

# Apriori 알고리즘 기반의 중소 베이커리 기업의 대응 전략

김도훈<sup>1</sup>, 이현준<sup>2</sup>, 이봉규<sup>3\*</sup>

<sup>1</sup>연세대학교 기술경영학협동과정 학생, <sup>2</sup>연세대학교 정보대학원 학생, <sup>3</sup>연세대학교 정보대학원 교수

## SME Bakery's Marketing Strategies Based on Apriori Algorithm

Do Hoon Kim<sup>1</sup>, Hyeon June Lee<sup>2</sup>, Bong Gyou Lee<sup>3\*</sup>

<sup>1</sup>Student, Management of Technology, Yonsei University

<sup>2</sup>Student, Graduate School of Information, Yonsei University

<sup>3</sup>Professor, Graduate School of Information, Yonsei University

**요약** COVID-19에 따른 온라인 마케팅의 활성화는 디지털 트랜스포메이션 가속화를 촉진하고 있다. 본 연구는 디지털 트랜스포메이션 역량이 부족하고 COVID-19로 수입 감소를 겪고 있는 중소 베이커리 기업의 10년간 매출 데이터로 Apriori 알고리즘을 사용하여 연관규칙분석을 수행했다. 분석 결과 오프라인 마켓 중심에서 온·오프라인 B2B, B2C 시장으로 전환하고, 다품종 소량 판매에서 소품종 대량판매 전략을 취하는 것이 경영 개선을 할 수 있는 것으로 나타났다. 향후 다각화된 마케팅 전략에 따른 다양한 채널의 판매 데이터를 분석하고 학습하면 많은 중소기업의 디지털 전략 대응을 위한 가이드라인이 될 수 있을 것이다.

**주제어** : 중소 베이커리 기업, 연관규칙, 매출 데이터, 데이터베이스 마케팅, 디지털 트랜스포메이션, 판매 전략

**Abstract** The importance of online marketing is emerging due to the prevalence of COVID-19. In order to respond to the changing business environment, we have collected ten years of sales data of SME bakery company that have experienced a decrease in sales due to the COVID-19. As a result of the analysis, we found that switching from offline markets to omnichannel B2B and B2C markets and taking 'small quantity batch production' to 'mass production in a small variety' can improve management. This study presented online and offline marketing strategies through data analysis of small and medium-sized bakery companies, which have relatively insufficient digital capabilities compared to large companies, and could be a guideline for many SMEs.

**Key Words** : SME Bakery, Association Rules, Sales Data, Database Marketing, Digital Transformation, Sales Strategy

\*This study was revised based on the author's(DoHoon Kim) thesis for Master's Degree in Business Administration at Yonsei University

\*Corresponding Author : Bong Gyou Lee(bglee@yonsei.ac.kr)

Received January 1, 2022  
Accepted April 20, 2022

Revised April 2, 2022  
Published April 28, 2022

## 1. 서론

### 1.1 연구배경

ICT(정보통신기술)의 발전은 많은 양의 다양한 형태의 데이터가 실시간으로 오고 가는 디지털 시대로의 발전을 도모하고 있다. 이와 맞물려 COVID-19(코로나바이러스감염증)의 범유행은 우리 사회 전 영역에 큰 영향을 주어 일상의 변화를 촉진 시키고 있다. 또한 'MZ'세대(20~30대)가 주력 소비층으로 자리매김하면서 기업은 마케팅 방식과 판매채널 등에서 다양화 변화를 시도하고 있다. 급격하게 변화하는 환경 속에서 인적, 물적 자원이 풍부한 대기업은 고객의 요구와 니즈를 충족시키기 위해 축적된 데이터 활용으로 대응하고 있지만, 역량이 충분하지 못한 소상공인을 포함한 중소기업은 이를 수행하는 데에 어려움을 겪고 있으며 도태되고 있다[1].

중소벤처기업부에서 2021년 발표한 소상공인 실태 조사 결과에서는 전자상거래 실적이 있는 사업체는 전년 대비 2% 증가한 8.9%를 기록하였으며 변화하는 시대에 부합하고 어려운 경제 상황 속 수익증대를 위해 스마트오더, 키오스크, 셀프계산대, 디지털 플랫폼 활용 등을 도입하였다. 그러나 도소매업과 숙박·음식점업을 제외한 산업 분야에서는 실적증감을 경험하였다고 응답했으나 그 비중은 낮은 수준이었으며, 타 산업 분야에서는 실적증감을 경험하지 못했거나 감소했다고 응답했다[2].

