 분류할 수 있다.

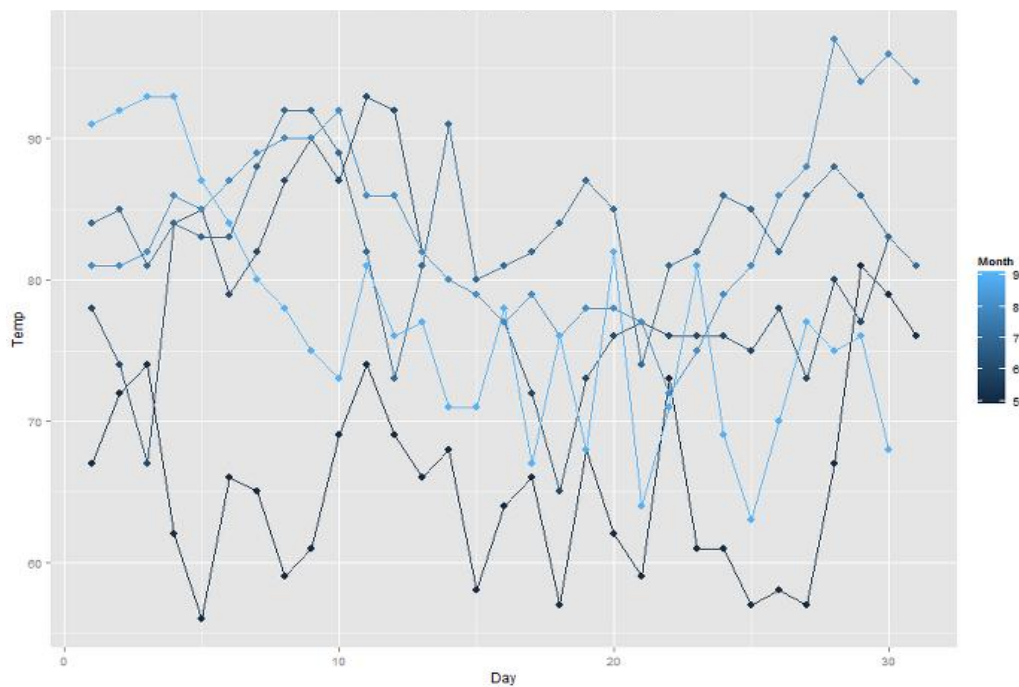
2.2.3 빅데이터 시각화

데이터 시각화(Visualization)는 R 프로그램을 활용해 앞에서 소개한 분석 기법으로 분석된 자료를 다양한 차트나 그래프로 시각화함으로써 데이터의 의미를 효율적으로 전달할 수 있다. 시각화는 데이터 분석에 능숙한 사람이 데이터에서 얻어 낸 정보를 능숙하지 않은 사람에게 설명하기 위해 필수적인 부분이다. 또한 데이터 분석을 시작할 때 데이터의 특성과 구성을 알아

보는 데에도 편리한 방법이다. 빅데이터는 단순히 자료의 양이 방대할 뿐만 아니라 자료의 차원이 크고, 종류가 다양하기 때문에 단순한 2차원 그래프로는 자료의 상관관계나 분포를 파악하기 어렵다. 따라서 의미 있는 정보를 찾아내기 위해 여러 각도의 시각화가 필요하며, 시각화를 통해 기존 통계 분석을 적용하기 어려운 데이터에 어떤 분석을 적용해야 할지 방향을 제공할 수 있다. 빅데이터 시각화 툴(Tool)로 주목받고 있는 R 프로그램 패키지는 ggplot2와 BigVis가 있다. BigVis는 억 단위의 큰 데이터 셋(Data Set)을 시각화하기 위한 전처리를 지원한다. 데이터를 잘게 자르고 각 조각의 요약 정보를 저장한 후, spline(스플라인)을 적합 시키며, spline의 정도 또한 최적화할 수 있다. 본 연구에서는 시각화를 위해 ggplot2 패키지를 사용하였다. ggplot2 패키지는 섬세한 시각화를 위해 만들어졌다. Grammar of Graphic에서 발전되었으며 특유의 문법을 가져 진입장벽이 높지만, 그래프의 섬세한 부분까지 조정할 수 있으므로 디자인과 색상을 조정하여 보기 좋은 그래프를 만드는 데 용이하다. [그림 2.1]과 [그림 2.2]는 R 프로그램 패키지 중 ggplot2로 그린 간단한 그래프 예시이다.



[그림 2.1] R 프로그램으로 그린 산점도



[그림 2.2] R 프로그램으로 그린 꺾은선 그래프

2.3 빅데이터 분석 활용 사례

2.3.1 정치, 사회, 문화 활용 사례

빅데이터의 정치 분야의 사례로는 2008년 미국 대통령 선거를 예로 들 수 있다. 미국 대통령 후보인 버락 오바마는 다양한 형태의 유권자 데이터베이스를 확보하였고 이를 분석하여 유권자 맞춤형 선거 전략을 진행하였다. 당시 오바마 캠프는 종교, 인종, 가구 형태, 소비수준, 나이와 같은 기본 인적 사항으로 유권자를 분류하는 것을 넘어서서 구독하는 잡지, 마시는 음료, 과거 투표 여부 등 유권자 성향까지 소셜 미디어나 전화 또는 개별 방문을 통해 유권자 정보를 수집하였다. 수집된 데이터는 오바마 캠프로 전송되었고 유권자 정보를 온라인으로 통합 관리하는 VoteBuilder.com(보우트빌더) 시스템의 도움으로 유권자에 대한 예측, 미결정 유권자 선별, 유권자 성향 분석을 해나갔다. 이를 바탕으로 유권자 맞춤형 선거 전략을 전개하였고 오바마는 비용 대비 효과적인 선거를 치렀다[17]. 사회 분야 활용 사례로는 서울시의 심야버스 노선 확정이 있다. 서울시는 심야버스 노선을 확정하는 데 빅데이터를 활용하였다. 우선 KT와 함께 심야에 핸드폰을 사용하는 사람들의 정보를 분석 및 수집하였다. 발신자 위치(출발지)와 소유자의 주소지(도착지)를 연결하여 심야에 이동해야 할 길을 찾고 서울시는 이 데이터를 바탕으로 새로운 노선을 확정하였다[6]. 문화 부분 활용 사례로는 머니볼 이론이 있다. 머니볼 이론이란 분석한 경기 데이터를 기반으로 적재적소에 선수들을 배치하여 승률을 높이는 게임 이론이다[5]. 영화로도 만들어진 머니볼 이론은 오 클랜드 애슬레틱스(Oakland Athletics)의 구단장인 빌리 빈은 출루율이 승패에 영향을 미친다는 가정하에 선수들의 데이터를 분석하였고 출루율은 높지만 저평가된 선수들을 고용하였다. 그 결과 메이저리그 최초로 20연승의 신기록을 세웠다. 정확한 의미에서 머니볼 이론은 빅데이터 분석은 아니지만 데이터 분석 수요가 미미했던 스포츠 분야에서 데이터 분석의 중요함을 보

여주었다는 점과 그 당시 비정형화된 데이터를 활용했다는 점에서 빅데이터의 개념과 유사한 사례라 볼 수 있다.

2.3.2 경영, 마케팅 활용 사례

경영 분야에서의 빅데이터 분석이 가장 많이 사용될 것으로 예측되는 분야를 조사한 결과 마케팅 49%, 생산관리 18%, 재무15%, 신사업14%, 인사 4% 등으로 나타났다[33]. 마케팅 분야에서 방대한 양의 데이터의 저장이 이루어진 예시로 상품 Scanner(스캐너)가 있다. 처음 Scanner는 계산 편리성과 가격 식별 등 매장 관리를 더욱 편하게 하기 위해 사용 되었고 이 이유만으로 그 당시 상당한 혁신이었다. Scanner가 설치된 후 저장되기 시작한 데이터의 중요성을 인식하였고 마케팅 분야에서 Scanner data를 활용한 연구들이 활발히 진행되었다[31]. Scanner data의 사용 이후 기업의 데이터 활용에 큰 변화가 생겼다. 인트라넷, 인터넷 등 네트워크 기술의 발달로 고객 관계 관리(CRM), 전사적 자원관리 시스템(ERP) 등 IT 기반 시스템을 위한 DB 구축이 진행되었고, Big Data가 그 뒤를 따르고 있다. 아마존은 소비자들의 구매 내역을 데이터베이스에 저장하고 이 기록을 분석하여 소비자의 관심사와 소비 취향을 파악한다[13]. 이러한 빅데이터의 활용을 통해 아마존은 소비자에게 알맞은 추천 상품을 표시한다. 고객 한 사람 한 사람의 성향을 파악해 가장 일치한다고 생각되는 상품을 자동으로 제시하는 것이다[4]. 아마존의 추천 상품 표시와 같은 방식으로 페이스북 및 구글도 빅데이터의 활용을 증대시키고 있다. 페이스북 및 구글은 이용자의 검색 조건 또는 사진과 동영상 같은 비정형 데이터를 즉각 처리하여 맞춤형 광고를 제공하는 등의 노력을 기울이고 있다. 인터넷으로 시작해서 인터넷으로 끝내는 생활, 스마트폰을 이용해 쇼핑 및 정보 검색을 하고 실시간으로 글을 남기는 등의 인터넷을 이용하는 동안 남은 데이터 흔적들을 분석하면 소비 성향 및 개인의 생활 패턴 등을 예측할 수 있고 기업들은 이러한 예측 데이터를 통해서

소비자의 취향을 파악 및 예측 할 수 있다[8].

2.4 온라인 판매 기업의 빅데이터 분석 활용 사례

2.4.1 국외 패션 기업 활용 사례

소비 취향의 다양화와 트렌드의 짧은 변화 주기로 인해 패션업계에서는 빅데이터가 주목의 대상이 되고 있으며 다양한 종류와 많은 양의 데이터를 보유하고 있는 글로벌 패션 기업들을 중심으로 마켓 분석이 시도되고 있다. 글로벌 스타일 사이트(Global Style Sight)의 대표 프랭크 바버(Frank Barbour)는 다음 해 트렌드를 예상하는 데 따르는 위험을 줄이려면 데이터의 도움을 받아야 한다고 하였다. 즉, 데이터는 트렌드에 민감한 의류 업체가 시장의 변화를 파악하고 예측하는 데 도움을 주며 신제품을 개발하거나 제품의 라인업(Line-Up)을 결정하는 데 중요한 역할을 한다는 것이다. 패션 정보관리와 경영에 빅데이터를 활용하는 EDITD(에딧디)는 인터넷상의 정보를 실시간으로 수집, 분석하여 의류 제조업체와 유통업체에 시장의 최신 현황을 알려준다. 또한 고객의 반응, 제품 가격 등을 경쟁사와 비교 및 분석할 수 있게 도와준다. 실제로 온라인 패션 유통업체인 Asos(에이소스)는 EDITD의 분석 데이터를 바탕으로 가격 경쟁력을 갖췄고 2013년 매출이 전년 대비 37% 증가했다. 결과적으로 이러한 분석은 매출과 수익을 최대화하고 재고관리를 최적화시킴으로써 경영 혁신을 주도하고 있다. 또한 패션 서비스 산업에서도 데이터 분석 기술을 활용하고 있다. 기존에는 전문가의 직관과 지식으로 제품을 추천하는 형태였지만 커머스 서비스가 등장함으로써 보다 합리적으로 소비할 수 있는 새로운 경험의 가치를 제공하는 형태로 패러다임의 변화가 나타나고 있다. Go Find AI(고 파인드 에이아이) 기업은 이미지 데이터 분석 기술을 활용하여 서비스를 제공하고 있다. 이미지 데이터 분석 기술은 사용자가 스마트폰으로 촬영한 의류와 최대한 일치되는 옷

을 찾아 쇼핑할 수 있도록 해주는 패션 큐레이션 서비스이다. 온라인 쇼핑몰 Stitch fix(시티치 픽스)는 빅데이터 분석을 토대로 맞춤형 패션 아이템을 선별하여 배송하는 서비스를 제공하고 있다. 이 배송 서비스는 개인의 취향에 맞춰 정기적으로 액세서리와 의류를 배송해 주는 서비스로 사용자들의 제품에 대한 만족도를 높이고 있다. 또 다른 온라인 쇼핑몰 기업인 Rent the runway(렌트 더 런웨이)는 스마트 대여 서비스를 제공하고 있다. 이 서비스는 사용자 체형 데이터를 총 10만개 이상의 제품에 적용하여 시각적으로 제시하는 새로운 스마트 서비스이다. 디자인을 선택할 때 자신의 신체 정보가 반영된 실제 착용 이미지를 제공받음으로서 최적의 사이즈와 스타일을 갖춘 제품을 선택할 수 있다는 점에서 긍정적인 반응을 얻고 있다. 패션 분야는 시즌별로 빠르게 변화한다. 따라서 빅데이터를 활용하면 새로운 트렌드와 소비자 선호를 쉽게 예측할 수 있고 그 점에서 패션 업계에서 빅데이터가 주목받고 있다. SPA 브랜드인 ZARA(자라)는 2015년 15% 이상의 영업이익률을 기록한 바 있다. ZARA가 높은 영업이익률을 기록 할 수 있었던 방법으로는 세계 70여 개국에 7천 개에 가까운 매장에서 판매된 1만 1000여 종의 제품에 대한 데이터를 분석하여 패션 트렌드의 최적화된 상품을 기획, 재고 관리, 디자인, 출시에 적용하였기 때문이다. 텍스툼의 Fashion Marketing Intelligence Service Platform은 패션 제품이나 서비스에 대한 사용자의 태도, 라이프스타일, 인식 등 데이터를 실시간으로 수집하여 데이터 분석을 하였고 분석 데이터를 바탕으로 패턴을 분석함으로써 트렌드 예측하고 있다. 결국 짧은 유행 주기 및 소비 취향의 다양화는 시장의 세분화를 이끌었으며, 많은 양의 비정형화된 데이터는 IT기술을 접목하여 빅데이터 분석을 통해 기업에 재고 부담과 손실로부터 적극적인 전략을 수립하도록 이끌고 있다 [12].