본 연구에서는 중소기업의 데이터 기반 경영 의사결정의 가능성을 알아보기 위해 중소 베이커리 기업 'A'의 실증 데이터를 분석하였다. 국내 베이커리 산업은 대형 프랜차이즈가 해당 산업의 대부분을 차지하고 있으며 [3], 디지털 트랜스포메이션을 통하여 수집한 데이터를 그간 축적된 데이터와 결합하여 경영 혁신을 도모하고 있는 대표적인 사례이다. 대형 프랜차이즈 베이커리 기업 중 하나인 SPC 그룹은 오프라인 점포에 온라인 서비스를 다면적으로 접목하는 ICT 전략을 추진하고 있으며 SPC 그룹의 자체 플랫폼 개발을 통해 매출과 주문량에서 가파른 성장세를 보인다. 소상공인을 포함한 중소 베이커리 기업은 이에 대응하기 위해 디지털 플랫폼을 통해 판매채널 다양화를 시도하고 있지만, 플랫폼 사업자의 수수료 및 과한 경쟁으로 많은 어려움을 호소하고 있다.

Boston Consulting Group의 주장과 같이 혁신과

기업 성장은 무분별한 디지털 기술의 활용이 아닌 넓은 시야와 전략관점에서 시작되어야 한다[4]. 따라서 본 연구에서는 변화하는 시대에 빠르게 대응하는 중소기업의 대응 방안을 탐색하기 위해 대형 프랜차이즈 기업이 한 산업 분야 대부분을 차지하고 있는 베이커리 산업 분야에 속한 'A' 중소기업의 10년간의 매출 데이터에 연관규칙 분석을 시행하여 경영 개선 가능성을 알아보고 더 나아가 다른 산업 분야에서도 데이터 기반의 의사결정이 위기 상황 타개와 기업 경영 개선에 효과적인 수단이 될 수 있다는 함의를 줄 수 있기에 본 연구가 필요하다고 할 수 있다.

## 2. 이론적 배경

### 2.1 데이터 기반 의사결정

ICT 기술과 대응량의 데이터 처리기술이 발전하면서 데이터 기반 전략 수립은 많은 사회 분야에서 이루어지고 있다[5]. 의사결정 과정 중 기존의 축적된 데이터 마이닝을 통해 쉽게 나타나지 않았던 유용한 정보들을 경영, 정책 수립 등에 활용할 수 있으며 고객들의 요구에 즉각적으로 대응하고 차별화를 위해 노력할 수 있다[6]. 일례로 데이터 기반 분석을 통해 맞춤형 교육 서비스를 제공하여 차별화 전략을 제시하기도 하고[7] 미디어 산업에서의 적용으로 대규모 영화 시장의 흥행 실적을 예측하기도 한다[8]. 마케팅 전략 분석을 통해 마케팅 트렌드 변화를 관찰하여 마케팅 전략 수립에 시사점을 제시하기도 한다[9].

데이터 기반의 의사결정은 기업 경영에 있어서 중요한 기업 경영에 있어 데이터 기반 전략 수립은 진행 중이며 그 중요성은 이전부터 강조되었다. 이준석 등(2015)의 연구는 무인 자동차 관련 특허문서 데이터 마이닝을 통해 개발현황을 분석하고 기술예측을 시행해 기술경영 전략 수립에 대한 시사점을 제시하였고[10] 이유재 등(2014)의 연구에서는 1개 기업의 마케팅 활동에 대한 분석과 데이터 활용 전략 분석을 통해 데이터 활용 마케팅 전략 수립 활용 가능성을 제시하였다 [11]. 안경찬 등(2012)의 연구는 구매 데이터 분석으로 기업 경영 의사결정을 위한 정보를 제공하여 빠른 대응을 위한 시스템을 개발하였다[12]. 조윤희(2020)의 연구는 소셜 데이터 분석과 의미 네트워크 분석을 통해 호텔 이용자의 인식을 탐색하고 서비스 마케팅 촉진 방

안을 제시하였고[13] A. Dagvadorj 등(2020)의 연구는 데이터 마이닝을 이용하여 베이커리 키워드 데이터를 분석하여 소비자의 인식을 분석하고 베이커리 마케팅 전략을 위한 시사점을 제시하였다[14].