2.4.2 국내 패션 기업 활용 사례

국내 패션사업 부문에 있어서 빅데이터 활용은 제품 기획에서부터 판매에 이르기까지 브랜드의 운영방식에 따라 다양하게 활용되고 있으며, 사업 부문과 운영 방식을 근거로 대표적인 사례를 통해 국내 패션사업 부문의 빅데이터 활용의 동향을 파악해 보면 그 내용은 다음과 같다. 먼저, 코오롱 인더스트리얼 FnC는 기존에 보유한 내부고객 데이터와 매장에서 축적한 데이터를 기반으로 매장별 입점 고객, 브랜드별 판매 시장, 고객의 가치, 과거 패턴, 온라인과 오프라인의 매출 상관관계 등을 분석하여 상품판매량과 고객 선호상품을 예측하는 데 활용하여 적중률을 높이고 있다. 코오롱 인더스트리얼 FnC의 빅데이터 활용 방식의 특징을 살펴보면, 먼저, 소셜 네트워크의 자료 분석의 방대함과 활용에 대한 제한성으로 온라인 쇼핑몰 위주의 소비자 데이터 분석에 집중하고, 구체적인 브랜드 틀에서 벗어나 제품이나 키워드를 중심으로 데이터를 분석한다. 예를 들면 코오롱 스포츠의 경우, 해당 브랜드에 국한하지 않고 캠핑, 텐트, 코펠 등 제품이나 키워드를 중심으로 데이터를 분석해 소비자 성향을 파악하고 키워드를 통해 열거된 제품의 불편한 점이나 좋은 점을 수집하여 분석한 결과로 서비스나 상품 개발에 아이디어를 전달한다. 또한, 매장과 유통채널, 고객을 중심으로 빅데이터를 분석하고, 브랜드의 웹 쇼핑몰 상에서 발생하는 데이터가 실제 오프라인 매출에 영향을 미치는지를 분석함으로써 제품에 대한 소비자의 관심도가 높은 상품과 매출과의 상관관계를 분석하여 이에 대한 예측치를 매장에 전달하며, 판매 가능수량 확보 및 대체 상품 제안에 대한 내용까지 미리 준비한다. 아울러, 쇼핑몰 회원의 데이터뿐 아니라 비회원의 방문 기록을 임의의 숫자를 붙여 관리하고 회원가입 순간 파격적인 혜택으로 고객 대상 프로모션을 구사하여 홍보하고 있다. 온라인 여성 의류 쇼핑몰 스타일난다는 온라인 사이트 가입 회원들과 방문자를 분석하고, 판매되는 상품 데이터 분석하여 매장 입지, 상품분배, 마케팅 방식 등을 결정하는데 빅 데이터를 활용하고 있다. 스타일난다

의 빅데이터 활용 방식은 온라인 사이트에 회원 가입하거나 방문한 소비자들의 로그 분석을 바탕으로 한다. Web log analysis(웹 로그 분석)이란 Web 사이트의 방문객이 남긴 데이터를 근거로 Web의 방문 형태 및 운영에 대한 정보를 분석하는 것으로서 방문객이 Web 사이트에 방문하게 되면 Web 서버에는 에러 로그, 액세스 로그, 에이전트 로그, 리퍼럴 로그 등의 형태로 자료가 파일 형식으로 기록된다. 에러 로그는 오류의 유무, 액세스 로그는 누가 어떤 것을 읽었는지, 에이전트 로그는 웹 브라우저의 이름, 운영 체제(OS), 화면 해상도, 버전 등의 정보, 리퍼럴 로그는 검색 엔진 키워드의 단서와 경유지 사이트 등의 정보로 기록된다. 결과적으로 웹 로그 분석에 의해 얻은 방문 유형, 방문자 수, 각 웹 페이지별 방문 횟수, 월·요일·시간·계절별 접속 통계 등의 자료는 마케팅 및 웹의 운영 자료로 유용하게 활용된다. 온라인 패션 유통기업 무신사는 온라인 사이트 내에서 상품을 구매하는 소비자의 사이즈 스펙 데이터를 분석하여, 이후 방문 시 자주 구매하던 상품 스펙과 유사한 상품들을 선별하여 추천 제안한다. 라이프 스타일과 오프라인으로 시각을 돌리는 기존 온라인 패션업체와는 달리 패션과 온라인에만 집중하는 기업으로 물류 시스템과 입점사를 위한 전산 시스템 개선하고 고객을 위한 빅데이터 기반의 딥러닝 기술을 통해 입점사와 고객들을 위한 서비스를 제공하는 특징을 지니고 있다. 구체적으로 무신사의 딥러닝 서비스란 유사 상품을 학습에 의해 자동 추천하는 시스템으로 고객이 자주 구매하는 상품들의 디자인과 패턴을 자동으로 분석해 제안하는 방식이다. 대부분의 일반 쇼핑몰이 상품의 카테고리를 직원들이 수동으로 설정하거나 데이터값(상품군/컬러)을 통해 관련 상품을 보여주지만, 무신사는 특정 상품을 클릭했을 때 과거의 구매 데이터를 기반으로 비슷한 상품을 찾아 추천해주며 클릭 수나 구매 수 등의 데이터를 기반으로 다시 한번 분류할 수 있는 기능을 부가한 시스템으로 서비스를 제공하고 있다. 데님 전문 브랜드인 피스 위커는 빅데이터를 상품 개발에 활용하는 업체로서, 전년 대비 4배 성장을 이끌며 주목을 받고 있다. 먼저, 피스 위커는 빅데이터를 제품개발에 적용함으로써 고객

의 니즈 분석에 활용하여 괄목할 만한 결과를 얻고 있다. 피스 워커의 빅데이터 활용의 특징은 텍스트 마이닝 기술 활용에 있다. 텍스트 마이닝은 텍스트 데이터에서 유용한 정보를 추출하고 가공하여 새롭고 유용한 정보를 찾아내고 온라인 패션 블로그, 패션 커뮤니티, SNS 소셜미디어에 올라온 수십만 개의 제품 관련 키워드를 수집·분석하여 도출된 결과를 차기 제품개발에 적용하는 방식이다. 실제로 피스 워커는 특정 아이템에 대한 소비자의 불만 및 요구사항을 다양한 인사이트를 도출하여 기존 데님의 부정 이슈에 대응하는 제품(크롭 핏의 데님 진)을 개발하여 매출 신장에 성공하였으며, 워싱턴을 구매하는 다수 고객의 동일 제품 반품 사례에 관한 데이터를 통해 레이저 워싱 기술을 통한 제품의 일관성과 공정대비 워싱 오차율 감소를 이끌어냈다. 결과적으로 피스 워커는 데이터를 통해 고객의 니즈와 정보에 집중함으로써 품질향상과 고객 만족, 매출 신장이라는 결과를 이끌어 내었으며, 빅데이터를 제품개발에 활용한 성공적인 사례로 들 수 있다. 통합 커머스 플랫폼 서비스 전문기업 SK플래닛은 프로젝트 앤이라는 빅데이터 사업으로 주목받고 있다. 옷장에 옷은 많지만 정작 입을 옷은 없다는 소비자들의 요구를 적극적으로 반영하여 필요한 옷을 필요할 때 입을 수 있게 한 패션 대여 서비스로 국내·외 유명 패션 브랜드부터 신진 디자이너의 아이템까지 모바일 앱을 통해 이용할 수 있다. 앱 사용자는 의류를 기준으로 한 달에 일정의 이용료를 지불하고 서비스를 받는다. 프로젝트 앤은 고객에게 잘 어울리는 스타일을 모바일 앱을 통해 실시간으로 추천하는 커스터마이징을 기본으로 한다. 따라서 제조자와 소비자에 대한 구체적인 정보가 양방향·실시간으로 축적된다는 장점이 있다. 또한 서비스의 부여된 가치는 단순한 의류 대여가 아닌 옷을 소비하는 새로운 방식으로 구매에 국한된 의류 소비 방식의 전환에 의미가 있다. 빅데이터 기반 패션 셀렉숍 알케미스트(Alchemist)는 인기 있는 패션 상품을 고객에게 효과적으로 큐레이팅하는 온라인 편집숍으로 최신 트렌드와 최고의 패션 아이템을 데이터로만 검증, 분석하며 고객 개개인을 대상으로 고객에게 어울릴 만한 상품을 제안하는 업체이다. 국내 시장에

서 정형데이터와 비정형데이터를 모두 이용해 의류를 판매하는 업체로 쇼핑 물에 출시된 상품의 통계자료를 모두 분석하고, SNS의 비정형데이터를 통해 제품을 설명하고 이를 소비자에게 제시한다. 또한 소비자 구매패턴 추적서비스(Purchasing Pattern Tracking Service)를 통해 고객의 구매 유형을 추적하여 편하고 합리적인 구매 활동을 유도하는 서비스를 운영한다[12][24]. 이렇듯 다양한 기업에서 빅데이터를 활용하고 있다. 하지만 Economist(이코노미스트)가 전 세계 약 600개 기업을 대상으로 실시한 빅데이터에 관한 조사에서 대상자의 46%는 빅데이터가 기업 의사결정의 중요한 요소로 작용할 것이라 응답했고 10%는 빅데이터가 기존 비즈니스 모델을 완전히 바꿀 것이라고 응답했다. 그러나 응답자의 53%는 일부만 활용하고 있다고 응답하였고 25%는 기업 내부에서 발생하는 데이터는 충분하지만, 대부분의 데이터를 방치하고 있다고 응답하였다. 이는 빅데이터를 활용한 부가가치 창출을 위해서는 더 많은 노력이 필요함을 시사하고 있다[24]. 특히 고객의 반응에 민감한 패션 업계에서는 빅데이터를 활용한다면 더 많은 부가가치를 창출할 수 있을 것으로 예상되지만, 국내·외 패션 기업에서 빅데이터를 활용하는 곳은 소수에 불과하며 주로 소비자 성향 분석, 상품 관리 등에만 빅데이터를 활용하고 있다.

III. 온라인 판매 시스템 설계 및 구현

3.1 연구 구조의 이론적 설명

3.1.1 시계열 예측 방법에 관한 이론적 고찰

시계열에 대한 최초의 예측연구는 Reid에 의해 이루어졌지만, 예측 결과에 대한 정확한 비교연구는 Makridakis&Hibon에 의해 수행되기 시작했다 [28][32]. 미래에 대한 일정한 상태를 설정하기 위해서는 미래 예측이 뒷받침되어야 한다. 그러나 의류는 다른 분야보다 미래를 예측하기 어려우며, 또한 제약요인이 많기 때문에 효과적인 미래 예측을 위해서는 정확한 수요 예측이 선행되어야 한다. 수요 예측 방법은 크게 두 가지로 나눌 수 있다. 첫 번째 관측된 과거 자료에 포함된 정보를 이용하여 예측 연구에 필요한 경험적 법칙을 추정하는 정량적 예측 방법과 전문가들의 주관적 견해를 사용하는 정성적 예측 방법으로 나눌 수 있다[9]. 정성적 예측 방법은 예측에 이용할 과거의 자료나 경험을 이용할 수 없을 때 사용하는 예측법으로는 기술적 예측법, 단순 예측법 등이 있다. 정성적 예측 방법은 과거에 대한 정보가 수집 및 관측을 할 수 있고, 양적인 자료로 나타낼 수 있으며, 과거의 패턴은 미래에도 지속할 것이라는 가정이 성립되어야 한다. 여기에는 공간 상호작용 모형, 회귀모형 그리고 시계열 예측 방법 등으로 나눌 수 있는데, 회귀모형은 많은 설명변수들과 다양한 모형을 적용하기 때문에 추정식의 편의는 줄어들지만, 설명변수들 사이의 다중공선성이 증가해서 분산이 늘어나기 때문에 추정식을 신뢰하기가 어렵다는 단점이 있다[18]. 따라서 특정 변수의 미래 예측을 수행하고자 할 경우 시계열 예측 방법이 유용한 방법이 될 수 있는데, 과거에 발생했던 사실이 미래에도 관련성이 있다는 가정하에 과거 자료의 경향 분석을 통해 미래 수요를 예측하기 때문이다[19]. 시계열 예측 방

법에는 분석대상이 되는 변수가 하나인 단일변량 시계열 예측 방법과 변수가 두 개 이상인 다변량 시계열 예측 방법으로 나눌 수 있다. 단일변량 시계열 예측은 확률과정의 이론을 도입하는 ARIMA 모형과 같은 확률적 시계열 예측 방법과 평활법이나 분해법과 같은 고전적 시계열 예측 방법으로 구분된다. 하지만 시계열 분석을 활용해 수요를 예측한 선행 연구 결과에 따르면 고전적 시계열 예측 방법보다 확률적 시계열 예측 방법이 더 정확한 예측을 하는 것으로 보고되고 있다[23]. 이러한 결과는 확률과 정의 이론을 도입하는 ARIMA 모형이 비정상 시계열 자료인 계절 요인 또는 추세요인과 같은 결정적 부분 제거, 변수변환, 차분을 이용해서 정상 시계열로 바꾸어 적합한 모형을 선택하여 예측하기 때문이다. ARIMA 모형은 여러 형태의 단일변량 시계열 자료를 확률 과정모형, 즉 AR, MA, ARMA, ARIMA 및 계절 ARIMA 등의 이론적 특성을 기초로 하여 시계열 자료로부터 최적의 모형을 결정하는 분석 방법으로, 비교적 완전한 이론적 체계를 갖추고 있으며, 거의 모든 형태의 시계열 자료에 적용할 수 있는 장점이 있다[1]. 그리고 다른 시계열 예측 방법보다 비교적 복잡한 산술적 및 통계적 과정을 거치기 때문에 때로는 사용자의 주관적 판단이 요구되기도 한다. 또한 다른 방법에 비해 예측치의 정확도가 높아 단기 또는 중기 예측에 적합한 방법이라고 할 수 있다. 따라서 본 연구에서도 이 분석 방법을 이용하여 온라인 판매 수요 예측 연구를 진행하려 한다.

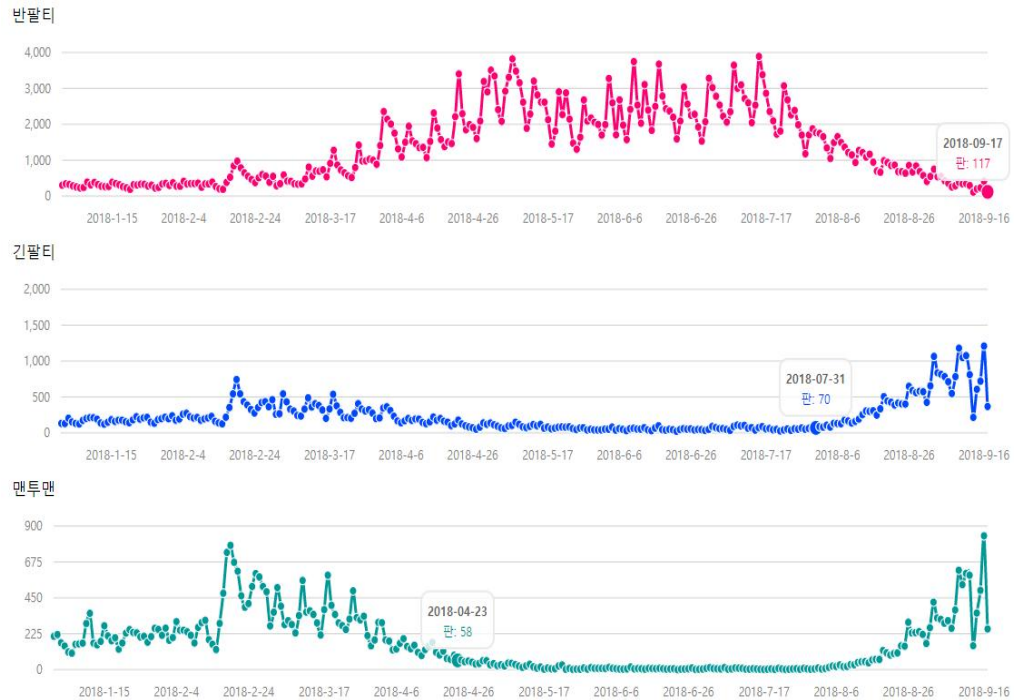
3.1.2 A기업의 온라인 쇼핑물 데이터

크롤링은 컴퓨터 내의 수많은 문서를 수집하는 기술이다. Web 환경에서의 각종 데이터를 자동 수집하는 기술을 Web Crawling(웹 크롤링)이라 한다. 뉴스, SNS 등 다양한 미디어들의 콘텐츠 증가로 Web Crawler(웹 크롤러)의 중요성이 더욱 부각되고 있다[34]. 본 연구에서는 2014년 1월 1일 ~ 2018년 12월 31일까지 최근 5년간 온라인 의류 쇼핑몰 ‘A’ 업체에서 실시간으로 누적된 데이터를 활용하였다. 수집된 데이터는 날짜별 상품의 가격, 조회 수, 판매량, 사용자 리뷰, 질의응답, 홈페이지 가입자 수 등 총 7가지 이상 변수들이 저장된 데이터를 사용하였다. 따라서 빅데이터의 기본적인 조건인 3V(Volume, Variety, Velocity)를 모두 만족하며 데이터 정형화 과정을 거친 후 분석된다. [그림 3.1], [그림 3.2], [표 3.1], [표 3.2], [표 3.3]은 온라인 의류 쇼핑몰 ‘A’에서 실시간으로 크롤링한 데이터의 예시를 나타낸다.