## 2.2 중소기업 경영환경

우리나라의 경제성장은 대기업과 중소기업의 수직적 구조로 승자가 독식하는 문제가 지속적으로 발생해왔으며[15], 이는 오늘날 대기업 및 대형 프랜차이즈 업체가 보유한 인력과 자원으로 그간 축적된 많은 양의 데이터를 활용하여 데이터 기반 경영 의사결정을 하면서 중소기업과의 격차가 벌어지고 있다. 기업이 빅데이터를 의미있게 활용하기 위해서는 위와 같이 인적자원과 자본이 뒷받침될수록 원활하게 이루어지지만 그렇지 못한 중소기업은 부족한 자원과 가속화되는 경쟁과 혁신의 부재로 도태되고 있다[16].

디지털 기술의 발전, 판매채널의 다양화, COVID-19로 인한 비대면 소비문화가 정착하면서 기업에서는 판단하고 분석해야 하는 데이터가 다양해지고 증가하고 있다[17]. 고객들의 다양하고 빠르게 변화하는 요구에 맞추어 대응하고 서비스 및 제품을 개발하는 내적인 역량 강화는 기업의 수익증대를 가져올 수 있으며 중소기업은 중견기업, 대기업으로 나아갈 기회로 작용할 수 있다[18]. 4차 산업혁명 기술의 활용과 디지털 트랜스포메이션을 통해 기업이 데이터베이스를 구축하고 이를 활용하여 기존의 비즈니스 프로세스를 혁신하려는 움직임으로 나타나고 있다[19].

이와 같은 현상은 본 연구의 대상이 된 'A' 기업이 속한 베이커리 산업 분야에서도 확인할 수 있다. 대표적인 대형 베이커리 프랜차이즈 기업인 SPC 그룹의 '해피오더'가 있으며 그 외 관련 산업 분야인 스타벅스의 '사이렌오더' 및 투썸플레이스의 '투썸오더' 등이 있다[20]. 대형 프랜차이즈 기업은 기존의 시장에 디지털 기술을 접목하여 많은 형태의 데이터를 축적하고 이를 활용하여 데이터 기반 경영 의사결정을 주도하고 있다.

COVID-19 대유행 이전만 하더라도, 이러한 흐름은 대형 프랜차이즈 기업 위주로 진행되었다[21]. 중소 베이커리 기업은 디지털 기술을 활용하여 데이터를 수집하거나 보유한 데이터를 활용하는 것에 소극적이었으나 COVID-19 대유행은 베이커리 시장의 판매방식에도 영향을 주어 대형 프랜차이즈 기업뿐만 아니라 중소

베이커리 기업의 디지털화와 데이터 기반 의사결정의 중요성을 부각하고 있다[21].

## 2.3 연관규칙분석

연관규칙분석은 Agrawal에 의해 소개된 데이터 마이닝 기법으로 마케팅 전략, 수익 증가 등 기업 경영에 의사결정을 위한 수단으로 활용하고 있다[22]. 연관규칙 분석은 기업이나 기관이 소유하고 있는 데이터 세트로부터 유의미한 패턴, 관계 및 결과를 도출하고자 할 때 유용하게 사용될 수 있으며 미래 고객 패턴 예측에도 중요한 인사이트를 얻을 수 있다[23, 24]. 예를 들어 매장에 방문하는 고객은 자신의 취향에 따라 다른 제품들을 구매하게 된다. 해당 매장은 축적된 데이터를 활용하여 특정 제품을 구매하는 고객, 어떤 제품이 같이 구매되는지, 무슨 이유로 구매하는지, 이에 적합한 마케팅이 어떤 것이 있는지에 대한 정보를 제공하고 고객 경험 증진에 도움을 주는 분석 방법이라 설명할 수 있다[25]. 또한, 연관규칙 분석은 데이터를 기반으로 특정 패턴을 찾는 머신러닝 기법으로써 기업 경영 외에도 다른 분야에 응용할 수 있다는 장점이 있는 분석 기법이다[26].

연관규칙분석에서 도출된 결과를 평가하는 지표는 지지도(support), 신뢰도(confidence), 그리고 향상도(lift)에 근거하며 도출된 결과의 관련성을 찾는다[27]. 연관규칙분석을 수행하기 위한 가장 대표적인 알고리즘으로는 Apriori 알고리즘이 있다. Apriori 알고리즘은 빈도수가 높은 항목을 발견하고 각각의 부분집합을 구하는 방법이다[28].