No.	이미지	상품명	판매량	조회수	구매전환(%)	No.	이미지	상품명	판매량	조회수	구매전환(%)
1		♥ (기모)OPWEE 후드티 / OPWEE기모후드티 공급사 : 글든어퍼월 원) 9,500 원 판) 13,900 원 등록일: 2018-11-02 [재고] ★음선택 보기 ◆장바구니	2373 (79.1)	97221	2.44	2		♥ [시즈간기모행복 후드티 / 기모행복후드티 공급사 : 글든어퍼월 원) 8,300 원 판) 11,900 원 등록일: 2018-09-06 [재고] ★음선택 보기 ◆장바구니	2092 (24)	374444	0.56
3		♥ 편독 기모플라티셔츠 / 무지름플라 공급사 : DJSONA(디제이소나) 원) 3,500 원 판) 5,000 원 등록일: 2018-10-29 [재고] ★음선택 보기 ◆장바구니	1957 (57.6)	50564	3.87	4		♥ 말랑말랑 후리스텐투렌 / 푸딩mtm 공급사 : 덕별 원) 7,000 원 판) 9,900 원 등록일: 2018-10-19 [재고] ★음선택 보기 ◆장바구니	1918 (43.6)	163577	1.17
5		♥ [2만장(편매플라)특가세팅] PEN012 배색루 드집업 / PEN012배색루드집업 공급사 : 글든어퍼월 원) 5,000 원 판) 7,900 원 등록일: 2018-09-14 [재고] ★음선택 보기 ◆장바구니	1612 (20.4)	308562	0.52	6		♥ [시즈간기모행복 맨투맨 / 록기모MTM 공급사 : 글든어퍼월 원) 6,500 원 판) 9,500 원 등록일: 2018-09-05 [재고] ★음선택 보기 ◆장바구니	1466 (16.7)	163302	0.90
7		♥ 10컬러 글자목플라 / 글자플라 공급사 : 라디오 원) 3,500 원 판) 5,000 원 등록일: 2016-12-22 [재고] ★음선택 보기 ◆장바구니	1365 (4.1)	146704	0.93	8		♥ (기모)스타이트 트레이닝팬츠 / 글든스타이트 트레이닝팬츠 공급사 : 글든어퍼월 원) 7,700 원 판) 11,500 원 등록일: 2017-12-29 [재고] ★음선택 보기 ◆장바구니	1213 (3.6)	238948	0.51
9		♥ (기모)가성비값 트레이닝팬츠 / 기모2선슈리 닝 공급사 : HS레트리 원) 3,500 원 판) 5,900 원 등록일: 2018-11-02 [재고] ★음선택 보기 ◆장바구니	1156 (38.5)	39830	2.90	10		♥ 리브 글자양말 / 리브8 공급사 : 라온 원) 700 원 판) 1,000 원 등록일: 2018-08-27 [재고] ★음선택 보기 ◆장바구니	1146 (11.8)	19724	5.81

[그림 3.1] 실시간으로 수집된 데이터

[그림 3.1]은 2018년 1월 1일에서 2018년 12월 31일까지 실시간으로 수집된 데이터의 예시이다.



[그림 3.2] 실시간으로 수집된 상의 판매량 그래프

[그림 3.2]는 실시간으로 수집된 2018년 상의 판매량 그래프이다. 위 그래프는 아래 [표 3.1]을 바탕으로 표현된 그래프이다.

[표 3.1] 실시간으로 수집된 상의 판매량 데이터

날짜	판매량			날짜	판매량			날짜	판매량			날짜	판매량			날짜	판매량		
	반팔티	긴팔티	맨투맨		반팔티	긴팔티	맨투맨		반팔티	긴팔티	맨투맨		반팔티	긴팔티	맨투맨		반팔티	긴팔티	맨투맨
2018-01-01	302	132	209	2018-02-28	384	364	521	2018-04-27	1602	51	47	2018-06-24	3037	59	3	2018-08-11	936	160	31
2018-01-02	341	130	220	2018-03-01	550	462	488	2018-04-28	2089	80	35	2018-06-25	2563	51	4	2018-08-11	936	160	31
2018-01-03	324	203	168	2018-03-02	287	259	274	2018-04-29	3192	138	39	2018-06-26	2250	56	6	2018-08-12	1270	192	45
2018-01-04	283	152	148	2018-03-03	344	266	360	2018-04-30	2903	119	56	2018-06-27	2279	41	10	2018-08-12	1270	192	45
2018-01-05	254	132	110	2018-03-04	587	543	514	2018-05-01	3507	137	56	2018-06-28	1926	47	1	2018-08-13	1211	254	49
2018-01-06	228	124	103	2018-03-05	418	432	398	2018-05-02	3346	112	31	2018-06-29	1531	45	2	2018-08-13	1211	254	49
2018-01-07	240	174	159	2018-03-06	411	330	282	2018-05-03	2407	92	38	2018-06-30	2070	33	4	2018-08-14	1085	304	50
2018-01-08	389	200	161	2018-03-07	342	303	307	2018-05-04	2081	70	25	2018-07-01	3282	48	8	2018-08-14	1085	304	50
2018-01-09	313	212	166	2018-03-08	330	242	283	2018-05-05	2923	61	30	2018-07-02	3022	90	13	2018-08-15	1165	299	43
2018-01-10	380	211	289	2018-03-09	339	234	230	2018-05-06	3307	95	23	2018-07-03	2780	74	7	2018-08-15	1165	299	43
2018-01-11	321	191	353	2018-03-10	476	331	339	2018-05-07	3823	103	41	2018-07-04	2536	59	6	2018-08-16	945	307	61
2018-01-12	273	137	166	2018-03-11	806	488	559	2018-05-08	3481	148	40	2018-07-05	2223	60	5	2018-08-16	945	307	61
2018-01-13	264	121	155	2018-03-12	554	360	361	2018-05-09	3156	119	32	2018-07-06	2059	47	13	2018-08-17	698	247	65
2018-01-14	268	148	177	2018-03-13	699	406	369	2018-05-10	2615	84	22	2018-07-07	2349	34	3	2018-08-17	698	247	65
2018-01-15	386	183	275	2018-03-14	692	381	346	2018-05-11	1888	73	14	2018-07-08	3646	90	5	2018-08-18	664	335	64
2018-01-16	350	160	211	2018-03-15	732	319	293	2018-05-12	2285	91	24	2018-07-09	3004	108	11	2018-08-18	664	335	64
2018-01-17	314	177	181	2018-03-16	537	202	217	2018-05-13	3203	117	34	2018-07-10	3098	98	9	2018-08-19	986	504	120
2018-01-18	260	176	200	2018-03-17	912	333	376	2018-05-14	2816	100	17	2018-07-11	2716	104	8	2018-08-19	986	504	120
2018-01-19	228	154	128	2018-03-18	1275	537	592	2018-05-15	2618	119	10	2018-07-12	2600	64	3	2018-08-20	935	443	105
2018-01-20	187	141	167	2018-03-19	866	376	403	2018-05-16	2615	65	16	2018-07-13	2045	73	3	2018-08-20	935	443	105
2018-01-21	319	181	229	2018-03-20	725	286	346	2018-05-17	2127	80	3	2018-07-14	2533	39	6	2018-08-21	850	426	92
2018-01-22	308	225	251	2018-03-21	647	211	293	2018-05-18	1449	57	10	2018-07-15	3892	73	4	2018-08-21	850	426	92
2018-01-23	331	191	233	2018-03-22	565	212	276	2018-05-19	1810	65	6	2018-07-16	3382	87	2	2018-08-22	864	387	101
2018-01-24	328	209	230	2018-03-23	517	203	251	2018-05-20	2907	78	6	2018-07-17	2862	56	1	2018-08-22	864	387	101
2018-01-25	281	216	205	2018-03-24	795	273	317	2018-05-21	2270	83	23	2018-07-18	2356	61	2	2018-08-23	680	416	105
2018-01-26	302	149	209	2018-03-25	1419	406	494	2018-05-22	2876	79	29	2018-07-19	2094	43	5	2018-08-23	680	416	105
2018-01-27	226	134	172	2018-03-26	980	331	326	2018-05-23	2143	83	2	2018-07-20	1723	48	1	2018-08-24	679	403	149
2018-01-28	241	184	207	2018-03-27	980	305	312	2018-05-24	1472	60	8	2018-07-21	1810	27	8	2018-08-24	679	403	149
2018-01-29	336	197	259	2018-03-28	1026	321	335	2018-05-25	1305	46	1	2018-07-22	3070	41	5	2018-08-25	637	401	147
2018-01-30	356	218	252	2018-03-29	998	274	213	2018-05-26	1639	67	3	2018-07-23	2679	54	1	2018-08-25	637	401	147

[표 3.1]은 2018년 실시간으로 수집된 상의 판매량 데이터의 일부이다.

[표 3.2] 2018년 온도에 따른 판매 데이터

날짜	요일	평균기온	최저기온	최고기온	강수량	평균판매가	총판매수량	총판매가
2018-12-01	토	5.5	0.8	12.4	0	10,529	1,457	15,341,300
2018-12-02	일	8.2	2.8	11.8	0	10,294	2,198	22,627,400
2018-12-03	월	11.5	9.3	13.5	9.5	12,169	2,237	27,223,694
2018-12-04	화	6.7	-0.3	11.8	5	11,508	1,787	20,565,073
2018-12-05	수	0	-4.3	3.4	0	11,287	1,766	19,933,286
2018-12-06	목	2.4	-1	6.8	0	11,046	1,832	20,237,141
2018-12-07	금	-7.3	-9.6	-1.1	0	10,791	1,405	15,161,823
2018-12-08	토	-8.3	-11.4	-4.4	0	10,603	1,614	17,113,500
2018-12-09	일	-7.2	-11.8	-2.1	0	10,267	2,503	25,699,300
2018-12-10	월	-2.5	-8.1	4.6	0	11,305	2,692	30,434,186
2018-12-11	화	0.2	-2.4	2.9	0	10,988	2,144	23,559,425
2018-12-12	수	-1.9	-5	2.4	0	11,089	2,205	24,452,330
2018-12-13	목	-2.2	-5.9	2.1	1.3	10,496	1,978	20,762,204
2018-12-14	금	-4.4	-7.8	0.6	0	10,773	1,743	18,777,652
2018-12-15	토	-2.4	-7.8	3.2	0	10,600	1,603	16,992,800
2018-12-16	일	0	0	0	0	10,435	2,499	26,078,900
2018-12-17	월	1.6	-1.5	5.9	0.3	11,665	2,936	34,249,411
2018-12-18	화	0	0	0	0	11,094	2,179	24,175,206
2018-12-19	수	4	1.3	9.5	0	10,350	2,198	22,751,011
2018-12-20	목	3.4	-1.1	10.4	0	10,795	2,004	21,633,375
2018-12-21	금	5.6	0.4	9.8	0	10,147	1,513	15,353,445
2018-12-22	토	6.9	2.9	12.4	0	9,730	1,367	13,302,100
2018-12-23	일	2.4	-3.4	7.5	0	10,225	1,932	19,754,800
2018-12-24	월	-2.3	-5.9	3.1	0	11,185	2,462	27,539,289
2018-12-25	화	0.6	-4	5.2	0	9,767	1,808	17,659,600
2018-12-26	수	-1.5	-7.1	3	0	11,292	2,081	23,498,657
2018-12-27	목	-10.1	-12.5	-6.5	0	10,556	1,383	14,599,024
2018-12-28	금	-10.5	-14.4	-6.6	0	10,673	1,186	12,658,691
2018-12-29	토	-7.6	-10.5	-3	0	9,996	1,109	11,085,700
2018-12-30	일	-6.8	-11.3	-0.9	0	10,119	1,379	13,954,200
2018-12-31	월	-5.4	-10.5	-0.5	0	11,597	1,644	19,066,780

[표 3.2]는 2018년 온도에 따른 판매 데이터이다. 평균 기온, 최저 기온, 최고 기온, 강수량, 평균 판매가, 총 판매 수량, 총 판매가의 정보를 담고 있다.

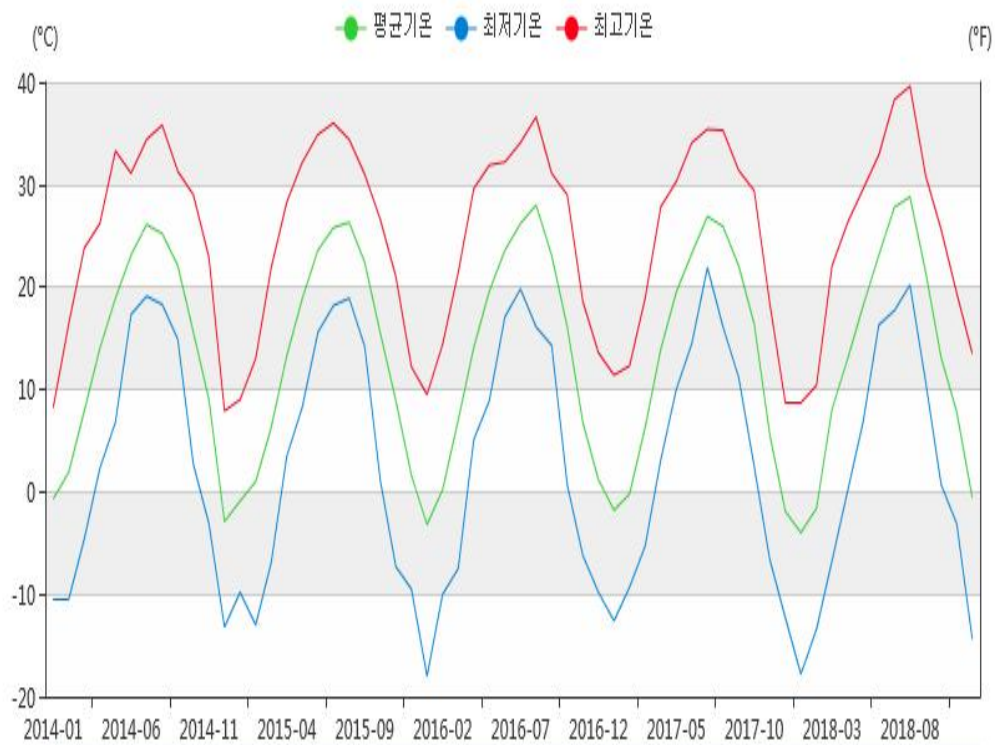
[표 3.3] 2017년 온도에 따른 판매 데이터

날짜	요일	평균기온	최저기온	최고기온	강수량	평균판매가	총판매수량	총판매가
2017-12-01	금	-2.2	-7.6	3.1	0	10,259	1,041	10,679,950
2017-12-02	토	1.9	-4.6	8.1	0	10,417	1,215	12,657,700
2017-12-03	일	4.9	3.1	8.1	4	10,312	1,957	20,181,000
2017-12-04	월	-1.2	-6.1	4.5	0.1	9,836	1,751	17,223,965
2017-12-05	화	-4.2	-8.2	-0.4	0.1	9,916	1,546	15,331,325
2017-12-06	수	0.2	-4.5	6	1.2	10,146	1,521	15,433,575
2017-12-07	목	0.7	-3.5	4.6	0	10,196	1,337	13,633,235
2017-12-08	금	-3.8	-7.1	-0.5	0	10,037	1,090	10,940,900
2017-12-09	토	-0.5	-5.6	5.4	0	10,159	1,427	14,498,300
2017-12-10	일	1.1	-2.5	6.7	4.1	10,408	2,084	21,691,300
2017-12-11	월	-7	-11	-2.5	0	10,219	1,912	19,538,728
2017-12-12	화	-9.4	-12.3	-5.2	0	9,910	1,608	15,936,415
2017-12-13	수	-8.7	-12	-4.4	0	9,873	1,679	16,577,495
2017-12-14	목	-6.6	-11.2	-1.9	0	9,896	1,485	14,696,944
2017-12-15	금	-3	-7.5	0.9	0	10,194	1,245	12,691,675
2017-12-16	토	-5.7	-8.1	-2.6	0	10,200	1,434	14,627,800
2017-12-17	일	-6.7	-10.8	-2.3	0	10,237	2,158	22,093,200
2017-12-18	월	0	0	0	0	10,205	1,714	17,492,635
2017-12-19	화	-3.9	-6.5	-1	0	10,130	1,815	18,387,180
2017-12-20	수	-3.9	-8.5	-0.1	0.2	9,994	1,494	14,931,775
2017-12-21	목	-0.9	-5.6	4.6	0	9,915	1,282	12,712,290
2017-12-22	금	3.4	-1.9	6.6	0	9,892	886	8,764,913
2017-12-23	토	6.1	5	8.7	0	9,854	1,022	10,071,200
2017-12-24	일	2.9	-0.6	5.7	20	9,972	1,123	11,199,300
2017-12-25	월	-2.1	-4.9	0.6	0	9,969	1,594	15,891,300
2017-12-26	화	-3.8	-7.9	-1.1	0	10,105	1,433	14,480,644
2017-12-27	수	-6.9	-10.3	-2.6	0	9,987	1,248	12,464,420
2017-12-28	목	-2.1	-6.8	0.8	0	9,850	1,106	10,894,495
2017-12-29	금	2.1	-0.4	6	0	10,047	934	9,384,254
2017-12-30	토	1.4	-1.2	4.5	0.7	9,917	1,020	10,115,900
2017-12-31	일	-0.3	-3.2	3.3	1	10,156	1,106	11,232,800

[표 3.3]은 2017년 온도에 따른 판매 데이터이다. [표 3.2]와 [표 3.3]을 통해 온도에 따른 판매량을 예측할 수 있다.

3.1.3 기상청 기후 데이터

본 연구에 사용된 기후 데이터는 기상 자료 개방 포털을 이용하였다[2]. 기상 자료 개방 포털에서 2014년 1월 1일부터 2018년 12월 31일까지의 평균 기온, 최저기온, 최고기온의 데이터를 수집하였다. [그림 3.2], [표 3.4]는 수집한 기상 데이터 자료의 예시이다.



[그림 3.3] 2014년 ~ 2018년 기온 그래프

[그림 3.2]는 2014년부터 2018년까지의 평균 기온, 최저 기온, 최고 기온을 나타낸 그래프 이다.

[표 3.4] 기상청 기후 데이터

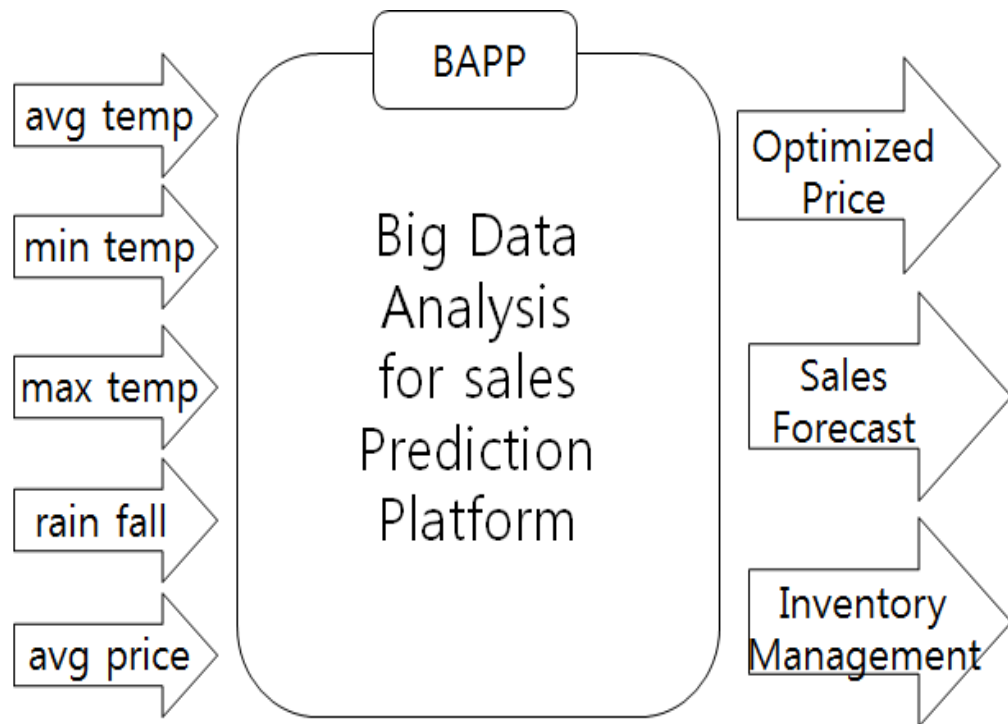
일시	평균기온(°C)	평균최고기온(°C)	최고기온(°C)	최고기온일자	평균최저기온(°C)	최저기온(°C)	최저기온일자
Jan-14	-0.7	3.5	8.2	2014-01-29	-4.9	-10.5	2014-01-13
Feb-14	1.9	6.6	16.4	2014-02-26	-1.8	-10.5	2014-02-04
Mar-14	7.9	13.1	23.8	2014-03-28	3.5	-4.6	2014-03-07
Apr-14	14	19.6	26.2	2014-04-24	9.4	2.3	2014-04-06
May-14	18.9	25.3	33.3	2014-05-31	13.8	6.8	2014-05-05
Jun-14	23.1	27.8	31.1	2014-06-06	19.7	17.3	2014-06-12
Jul-14	26.1	30.6	34.4	2014-07-31	22.9	19.1	2014-07-24
Aug-14	25.2	29	35.8	2014-08-02	22.1	18.3	2014-08-11
Sep-14	22.1	27	31.3	2014-09-07	18	14.9	2014-09-19
Oct-14	15.6	21.4	29	2014-10-12	10.7	2.7	2014-10-28
Nov-14	9	13.8	22.9	2014-11-01	4.9	-3.1	2014-11-13
Dec-14	-2.9	1.5	7.9	2014-12-01	-6.7	-13.2	2014-12-18
Jan-15	-0.9	3.6	9	2015-01-15	-4.8	-9.8	2015-01-01
Feb-15	1	6	13	2015-02-15	-2.9	-13	2015-02-09
Mar-15	6.3	12.4	21.9	2015-03-20	1	-6.9	2015-03-10
Apr-15	13.3	18.8	28.3	2015-04-30	8.4	3.5	2015-04-08
May-15	18.9	24.9	32.2	2015-05-28	13.6	8.3	2015-05-05
Jun-15	23.6	29.5	34.9	2015-06-10	18.9	15.6	2015-06-21
Jul-15	25.8	30	36	2015-07-11	22.6	18.2	2015-07-04
Aug-15	26.3	30.8	34.4	2015-08-07	22.8	18.9	2015-08-26
Sep-15	22.4	28.1	31	2015-09-22	17.5	14.2	2015-09-27
Oct-15	15.5	21.2	26.6	2015-10-17	10.8	1.1	2015-10-31
Nov-15	8.9	12.7	21	2015-11-12	5.9	-7.3	2015-11-27
Dec-15	1.6	6	12.2	2015-12-13	-2.1	-9.5	2015-12-28
Jan-16	-3.2	1.1	9.5	2016-01-02	-6.8	-18	2016-01-24
Feb-16	0.2	5.1	14.4	2016-02-13	-4.1	-10	2016-02-15
Mar-16	7	12.6	21.4	2016-03-31	2.4	-7.5	2016-03-01
Apr-16	14.1	20.3	29.6	2016-04-26	9.1	5.1	2016-04-04
May-16	19.6	25.7	31.9	2016-05-19	14.1	8.9	2016-05-03
Jun-16	23.6	28.7	32.2	2016-06-05	19.5	17.1	2016-06-01
Jul-16	26.2	29.8	34.1	2016-07-22	23.3	19.8	2016-07-16

[표 3.4]는 기상청에서 수집 가능한 기후 데이터의 일부분이다.

3.2 판매 수요 예측 시스템 구성

본 연구에서 진보된 수요 예측 시스템인 BAPP(Big Data Analysis for sales Prediction Platform)을 제안한다. 시스템의 구성은 [그림 3.3]과 같이 크게 데이터 수집, 데이터 분석, 데이터 출력으로 구분된다. 첫 번째, 데이터 수집 영역에서는 판매량 수요 예측을 위한 과거 판매 실적 데이터를 수집하여 저장하고 매 수집 주기마다 데이터를 업데이트 하여 저장한다. 본 논문에서는 크롤링을 통하여 데이터 수집을 진행하였고 수집된 데이터를 서버에 저장하였다. 두 번째, 수집한 데이터를 분석 목적에 맞게 정제하고 R 프로그램을 활용해 분석을 진행하였다. 세 번째, Optimized Price(최적의 가격)을

도출하여 Inventory Management(효율적인 재고 관리)를 위한 Sales Forecast(판매량 예측)을 하였다.



[그림 3.4] BAPP 구성도

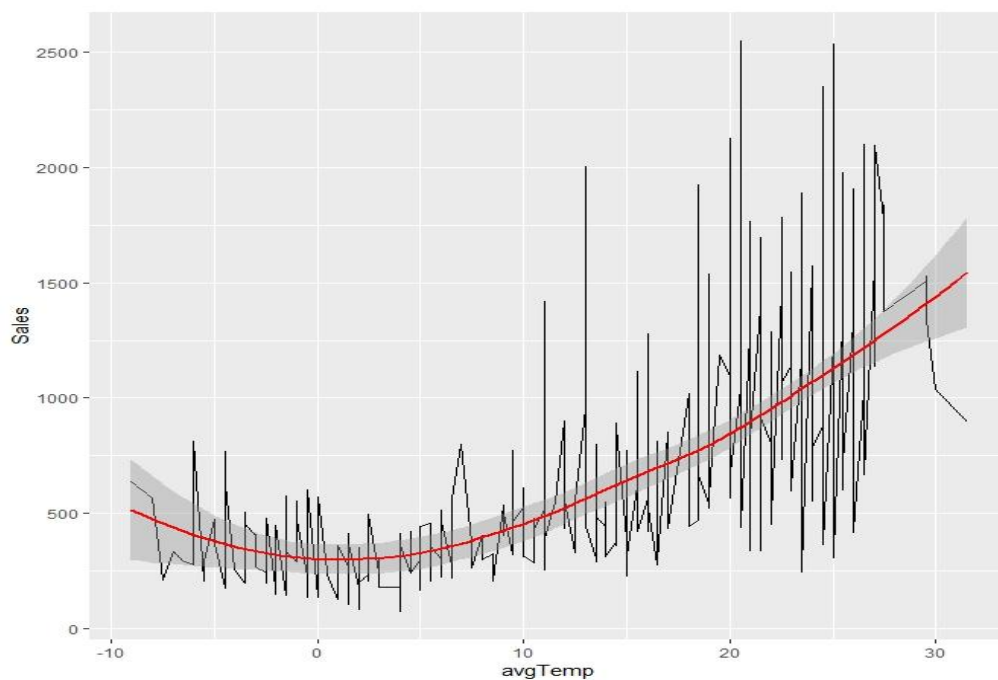
BAPP는 Date:날짜, avg Temp:평균 온도, min Temp:최소 온도, max Temp:최고 온도, rain Fall:강수량, avg Price:평균 가격, Sales:판매량의 데이터를 통해 수요 예측 시스템을 구성한다.

IV. 온라인 쇼핑몰 시스템 결과 및 분석

4.1 효과 및 검증

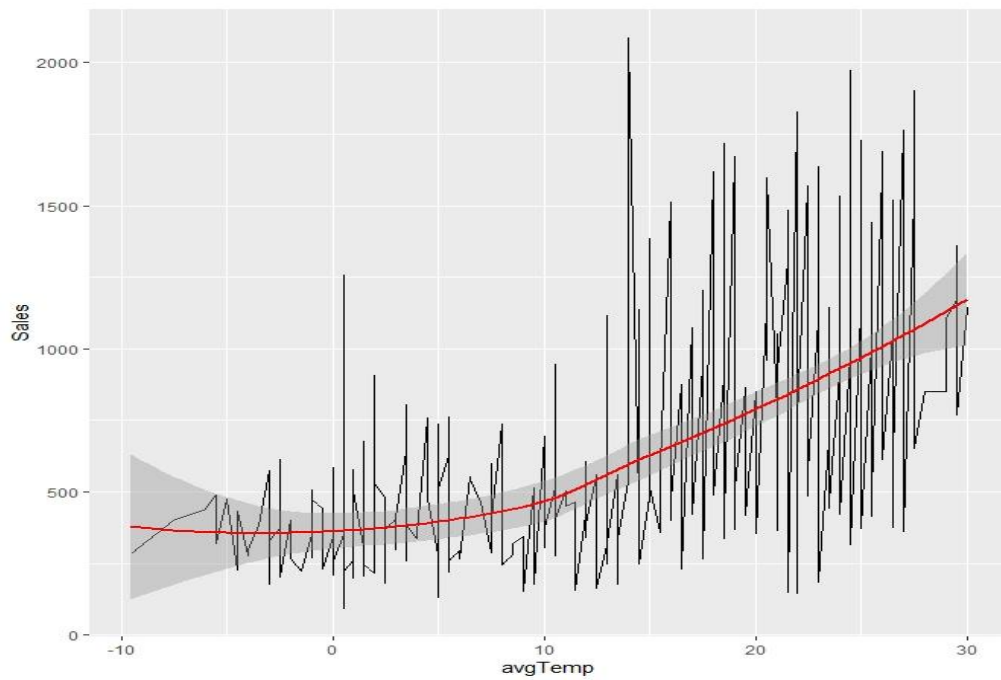
4.1.1 평균 온도 변화에 따른 판매량 예측

이번 장의 내용은 R 프로그램을 활용한 데이터 분석이다. 평균 온도, 평균 가격 변화에 따른 판매량에 대한 분석이며 반팔 티셔츠, 아우터 웨어(Outer Wear) 2가지 제품에 대한 분석을 다뤘다. 빨간색 그래프는 판매 데이터의 추세선을 나타낸다.



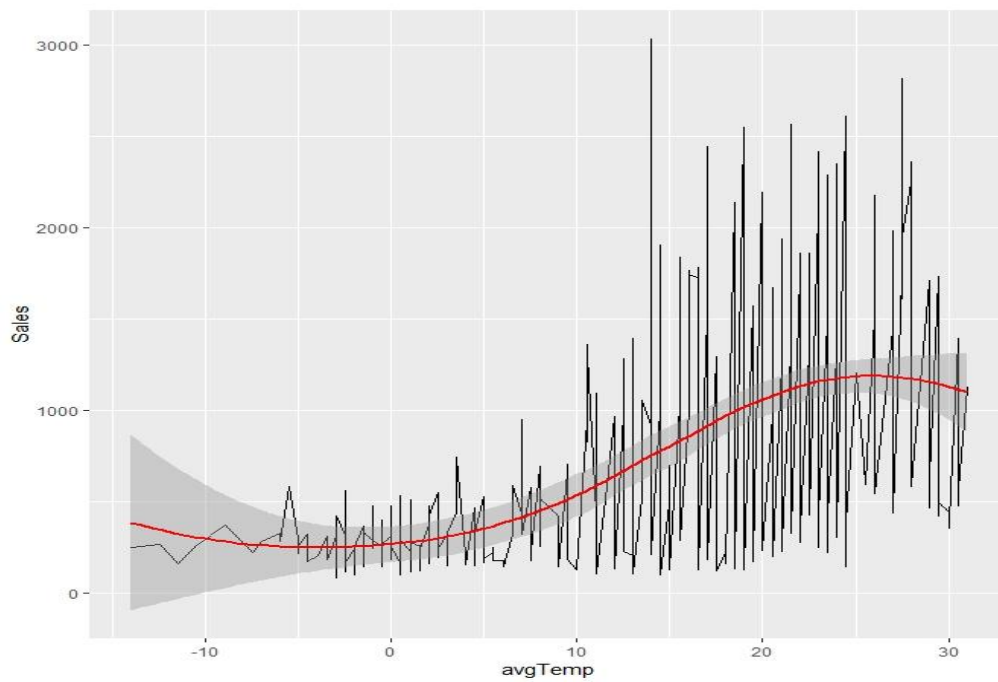
[그림 4.1] 2014년도 온도 변화에 따른 반팔 티셔츠 판매량

기상청 기후 데이터에 따르면 2014년도에는 평균 온도가 0도 이하인 53번, 15도 이상은 185번, 30도 이상은 2번이다.