앞서 언급한 바와 같이 연관규칙분석은 소규모 자영업자, 기업, 기관 등의 다양한 문제를 예측하기 위한 연구에 활용되었다. 연관규칙 분석기법을 사용하여 농촌 진흥청에서 제공하는 소비자들의 구매 정보를 활용하여 소비자들의 유의미한 계절별 식품구매 패턴을 발견하였으며 계절별 매대 구성 및 프로모션 방법을 제안하였다[29]. 연관규칙 분석을 통해 고객들의 충성도와 구매패턴, 간격을 발견하고 중소기업들이 이에 걸맞은 고객 서비스를 제공할 수 있음을 시사했다[30]. 항공산업 데이터를 활용하여 연관규칙 기법을 시행하고 항공사에서 더 신경을 써야 하는 고객층들을 도출하였다[31].

### 3. 연구 방법

#### 3.1 데이터

본 연구의 대상이 되는 ‘A’ 베이커리 기업은 B2B(Business to Business) 기반의 중소기업으로 B2C(Business to Customer) 영역으로 확장을 위해 온라인 쇼핑몰을 운영하고 있다. 해당 기업은 총 58개의 베이커리 제품을 판매하고 있다. 연구를 수행하기에 앞서 온라인 쇼핑몰 창립 시점인 2011년 7월부터 2021년 6월까지의 총 4,675건의 Table 1의 정보가 포함된 판매 데이터가 수집되었으며 분석에 사용되었다.

Table 1. Sales Data

Period	Total Data
2011.07-2021.06	4,675
Collected Data	
Order	order number
	date
	type
	product code
	number of products
Payment	delivery
	total purchase
	payment method
Personal Information	redeem point
	name
	address
	number of visits
	connection path
	number of purchase
	etc.

#### 3.2 분석 방법

본 연구는 Fig. 1.과 같이 진행된다. 수집된 데이터의 경우 2011년 온라인 쇼핑몰 개설 이후 제품명과 카테고리가 혼재되어 있어 연관규칙 분석을 바로 수행하기에 어려움이 있다. 데이터 분석을 진행하기 위해 Python 3와 Pandas, Numpy 패키지를 사용한다. 데이터 분석에 사용되지 않는 정보들을 제거하고 2011년 7월부터 2021년 6월까지의 약 10년간 매출 데이터를 분석 가능한 형태로 정제한다. 데이터 전처리를 위해 판매가 중단된 제품이 포함된 거래를 제거하였으며, 동일한 제품들이 제품명이 아닌 제품번호로만 입력되어있는 값을 분석 가능한 형태로 전처리하였으며

이후 연관규칙분석(Association Rules)을 수행하기 위해 Apriori 알고리즘을 사용한다. 마지막으로 도출된 결과의 시각화 작업을 위해 Matplotlib 패키지를 사용한다.

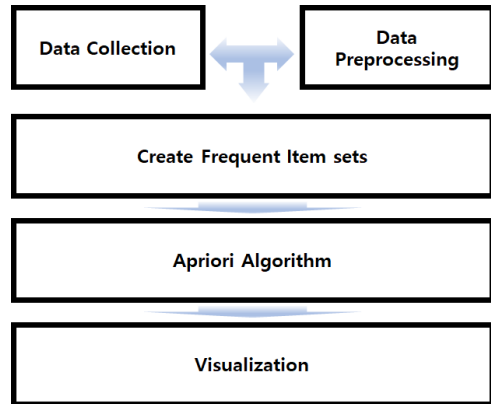


Fig. 1. Research Design

### 4. 연구 결과

#### 4.1 분석 기준

‘A’ 베이커리 기업의 매출 데이터를 Apriori 알고리즘을 사용해 연관규칙 분석을 시행하였고 도출된 결과의 유용성을 확인하기 위한 척도로 Bootstrap-based method[32], 회귀분석모형[33], 지지도-신뢰도 산포 그래프[34]등 다양한 방법이 있지만 본 연구에서는 지지도, 신뢰도, 향상도 기반으로 도출된 연관규칙의 유용성을 확인하고자 한다. 지지도는 상품의 구매가 발생할 확률을 나타낸 것으로 전체 거래 중 연관규칙이 발생한 거래의 확률로 나타낸다. 지지도가 높다는 것은 상대적으로 해당 규칙이 적용된 거래의 발생확률이 높다는 것을 의미하며 이는 연관규칙의 유용성을 파악할 수 있다. 신뢰도는 상품 A를 구매한 거래 중, 상품 A와 B가 동시에 나타나는 비율을 나타내는 것으로 지지도의 단점을 보완하여 연관규칙의 강도를 파악할 수 있다. 향상도는 규칙들의 상관관계를 파악하기 위한 척도로 향상도가 1이면 상품 A, B가 독립되어 있음을 뜻하여 서로 상관관계가 없음을 나타내고, 향상도가 0보다 크고 1보다 작으면 음의 상관관계를 나타낸다. 향상도가 1보다 크면 상품 A의 구매가 상품 B의 구매로 이어지는 양의 상관관계가 있음을 의미한다.

Apriori 알고리즘에서 유의미한 결과를 도출하기 위

해서는 최소 지지도, 최소 신뢰도, 최소 향상도를 설정해야 한다. 최소지지도를 낮게 설정한다면 유의미하지 않은 연관규칙까지 포함하여 유의미한 규칙을 찾기 어렵고, 높게 설정한다면 유의미한 규칙이 포함되지 않을 가능성을 가지게 된다. 안민욱 등(2018)의 연구에서는 최소 지지도를 2%와 최소 신뢰도 20%, 향상도 1을 초과하는 연관규칙을 추출하였고[35] 오승식 등(2019)의 연구는 유의미한 규칙을 도출하기 위해 최소 지지도 1%, 최소 신뢰도 50%, 향상도 1 이상의 규칙들을 설정하였다[36]. 원은송 등(2019)의 연구는 유의미한 연관규칙을 추출하기 위해 최소 신뢰도 20%, 최소 지지도는 10%로 설정하고 최소 향상도가 1을 초과하는 규칙만 선별하였다[29]. 본 연구에서는 기존 연구들보다 높은 최소 지지도 10% 이상, 최소 신뢰도 85% 이상, 향상도 1 이상으로 설정하여 높은 지지도, 신뢰도를 나타내는 연관규칙 분석 결과를 도출하였다.

본 연구에서는 연관규칙 분석 결과 중 유의미한 항목을 추출하기 위해 지지도 10% 이상, 신뢰도 50% 이상, 최소 향상도 1 이상인 연관규칙분석을 수행하여 지지도 순으로 상위 10개 규칙을 정렬하였고 Table 2.와 같다.

Table 2. Results of association rules analysis

#	Pattern		Support	Confidence	Lift
	Product A	Product B			
1	Scone	→ Muffin	0.342	0.845	1.178
2	Plain Cheesecake	→ Muffin	0.308	0.937	1.306
3	Choco Cheesecake	→ Muffin	0.242	0.904	1.260
4	Blueberry Cheesecake	→ Muffin	0.233	0.908	1.265
5	Plain Cheesecake	→ Scone	0.179	0.547	1.349
6	Plain Cheesecake, Scone	→ Muffin	0.164	0.916	1.277
7	Blueberry Cheesecake	→ Scone	0.149	0.579	1.429
8	Choco Cheesecake	→ Scone	0.144	0.535	1.321
9	Choco Cheesecake, Scone	→ Muffin	0.129	0.896	1.248
10	Blueberry Cheesecake, Scone	→ Muffin	0.128	0.863	1.203

#### 4.2 연관규칙 분석 결과

연관규칙 분석은 그 척도에 따라 다양한 결과가 나

타나게 된다. Table 2.와 같은 지지도 순 정렬은 전체 거래 중 연관규칙이 발생한 거래의 확률이 높은 순서를 알 수 있다. 가장 높은 지지도인 0.342는 'Scone, Muffin'의 구매이며 'Scone'을 구매한 고객 중 0.845는 'Muffin'을 동시에 구매하였고 두 제품은 1.178의 연관성을 가진다.

다음으로 높은 지지도는 0.308로 'Plain Cheesecake, Muffin'의 구매이며 'Plain Cheesecake'을 구입한 고객 중 0.937은 'Muffin'을 구매하였고 두 제품의 연관성은 1.306이다. 'Choco Cheesecake, Muffin'의 구매는 0.242로 세 번째로 높은 지지도를 가지며 'Choco Cheesecake'를 구매한 고객의 0.904는 'Muffin'을 구매하고 두 제품은 '1.260'의 연관성을 가진다.

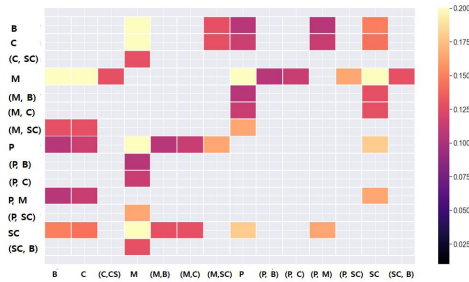
'Blueberry Cheesecake, Muffin'의 구매는 0.233의 지지도를 가지고 'Blueberry Cheesecake'를 구매한 0.908의 고객은 'Muffin'을 구매하는 경향을 보이며 두 제품은 1.265의 연관성을 가진다.