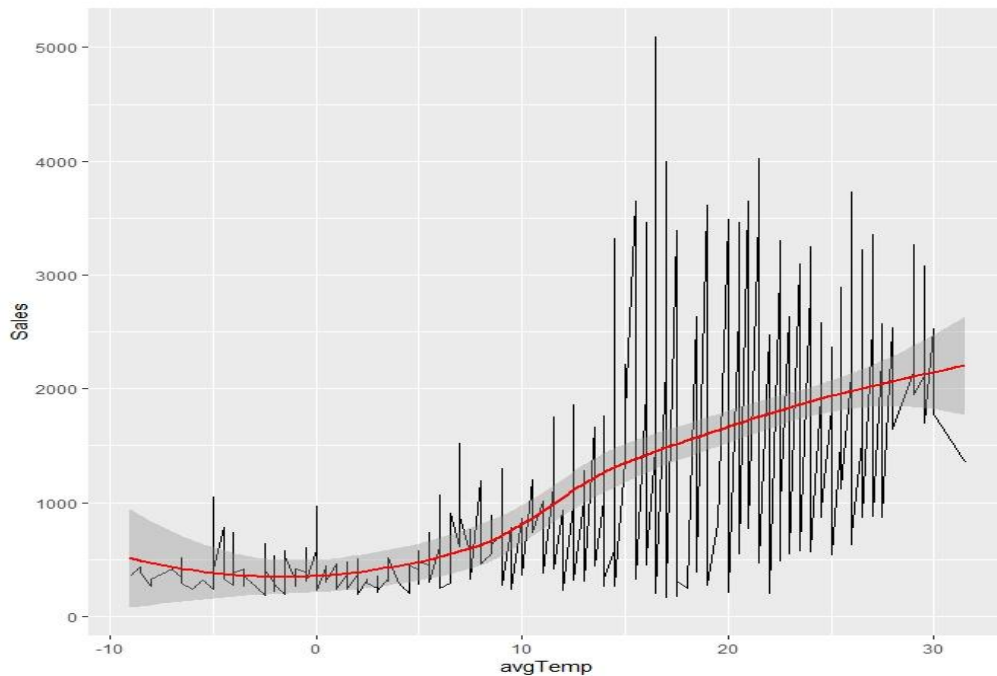
[그림 4.2] 2015년도 온도변화에 따른 반팔 티셔츠 판매량

기상청 기후 데이터에 따르면 2015년도에는 평균 온도가 0도 이하는 48번, 15도 이상은 181번, 30도 이상은 1번이다.



[그림 4.3] 2016년도 온도변화에 따른 반팔 티셔츠 판매량

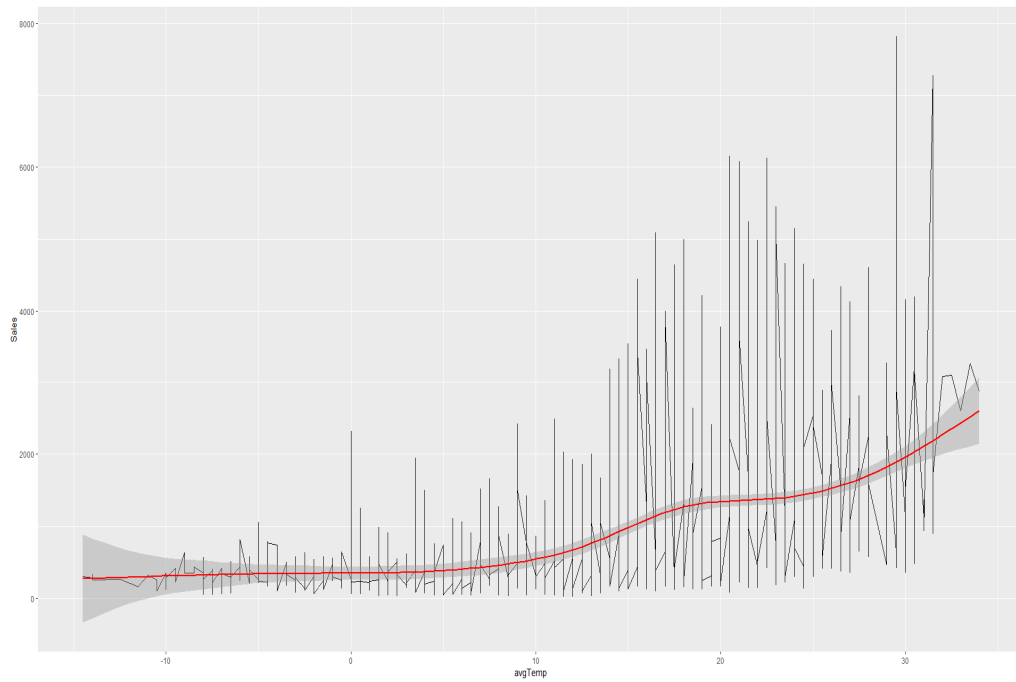
기상청 기후 데이터에 따르면 2016년도에는 평균 온도가 0도 이하는 58번, 15도 이상은 184번, 30도 이상은 11번이다.



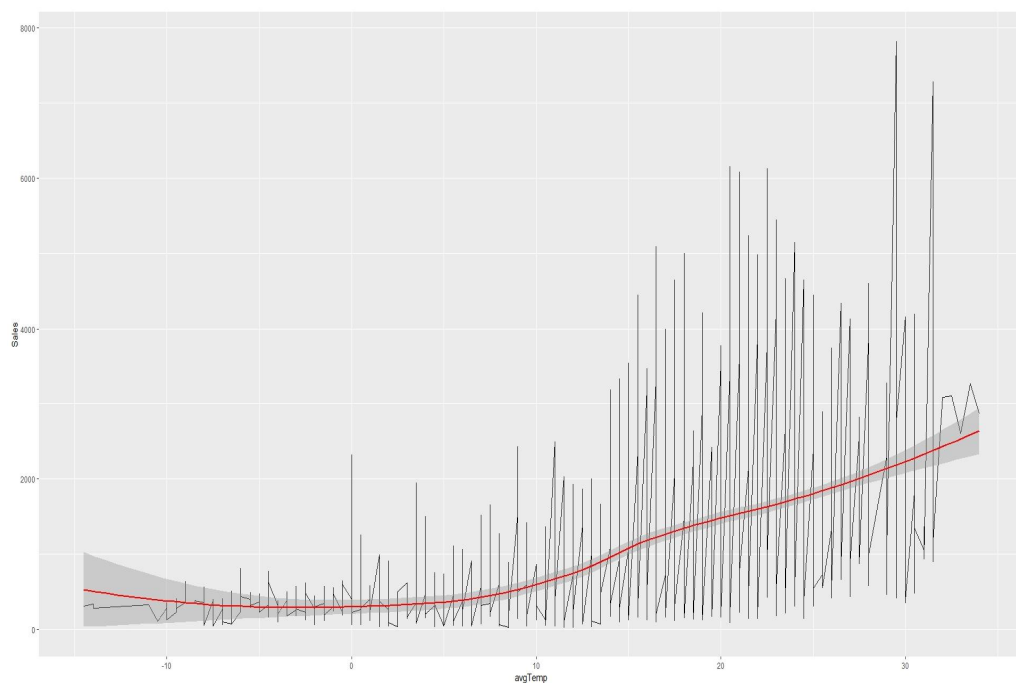
[그림 4.4] 2017년도 온도변화에 따른 반팔 티셔츠 판매량

기상청 기후 데이터에 따르면 2017년도에는 평균 온도가 0도 이하는 69번, 15도 이상은 181번, 30도 이상은 4번이다.

[그림 4.1], [그림 4.2], [그림 4.3], [그림 4.4] 평균 온도 변화에 따른 반팔 티셔츠 판매량이다. 매년 비슷한 온도가 유지되고 있기 때문에 2014년부터 2017년도 반팔 티셔츠 판매량은 매우 유사한 그래프 형태를 보인다. 반팔 티셔츠는 날이 더울 때뿐만 아니라 평소에도 인 웨어(In Wear)로 많은 사람이 활용하기 때문에 꾸준한 판매량을 보인다. 또한 날이 더워지기 시작하는 온도인 14도 ~ 16도 사이에 매년 최고치의 판매량을 기록하고 있다. 지난 4년간의 데이터를 통해 2018년 평균 온도 변화에 따른 반팔 티셔츠의 판매량을 예측하면 [그림 4.5]와 같다.

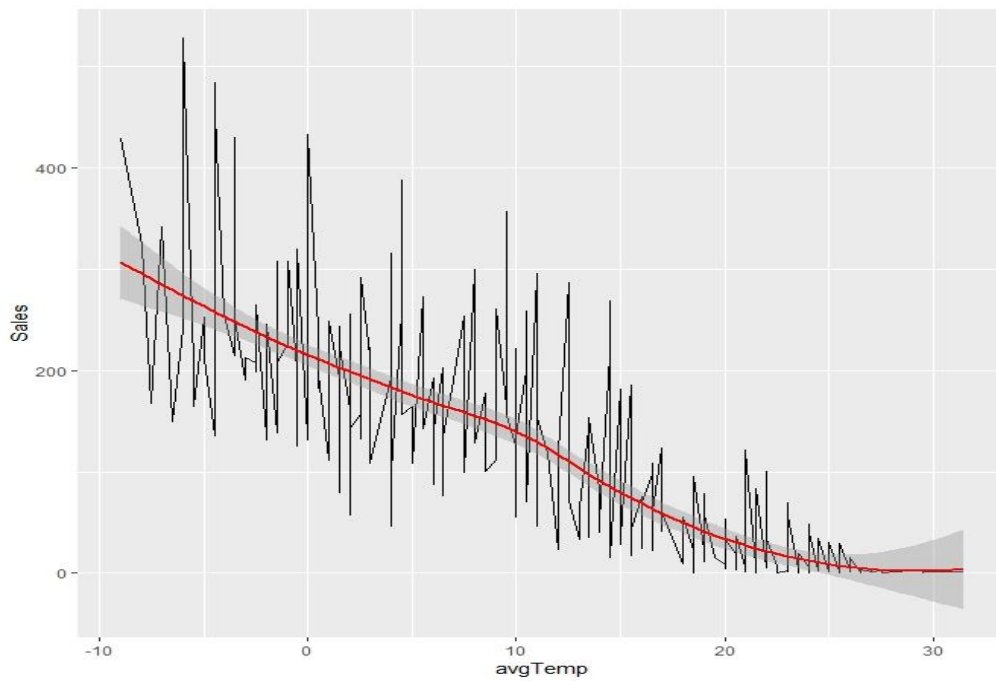


[그림 4.5] 2014년 ~ 2017년도 데이터로 예측한 2018년
온도 변화에 따른 반팔 티셔츠 판매량



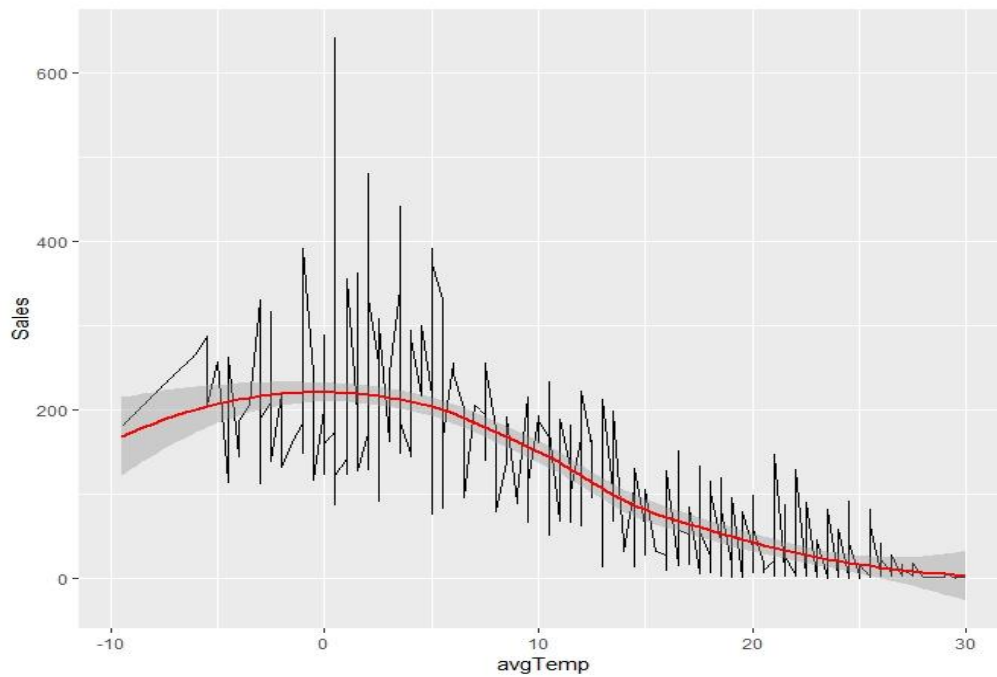
[그림 4.6] 실제로 판매된 2018년도 온도 변화에 따른 반팔 판매량

[그림 4.5]는 2014년부터 2017년까지의 데이터로 예측한 2018년 온도 변화에 따른 반팔 티셔츠 판매량이다. [그림 4.6]은 2018년 실제 판매된 데이터이다. [그림 4.5]와 [그림 4.6]을 통해 분석한 결과 $\pm 1.5\%$ 의 오차가 발생했다. 반팔 티셔츠는 날이 더울 때뿐만 아니라 평소에도 인 웨어로 많은 사람이 활용하기 때문에 꾸준한 판매량을 보이며 0도 이하에서의 판매량이 존재한다. 반팔 티셔츠가 주 상품이 되는 시기인 14도 ~ 16도 이상의 예측 데이터와 실제 판매 데이터 그래프를 보면 일치도가 높다. 기상청 기후 데이터에 따르면 2018년도에 평균 온도 0도 이하는 67번, 15도 이상은 170번, 30도 이상은 25번이다. 지난 4년에 비해 30도 이상의 온도 횟수가 더 많았고 역대 최고 폭염을 기록한 년도였다. 한 달 이상 지속된 폭염 탓에 반팔 티셔츠는 높은 판매량을 기록하였다. 하지만 4년간의 데이터를 바탕으로 수요 예측 분석을 진행 후 미리 많은 물량을 확보할 수 있었고 'A' 업체는 2018년 여름 높은 매출을 기록하였다. 따라서 성공적인 수요 예측을 했다고 판단할 수 있다.



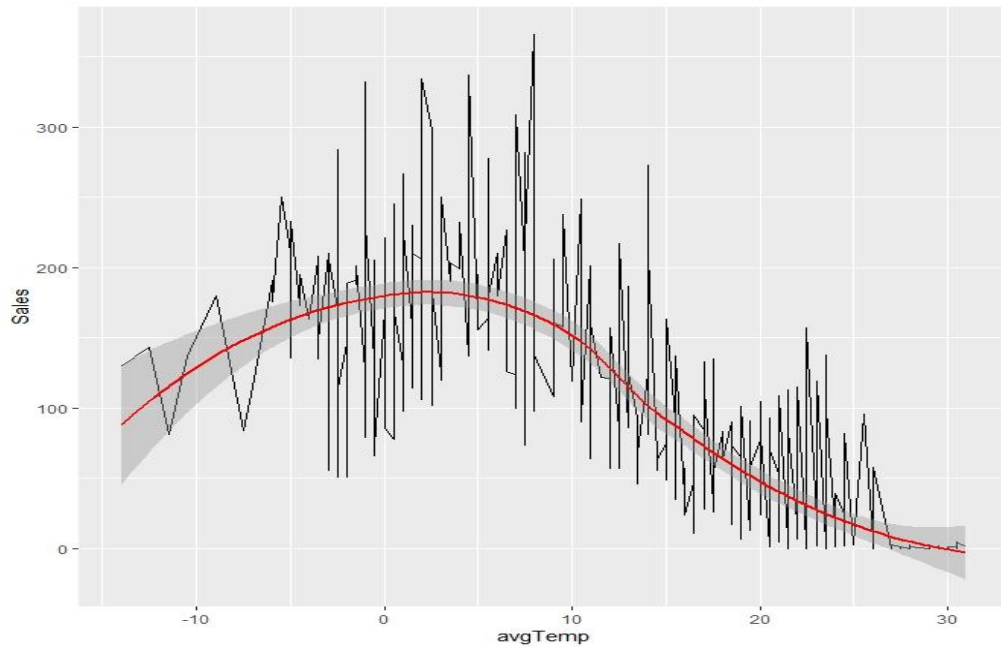
[그림 4.7] 2014년도 온도 변화에 따른 아우터 웨어 판매량

기상청 기후 데이터에 따르면 2014년도에는 평균 온도가 0도 이하는 53번, -5도 이하는 11번, -10도 이하는 0번이다.