'Plain Cheesecake, Scone'을 구매하는 지지도는 0.179이며 'Plain Cheesecake'의 구매자의 0.547은 'Scone'을 구매하고 두 제품은 1.349의 연관성을 가진다. 'Plain Cheesecake, Scone, Muffin'을 구매하는 지지도는 0.164로 여섯 번째 순서이며 'Plain Cheesecake, Scone'을 구매하는 사람 중 0.916은 'Muffin'을 구매하며 1.277의 연관성을 가진다. 그 뒤로 'Blueberry Cheesecake, Scone' 구매가 0.149의 지지도를 가지며 'Blueberry Cheesecake'를 구매하는 사람의 0.579는 'Scone'을 구매하며 두 제품은 1.429의 연관성을 가진다. 'Choco Cheesecake, Scone' 구매는 0.144의 지지도를 가지고 'Choco Cheesecake'을 구매한 사람의 0.535는 'Scone'을 구매하며 'Choco Cheesecake'와 'Scone'은 1.321의 연관성을 가진다.

'Choco Cheesecake, Scone, Muffin'의 구매는 0.129의 지지도를 가지고 'Choco Cheesecake, Scone'을 구매하는 사람의 0.896은 'Muffin'을 구매하며 1.248의 연관성을 가지게 된다. 'Blueberry Cheesecake, Scone, Muffin'의 구매는 0.128의 지지도를 가지고 'Blueberry Cheesecake, Scone'을 구매한 고객의 0.863은 'Muffin'을 구매하며 1.203의 연관성을 가진다.

지지도 기준 분석을 통해 치즈케이크 종류에 따라

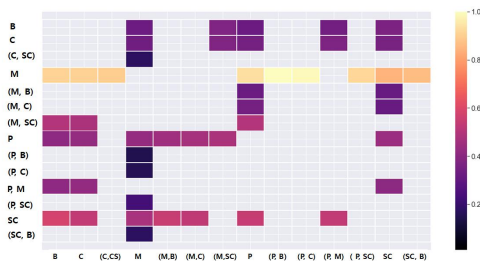
함께 구매하는 상품의 선호도가 달라지는 경향을 확인할 수 있다.



B=Blueberry Cheesecake, C=Choco Cheesecake, SC=Scone, M=Muffin, P=Plain Cheesecake

Fig. 2. Visualize association rules(Support)

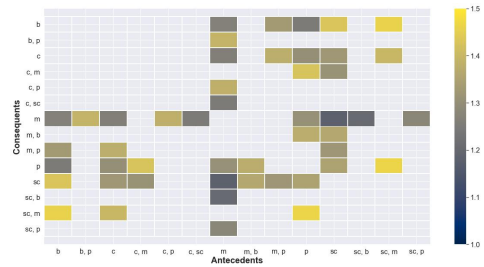
가장 높은 신뢰도를 나타내는 규칙은 ‘Plain Cheese Cake, Muffin’의 구매이며 ‘Plain Cheese Cake, Scone’을 구매한 고객이 ‘Muffin’의 구매가 포함되는 거래의 신뢰도가 0.916으로 높게 형성되어 있었다. 신뢰도가 높을수록 유용한 규칙이라고 할 수 있으며 이는 ‘Muffin’이 함께 구매되는 패턴이 가장 유용성 있는 구매패턴이라고 이해할 수 있다. 치즈케이크 제품과 ‘Scone’을 구매하는 신뢰도는 ‘Muffin’을 구매하는 경우보다 비교적 낮게 도출되었다. 그러나 연관규칙 분석 결과 지지도 기준으로 정렬하였을 때 모든 상위 10건의 규칙에서 ‘Muffin’이 발견된 점을 보았을 때 사람들은 해당 베이커리 기업의 ‘Muffin’을 더 선호하기 때문에, 모든 규칙에서 ‘Scone’이 포함된 거래보다 높은 신뢰도를 나타내고 있는 것으로 추측할 수 있다. 이는 지지도를 기준으로 분석했을 때의 결과를 다시 한번 설명해주고 있으며 이를 Heatmap으로 표현한 결과는 Fig. 3.와 같다.