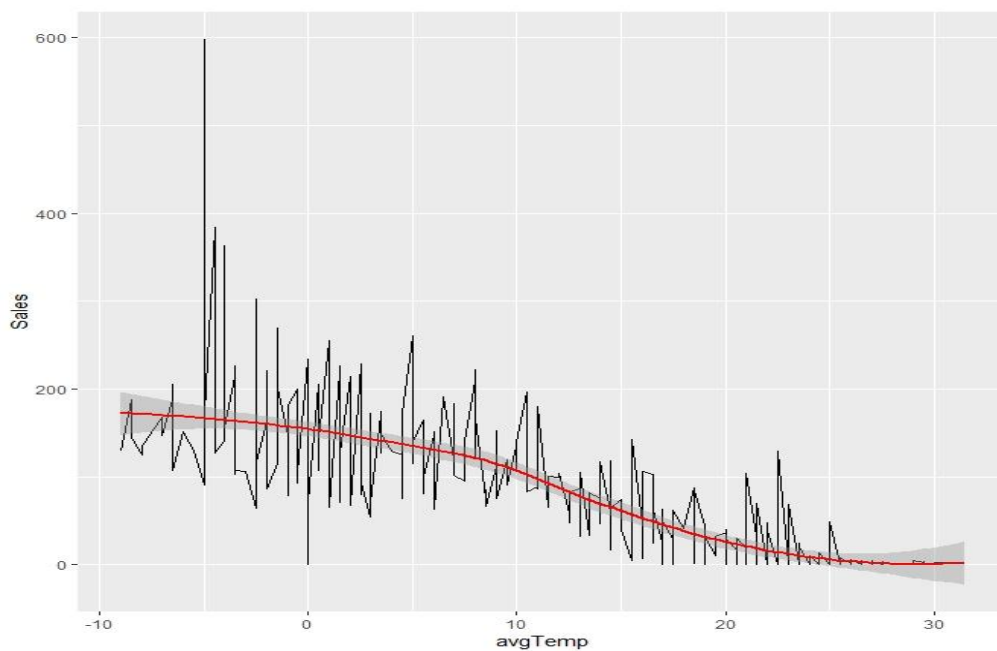
[그림 4.8] 2015년도 온도 변화에 따른 아우터 웨어 판매량

기상청 기후 데이터에 따르면 2015년도에는 평균 온도가 0도 이하는 48번, -5도 이하는 18번, -10도 이하는 0번이다.



[그림 4.9] 2016년도 온도 변화에 따른 아우터 웨어 판매량

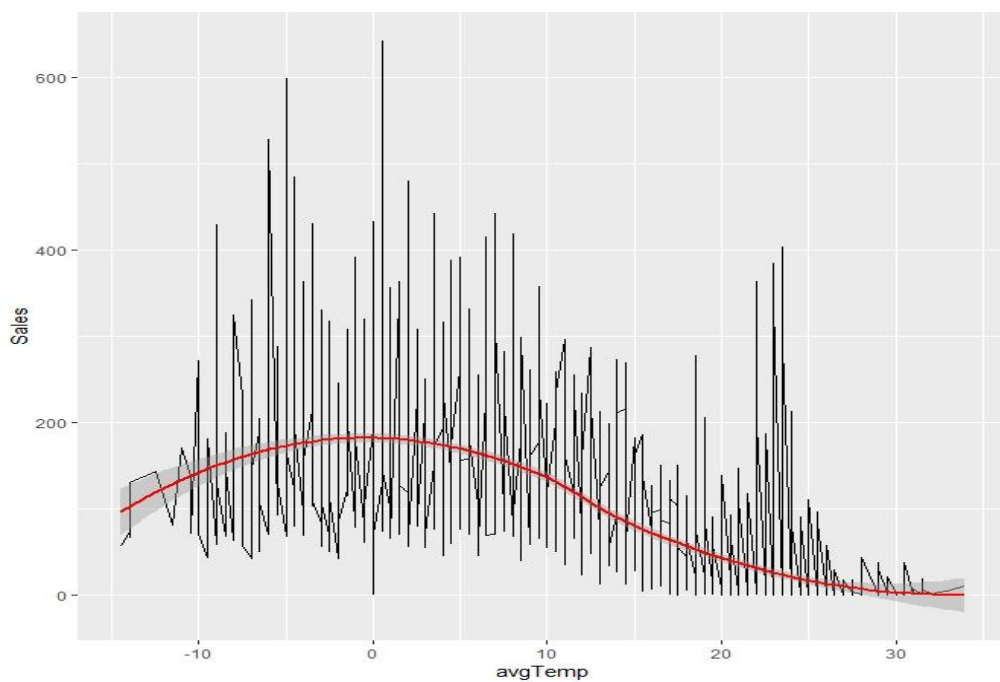
기상청 기후 데이터에 따르면 2016년도에는 평균 온도가 0도 이하는 58번, -5도 이하는 13번, -10도 이하는 3번이다.



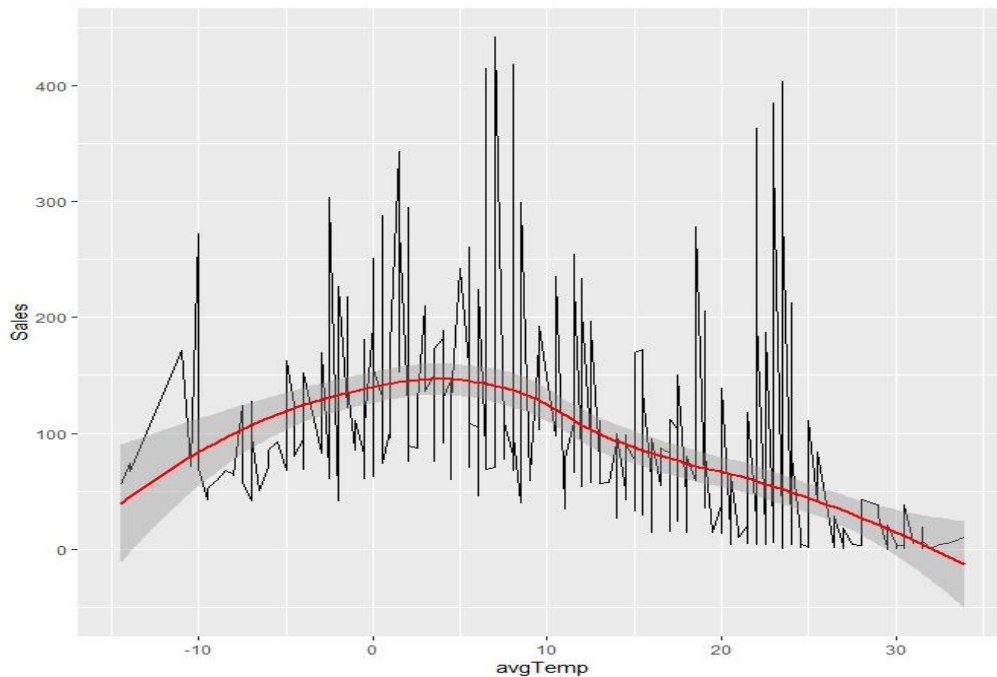
[그림 4.10] 2017년도 온도 변화에 따른 아우터 웨어 판매량

기상청 기후 데이터에 따르면 2017년도에는 평균 온도가 0도 이하는 69번, -5도 이하는 18번, -10도 이하는 0번이다.

평균 온도 변화에 따른 [그림 4.7], [그림 4.8], [그림 4.8], [그림 4.9], [그림 4.10] 4년간의 데이터를 통해 2018년 평균 온도 변화에 따른 아우터 웨어 판매량을 예측 하면 [그림 4.11]과 같다.



[그림 4.11] 2014년 ~ 2017년도 데이터로 예측한 2018년 온도변화에 따른 아우터 웨어 판매량



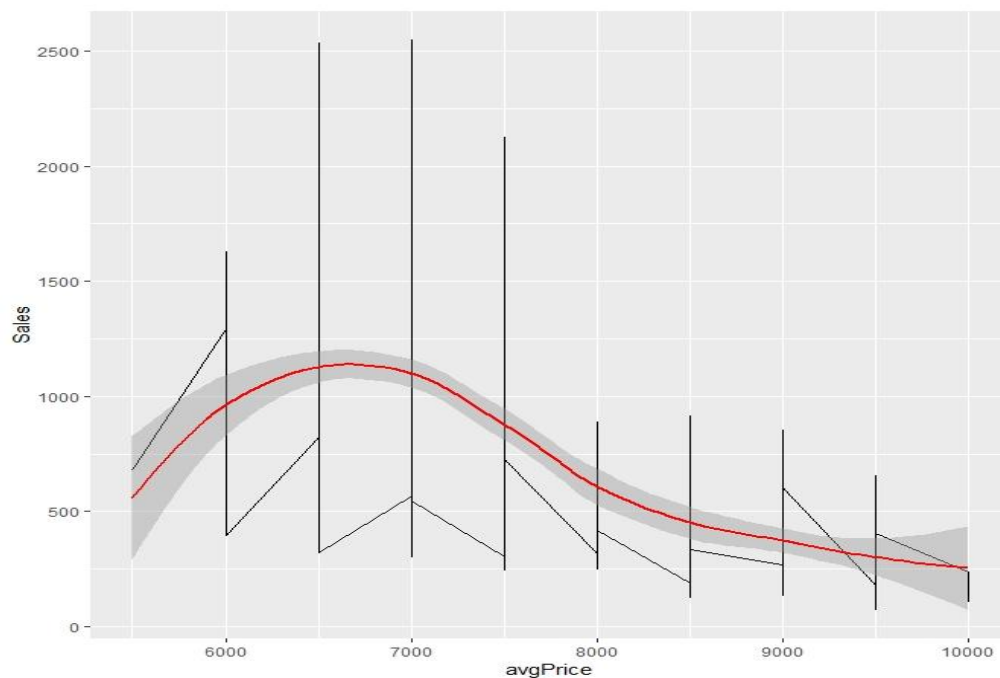
[그림 4.12] 실제로 판매된 2018년도 온도 변화에 따른 아우터 웨어 판매량

[그림 4.11]은 2014년도부터 2017년도까지의 데이터로 예측한 2018년 평균 온도 변화에 따른 아우터 웨어 판매량이다. [그림 4.12]는 2018년 실제 판매된 데이터이다. 아우터 웨어의 경우 겨울에 많이 팔리는 상품이기 때문에 0도 이하의 판매량에 초점을 두어 본다. [그림 4.8] 2015년 판매량 [그림 4.10] 2017년 판매량을 보면 다른 년도에 비해 유독 판매량이 높은걸 알 수 있다. 기상청 기후 데이터에 따르면 2015년과 2017년의 경우 -5도 이하의 온도가 18번으로 다른년도에 비해 추웠기 때문이다. 아우터 웨어가 0도 이상일 때 판매율이 존재하는 이유는 후드 집업, 가디건과 같은 환절기에 입는 상품들이 판매되기 때문이다. 예측 판매 데이터와 실제 판매 데이터에서 $\pm 8\%$ 의 오차가 발생하였다. 그 원인으로는 2016년 10월 29일부터 2017년 3월 10일까지 촛불 집회가 진행되어 아우터 웨어 판매량이 많았다. 수요 예측을 진행할 때 이러한 사회적 이슈를 배제하고 진행하여야 하지만 R 프로그램은 그러지 못하였고 전년도 겨울의 높은 판매량이 그대로 반영되어 2018년도에도 아우

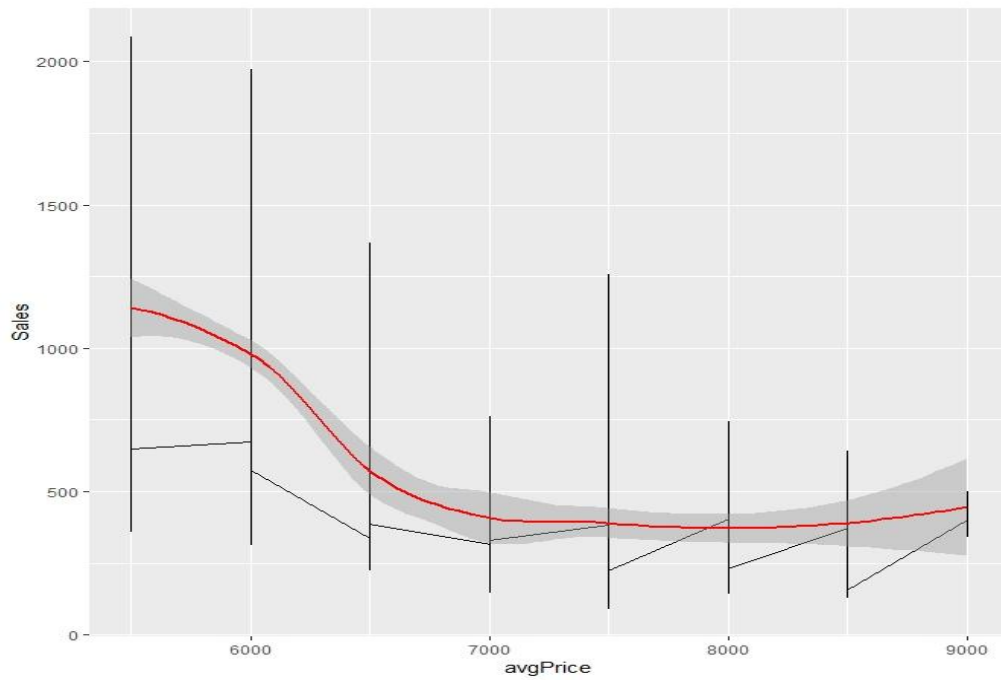
터 웨어가 많이 판매되리라 예측하였다. 하지만 실제 판매량은 예측 판매량보다 저조하였다. 그러나 전년도 판매량만 반영하여 예측을 진행한 것이 아닌 4년 치의 판매량을 바탕으로 예측을 진행하였기 때문에 쇼핑몰에 큰 타격을 주는 오차는 아니었다. 이러한 수요 예측 분석 없이 전년도의 수요만 보고 온라인 쇼핑몰에서 많은 양의 아우터를 판매했다면 엄청난 손해가 있었을 것으로 사료된다.

4.1.2 평균 가격 변화에 따른 판매량 예측

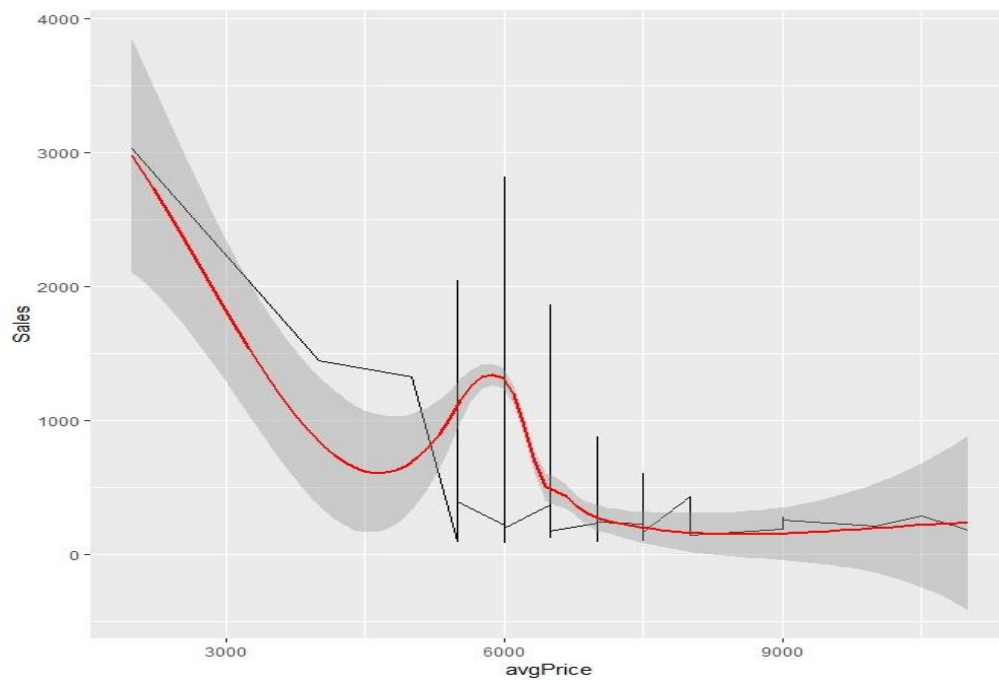
평균 가격 변화에 따른 판매량에 대한 분석이며 반팔 티셔츠, 아우터 웨어(Outer Wear) 2가지 제품에 대한 분석을 다뤘다. 빨간색 그래프는 판매 데이터의 추세선을 나타낸다.