B=Blueberry Cheesecake, C=Choco Cheesecake, SC=Scone, M=Muffin, P=Plain Cheesecake

Fig. 3. Visualize association rules(Confidence)

본 연구에서 수행한 연관규칙 분석의 향상도는 모두 1 이상의 값을 가지고 있다. 모든 연관규칙이 양의 상관관계를 가지고 있음을 뜻하며 이를 Heatmap으로 표현한 결과는 Fig. 4.와 같다.



B=Blueberry Cheesecake, C=Choco Cheesecake, SC=Scone, M=Muffin, P=Plain Cheesecake

Fig. 4. Visualize association rules(Lift)

신뢰도 기준의 분석 결과와 지지도 기준의 분석 결과에서는 다른 인사이트를 얻을 수 있다. 치즈케이크 제품의 구매가 ‘Muffin’의 구매로 이어지는 규칙에서 지지도 순으로 정렬하면 ‘Plain Cheese cake’(0.308), ‘Choco Cheesecake’(0.242), ‘Blueberry Cheesecake’(0.233) 순서로 나타난다. 그러나 신뢰도 순으로 정렬한다면 ‘Plain Cheesecake’(0.937), ‘Blueberry Cheesecake’(0.908), ‘Choco Cheesecake’(0.904) 순으로 나타나게 된다. ‘Choco Cheesecake’와 ‘Muffin’ 구매를 선호하지만 ‘Blueberry Cheesecake’와 ‘Muffin’을 구매하는 규칙의 유용성이 높은 것이다. 이러한 경향은 치즈케이크류 제품 구매가 ‘Scone’ 구매로 이어지는 패턴에서도 발견되었다. 지지도 기준으로 정렬하면 ‘Plain Cheesecake’ (0.179), ‘Blueberry Cheesecake’(0.149), ‘Choco Cheesecake’(0.144)의 순서로 나타나지만, 신뢰도 기준으로 정렬했을 때, ‘Blueberry Cheesecake’(0.579), ‘Plain Cheesecake’ (0.547), ‘Choco Cheesecake’ (0.535) 순으로 나타나게 된다. 이처럼 그 기준에 따라 규칙의 순서가 달라지므로 지지도, 신뢰도를 통합적으로 고려하여 전략을 수립해야 한다.