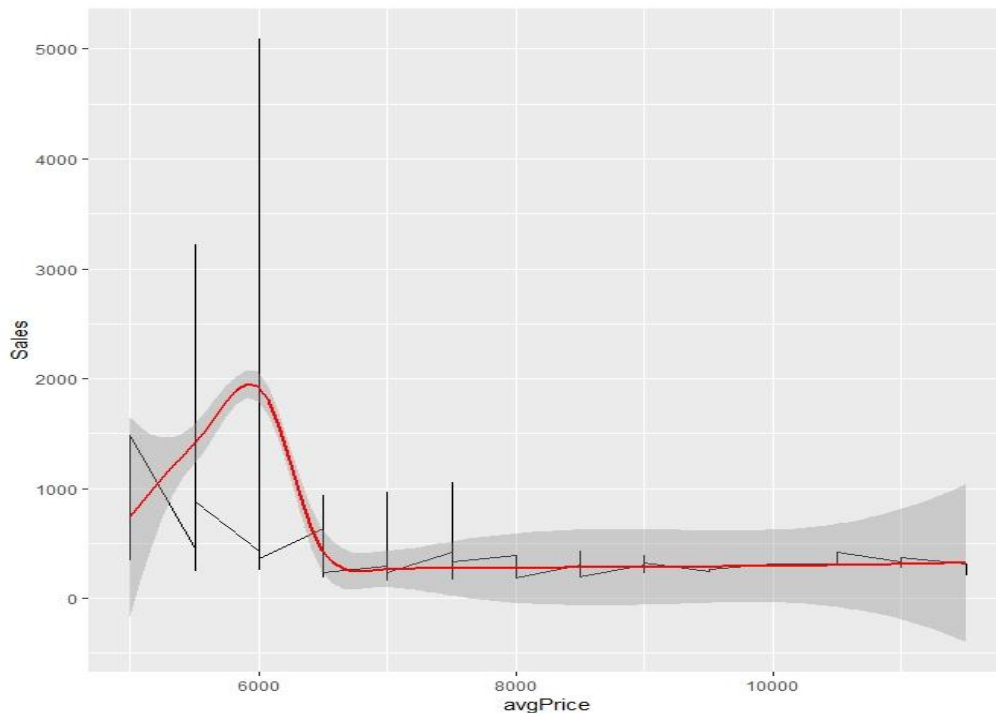
[그림 4.13] 2014년도 가격 변화에 따른 반팔 판매량



[그림 4.14] 2015년도 가격 변화에 따른 반팔 판매량

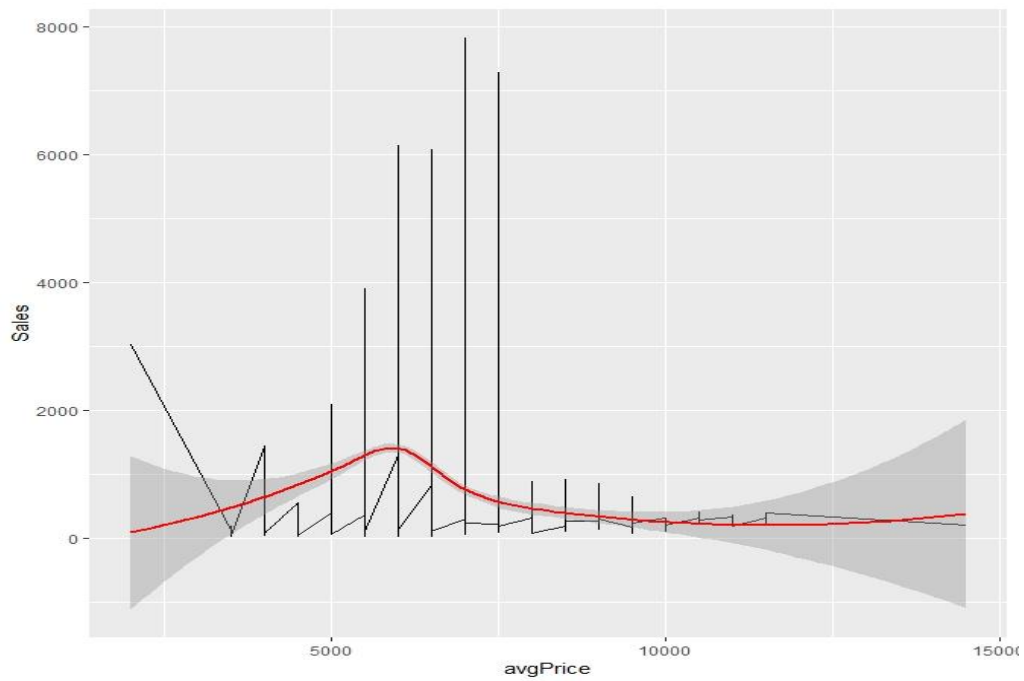


[그림 4.15] 2016년도 가격 변화에 따른 반팔 판매량

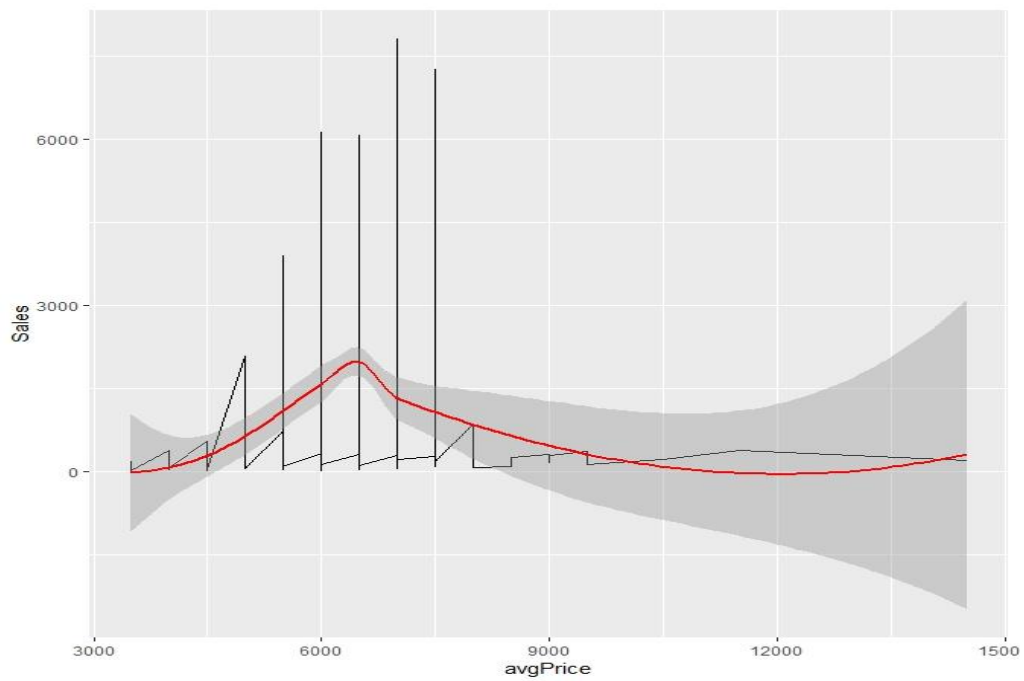


[그림 4.16] 2017년도 가격 변화에 따른 반팔 판매량

평균 가격 변화에 따른 [그림 4.13], [그림 4.14], [그림 4.15], [그림 4.16] 반팔 티셔츠 판매량을 보면 가격에 따른 판매량 변화가 큰 것처럼 보인다. 하지만 자세히 보면 반팔 티셔츠의 경우 5,000원 이하에서의 판매량이 많고 그 이상의 가격에서는 판매량이 낮은 것을 알 수 있다. 반팔 티셔츠의 경우 저렴한 가격으로 인해 소비자가 직관적 소비 성향을 띄운다. 그렇기 때문에 스펙트럼이 좁고 트렌드에 민감하지 않다는 것을 알 수 있다. 4년간의 데이터를 통해 2018년 평균 가격 변화에 따른 반팔 티셔츠 판매량을 예측하면 [그림 4.17]과 같다.

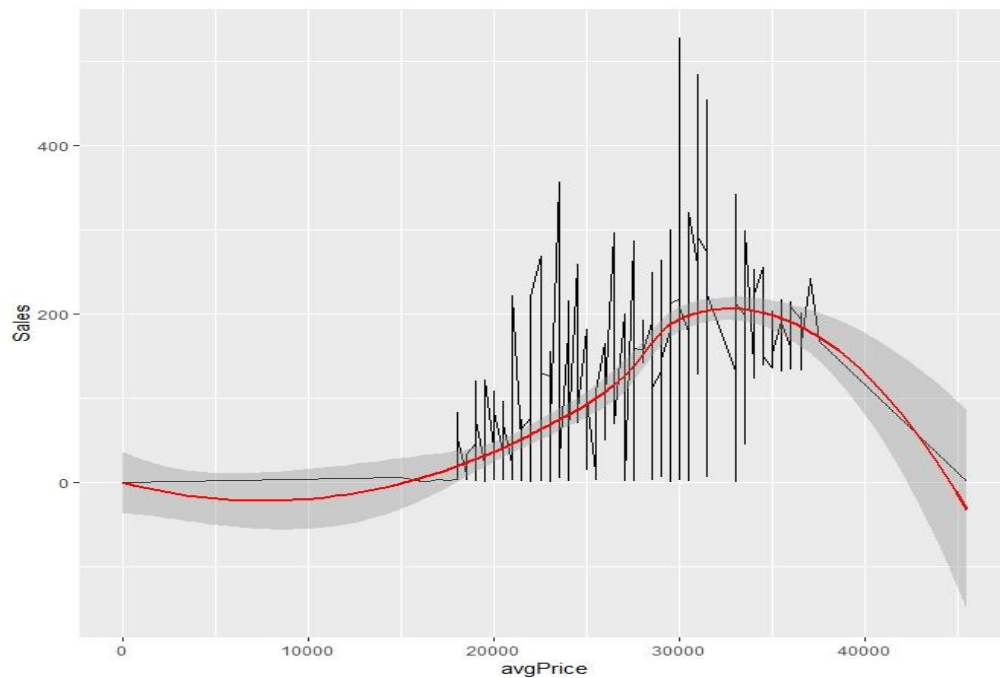


[그림 4.17] 2014년 ~ 2017년도 데이터로 예측한 2018년 가격 변화에 따른 반팔 판매량

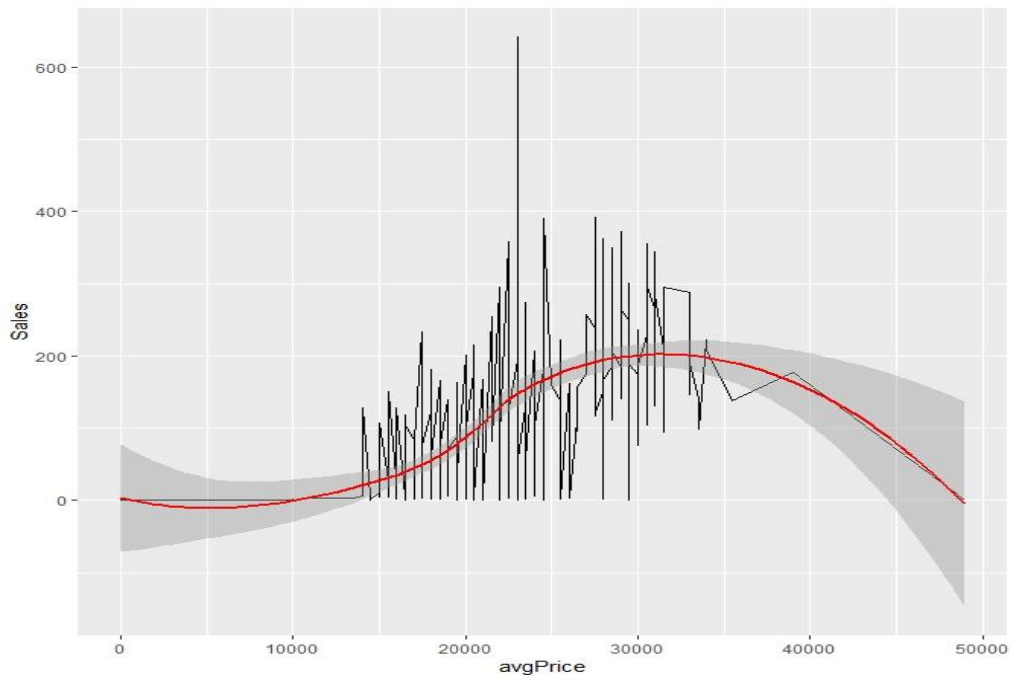


[그림 4.18] 실제로 판매된 2018년도 가격 변화에 따른 반팔 판매량

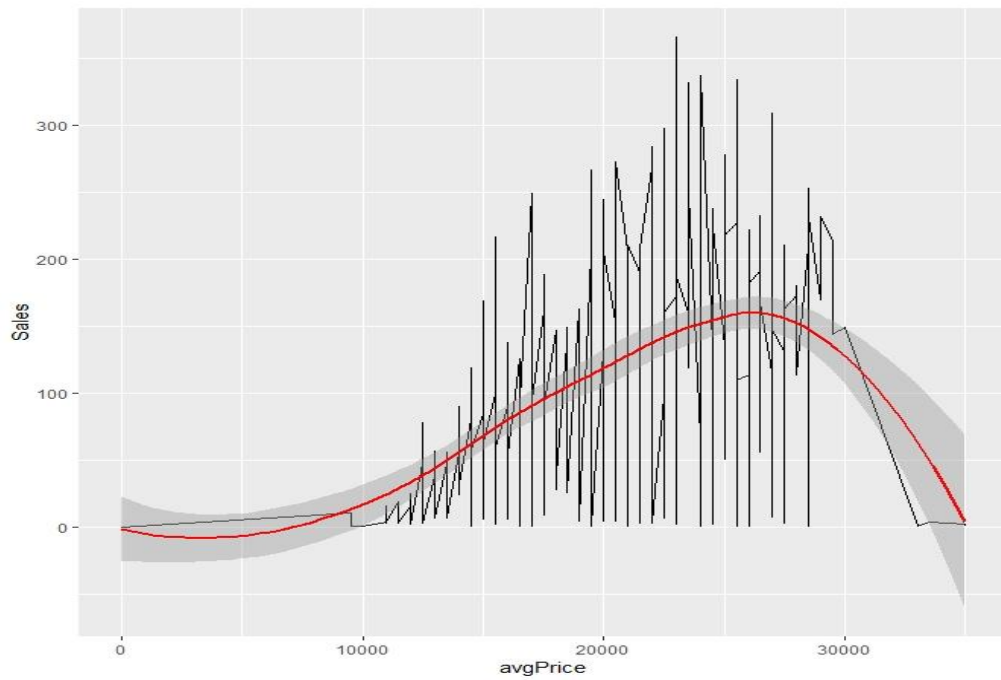
[그림 4.18]은 2014년도부터 2017년도까지의 데이터로 예측한 2018년 평균 가격 변화에 따른 반팔 판매량이다. [그림 4.19]는 2018년 실제 판매된 데이터이다. [그림 4.18]과 [그림 4.19]를 보면 $\pm 4.5\%$ 의 오차로 그래프가 매우 유사함을 알 수 있고 어느 가격대로 반팔을 판매해야 소비자들이 선호하는지도 알 수 있다. 반팔은 날이 더울 때뿐만 아니라 평소에도 인 웨어로 많은 사람이 입기 때문에 저렴한 가격의 상품들이 많이 팔린다.



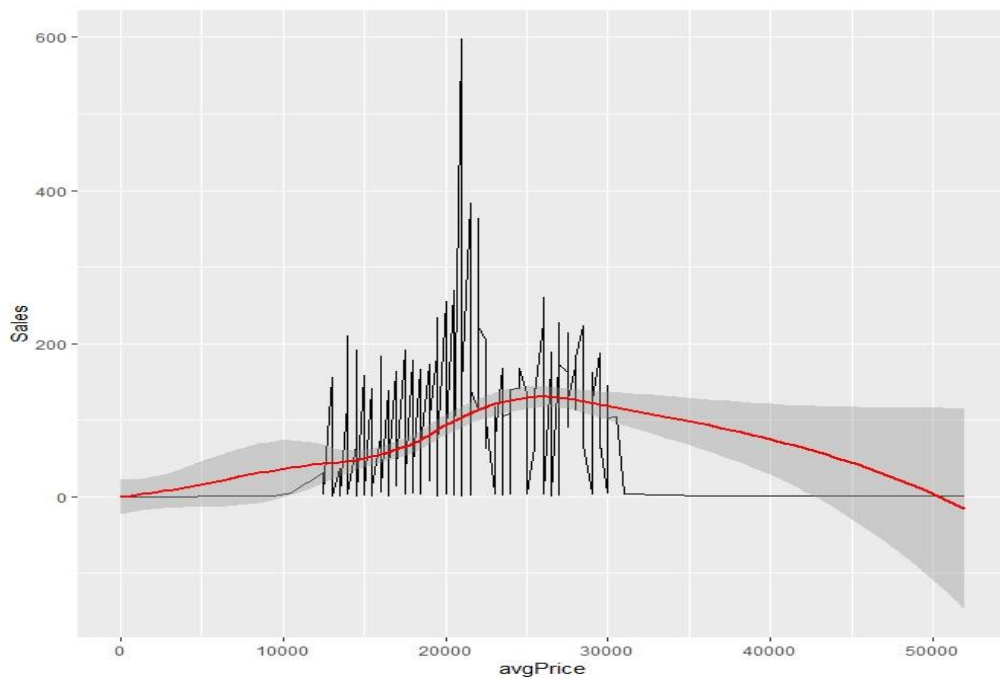
[그림 4.19] 2014년도 가격 변화에 따른 아우터 웨어 판매량



[그림 4.20] 2015년도 가격 변화에 따른 아우터 웨어 판매량

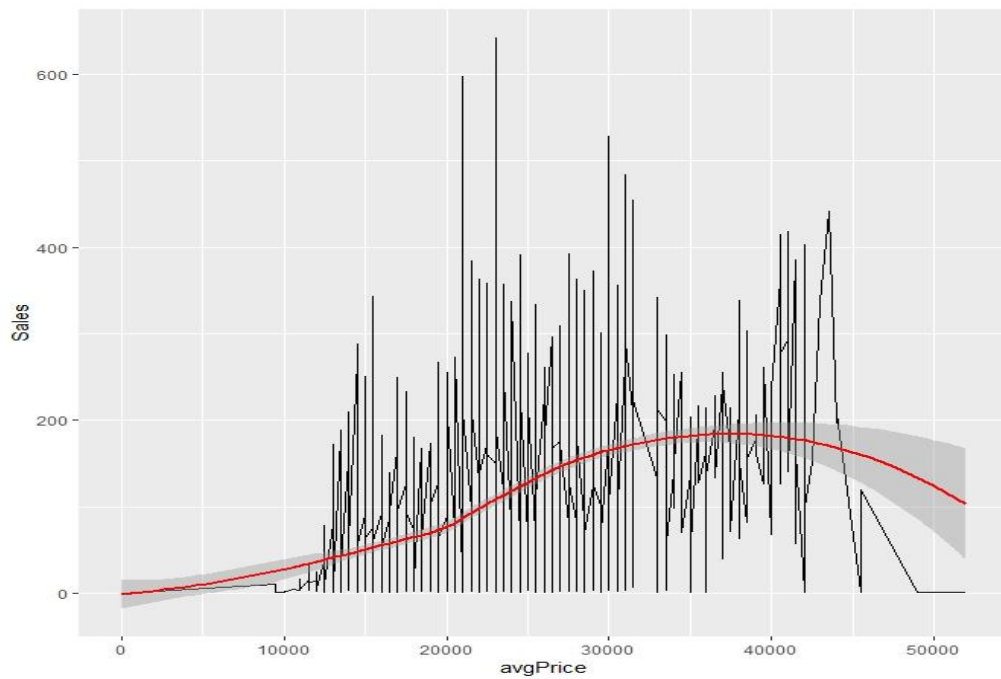


[그림 4.21] 2016년도 가격 변화에 따른 아우터 웨어 판매량

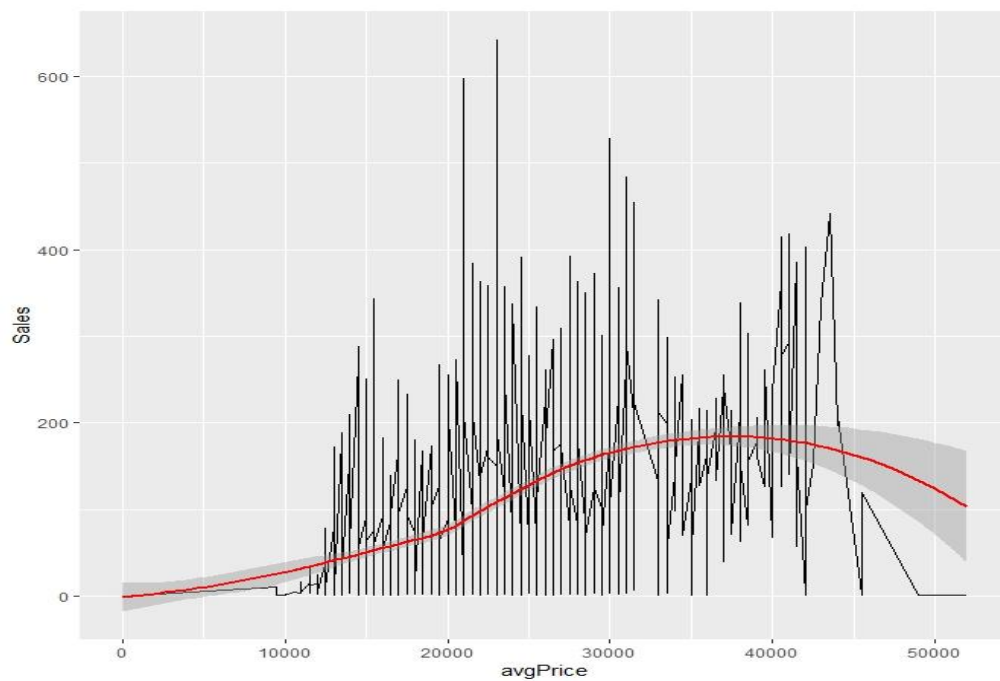


[그림 4.22] 2017년도 가격 변화에 따른 아우터 웨어 판매량

평균 가격 변화에 따른 [그림 4.19], [그림 4.20], [그림 4.21], [그림 4.22] 평균 가격 변화에 따른 2014년부터 2017년까지의 아우터 웨어 판매량을 보면 가격에 따른 판매량 변화가 적다는 것을 알 수 있다. 그만큼 아우터 웨어를 구매 할 때는 소비자들이 신중하게 구매한다는 것을 알 수 있으며 트렌드에 예민하고 스펙트럼의 폭이 좁다. 4년간의 데이터를 통해 2018년 평균 가격 변화에 따른 아우터 웨어 판매량을 예측하면 [그림 4.23]과 같다.



[그림 4.23] 2014년 ~ 2017년도 데이터로 예측한 2018년 가격 변화에 따른 아우터 웨어 판매량



[그림 4.24] 실제로 판매된 2018년도 가격 변화에 따른 아우터 웨어 판매량

[그림 4.23]은 2014년도부터 2017년도까지의 데이터로 예측한 2018년 평균 가격 변화에 따른 아우터 웨어 판매량이다. [그림 4.24]는 2018년 실제 판매된 데이터이다. [그림 4.23]과 [그림 4.24]를 보면 $\pm 1.5\%$ 의 오차로 오차가 매우 작은 편이다. 그만큼 아우터 웨어는 가격의 영향을 많이 받지 않는 상품이며 구매할 때는 소비자들이 신중하게 구매한다는 것을 알 수 있다. 예측 판매 데이터와 실제 판매 데이터가 매우 유사하며 판매량 예측이 성공적이라고 볼 수 있다.

V. 결론

본 논문에서 정의하였듯 빅데이터를 이용할 경우 운영상의 문제점에 대한 대응법과 미래 예측을 통해 실제 유통을 통해 문제가 발생하는 재고와 매출 향상에 실제적인 영향을 줄 수 있음을 알 수 있다.

‘A’ 업체의 제한적인 데이터와 현재의 상황분석을 통해 미래 예측을 하는 단편적인 결과를 낳지만 이후 장기간에 걸쳐 빅데이터가 누적되면 될수록 기간적인 예측에 있어 더 정확한 예측이 가능하며 이 결과로 리스크를 최소한으로 줄일 수 있을 것으로 기대된다.

특히 본 논문에서 다루는 패션 업계에 빅데이터를 활용한다면 패션 업계의 재고 관리, 수요 예측이 매우 정교해지므로 매출 확대와 관리비 절감이 되어 생산성 향상 및 수익 제고에 도움이 될 수 있음을 입증했다.

마지막으로 최초로 온라인 쇼핑몰을 창업하는 청년들에게는 재고 관리를 위한 가이드라인을 제시하였고 해외 진출 기업에도 긍정적인 도움을 줄 수 있을 것이라 사료된다. 또한 향후 후속 연구에 있어서는 평균 온도와 평균 가격 외에 더 다양한 요인들을 추가한다면 더 의미 있는 연구가 될 것이라 기대된다.

참고 문헌

- [1] 구본기, 손은호(2006), “계절 ARIMA 모형을 이용한 항공권판매액 예측”, 대한관광경영학회지 21(1): 81-96.
- [2] 기상 자료 개방 포털 <https://data.kma.go.kr/cmmn/main.do>.
- [3] 김범(2013), “빅데이터 분석 동향”, 한국데이터산업진흥원, 데이터베이스 백서.
- [4] 테이코산업연구소(2010), “글로벌 전자책시장 실태와 관련 산업 동향(급성장하는 MID단말 시장)”, 테이코산업연구소출판부.
- [5] 머니볼(Moneyball) 이론(2012), “에듀윌 정보통신/오늘의 일반상식”, 평생교육 No.1 에듀윌 http://blog.eduwill.net/1079#comment_area.
- [6] 서울뉴스(2013), “서울시 빅데이터로 심야버스 노선 구축” <http://news1kr/articles/1219217>.
- [7] 스킵 리스케, 천채정 역(2012), “빅 데이터 비즈니스”, 서울: 도서출판 더숲.
- [8] 연합뉴스 기획취재팀(2013), “빅데이터 사회”, 연합뉴스(월간마인더스).
- [9] 오광우, 이우리(1995), “시계열예측 방법과 응용. 자유아카데미”.
- [10] 오지연, 백승훈, 이양규, 홍준기, 홍성찬(2019), “빅데이터 분석을 활용한 온라인 판매 수요 예측”, 「춘계학술발표대회」 제20권 제1호, 한국인터넷정보학회.
- [11] 유병준(2018), “한국 온라인 창업 성장 리포트”: 네이버 스마트 스토어 사례분석.
- [12] 윤을요(2017), “국내 패션 업계의 빅데이터 활용에 대한 고찰”.
- [13] 장동인(2014), “빅데이터로 일하는 기술”, 서울: 한빛미디어(주), p.137.
- [14] 장영재(2012), “아마존닷컴, 현대의 서점 아저씨”, 비즈니스북스, pp. 119.
- [15] 정용찬(2012), “빅데이터 혁명과 미디어 정책 이슈”, (KISDI Premium Report 12-02). 정보통신정책 연구원.
- [16] 정우진(2014), “빅데이터를 말하다”, 서울: 클라우드북스, pp.64-65
- [17] 차재필(2012), “빅데이터 시대의 국민공감 선거전략-미 대선사례를 중심으로”, 한국정보화진흥원.
- [18] 최병선(1995), “단변량시계열분석I”, 세경사.
- [19] 최영문(1997), “관광수요예측모형의 예측정확성 평가”, 경기대학교 대학원 박사학위논문.

- [20] 한국경제신문(2017), “온라인 쇼핑몰 창업한 10명중 9명은 유지 못했다.”
<http://plus.hankyung.com/apps/newsinside.view?aid=201701179332A&sns=y>.
- [21] 허준석(2018), “빅데이터를 활용한 패션 브랜드 웹기반 소비자 평가 트렌드 추이 분석”, 국민대학교 대학원 석사 학위논문.
- [22] LG CNS AA센터(2013), “구현방식에 따른 빅데이터 기술 분류”.
- [23] Choy, Dexter, J.L. 1984. Forecasting Tourism Revisited. *Tourism Management*, Sept, pp. 171-176.
- [24] Economist Intelligence Unit(2011). Big Data Harnessing a game-changing asset.
- [25] Gartner, “op 10 Strategic Technology Trends for 2013,” Special Report, 2012,
<http://www.gartner.com/technology/research/top-10-technology-trends/>.
- [26] George, G., Haas, M. R., and Pentland, A. 2014. Big data and management. *Academy of Management Journal* 57: 321-326.
- [27] Kim, J. S. 2013. Big Data Utilization and Analysis Techniques. Master Thesis Dissertation, Korea University, Seoul, Korea.
- [28] Makridakisdh, S. and Hibon, M. 1979. Accuracy of Forecasting: An Empirical Gation. *Journal of Statistical Society, Series A* 142: 97-145
- [29] McAfee, A. and E. Brynjolfsson(2012), “Big Data: The management revolution,” *Harvard Business Review*, 90(10), 60-66.
- [30] McKinsey Global Institute(2011), Big data: TheNext Frontier for Innovation, Competition, and Productivity, Lexington, KY: McKinsey & Company.
- [31] Pauwels, K., D. M. Hanssens, & S. Siddarth(2002), “The long-term effects of price promotions on category incidence, brand choice, and purchase quantity,” *Journal of Marketing Research*, 39(4), 421-439.
- [32] Reid, D.J. 1969. A Comparative Study of Time Series Prediction Techniques on Economic Data, Phd. University of Nottingham.
- [33] Schroeck, M., R. Shockley, J. Smart, D. Romero-Morales, & P. Tufano (2012), *Analytics: The Real-world Use of Big Data*, NY: IBM Institute for Business Value.
- [34] Seo, D. M., and Jung, H. M., “Intelligent Web Crawler for Supporting Big

Data Analysis Services,” Journal of the Korea Contents Association, Vol. 13, No. 12, pp. 575-584, 2013.

- [35] Snijders, C., Matzat, U., & Reips, U.-D. (2012). ‘Big Data’: Big gaps of knowledge in the field of Internet. International Journal of Internet Science, 7, pp.1-5.

Abstract

Prediction of Online Sales Demand Using Big Data Analysis

JI-YEON OH

Department of Information &
Telecommunication
Graduate School of
Hanshin University

Advisor : Professor.

SUNG-CHAN HONG

In line with the era of the Fourth Industrial Revolution, companies are in a business environment where they can accumulate and use more data than ever before. In a variety of business environments, there is an immeasurable amount of data accumulation in real time, which is called big data. With the recent development of data extraction technology and analysis programs, a growing number of companies are looking to leverage big data in strategic decision-making. Big data analytics can generate new information that was not available with traditional data collection methods, such as inventory management, demand prediction,

production, pricing, market flow identification, and consumer awareness and trend analysis. In the fashion industry handled by this paper, more and more companies are trying to utilize web-based big data analysis and extraction in product planning, customer management and sales. However, there have been only a handful of studies related to this so far.

Taking this into account, this study proposes the Big data Analysis for Sales Prediction Platform (BAPP), an online sales demand prediction model using big data for efficient inventory management and increased sales by analyzing sales by product based on real-time cumulative data from online shopping mall 'A' companies over the past five years. In particular, this study used the R program to predict 2018 sales based on sales from 2014 to 2017 and compared and analyzed forecast and actual sales. According to the comparative analysis, the short-sleeved T-shirts were sold steadily not only when it was hot, but also because many people used them as in wear, resulting in only ± 1.5 percent error in forecast and actual sales. For Outer Ware, there was a $\pm 8\%$ error in forecast and actual sales. As a cause, candlelight vigils were held from Oct. 29, 2016 to March 10, 2017, with the sales of autor ware soaring momentarily. On the other hand, while forecasting demand, we should exclude these social issues, but the R program failed to do so and predicted that the high sales volume for the winter of the previous year would be reflected and that many Outer Ware would be sold in 2018. However, actual sales were lower than forecast sales. but, the forecast was not a big blow to the shopping mall as it was based on four years' worth of sales, not just on the previous year's sales. Moreover, the analysis of consumers' preferred price range and the sale was carried out in accordance with the price range, resulting in a error of $\pm 1.5\%$ for the

half-sale T-shirt with a error of $\pm 4.5\%$. It can be judged that the error was small enough to make a successful demand forecast.

In summing up these results, it is believed that this study will provide positive assistance to a wide range of companies, including the fashion and restaurant industries that sell goods, and to companies that make inroads into overseas markets.

Keyword : Big Data, Demand prediction, R program, Online shopping malls, Fashion, Clothing