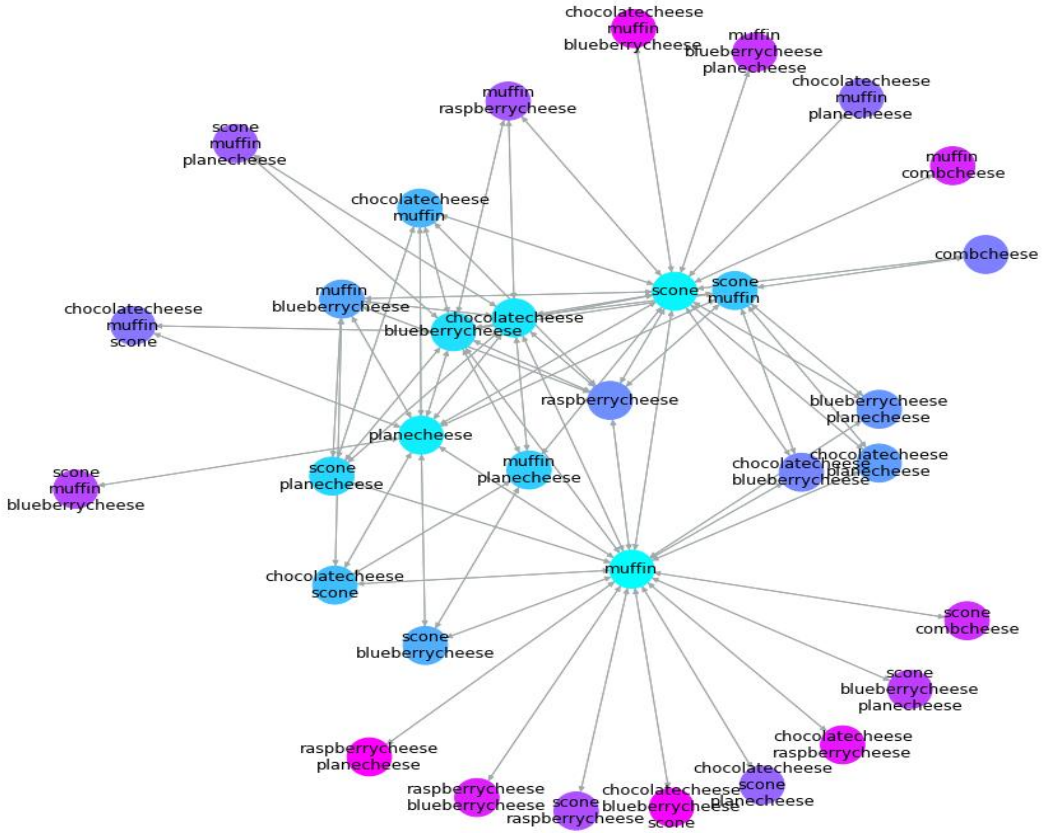


Fig. 5. Network Diagram of Rules

Note. The color displayed in figure.5 does not show a correlation

### 4.3 분석 결과 시각화

분석한 결과를 시각화하기 위해 Python3의 NetworkX 패키지를 사용하여 그린 관계도 그래프는 Fig. 5와 같으며 구매되는 모든 치즈케이크 제품군과 'Muffin'을 함께 구매하는 것을 볼 수 있었다. 또한, 해당 베이커리 기업은 당근케이크, 크림 케이크를 포함한 다양한 베이커리 제품들을 판매하고 있음에도 발견된 규칙들이 'Muffin', 'Scone'으로 집중되고 있었고 치즈 케이크류의 제품들이 지지도 기준 상위 규칙으로 이어지고 있는 것을 확인할 수 있었다. 해당 기업은 'Muffin', 'Scone'과 치즈케이크 제품들과의 묶음판매 판매 전략을 피할 수 있다. 또는, 치즈케이크 제품들이 많이 판매되는 것으로 보았을 때 전문성을 살려 치즈케이크 제품들만을 판매하는 소품종 대량생산을 통하여

원활한 재고관리로 불필요한 지출을 줄여나가는 전략을 취할 수 있을 것이다.

## 5. 결론

ICT 발전과 함께 다양한 데이터의 수집과 활용이 가능해지면서 기업 경영전략과 산업 생태계를 변화시키고 있다. 소비자들의 요구와 취향이 더욱 빠르게 변화하고 있으며 이에 따라 기존의 산업은 이러한 현상에 대응하기 위한 전략이 필요하다. 본 연구에서는 중소기업에 속한 'A' 베이커리 기업의 실제 데이터 분석을 통해 소상공인을 포함한 타 산업 분야 중소기업에서도 축적된 데이터 분석 및 데이터베이스 구축을 통해 경영환경 개선을 위한 전략을 수립할 수 있는 시사점을 제시

하였다.

오늘날 기업들은 산업 내 경쟁력을 키우고 새로운 성장 동력을 발굴하기 위한 수단으로 디지털 트랜스포메이션과 데이터 분석을 활발하게 진행하고 있다. 그러나 많은 기업이 기술 우선주의 함정에 빠져 조직 내 혁신을 끌어내지 못하는 상황이 빈번하게 발생하고 있다. Boston Consulting Group은 기업이 디지털 트랜스포메이션을 성공적으로 수행하기 위해서는 넓은 시야를 가지고 지속적인 모니터링이 필요하다고 언급했으며[4], 디지털 기술 도입을 통한 혁신이 아닌 전략관점에서 주도해야 한다고 말했다[37]. 기술력과 자본이 충분한 대기업의 경우 많은 양의 데이터 분석과 기술혁신을 통해 고객 경험 증진과 혁신을 이끌어내고 있다. 이와 반대로 대다수 중소기업의 경우 규모, 역량, 기술, 자원의 한계로 인하여 데이터 기반의 경영 의사결정이 더디게 일어나고 있다.

이전의 추천 시스템과 관련된 연구로 콘텐츠 기반 필터링과 협력 필터링 모델을 활용한 연구가 있다. 연관규칙분석을 활용한 기업의 교차판매 모델 개발 사례에 대해 분석하여 경영 효율화를 시사한 연구[38], 온라인 쇼핑물 구매 데이터의 연관규칙분석 수행으로 효율적인 운영 방법들을 제시하는 연구[39]가 진행되었다. 김형수 등(2021)의 연구는 협력 필터링과 소셜 네트워크 분석기법을 응용한 상품 네트워크 분석기법을 연계하여 장기적인 고객과의 관계 유지를 통한 고객 충성도 개선을 시사하였고[40]. 박성중(2019) 등의 연구는 구매 데이터를 기반으로 협력 필터링과 분류모형 중 스택킹 방법을 적용한 맞춤형 추천 시스템을 개발하여 새로운 경영전략을 수립하였다[41]. 김정훈 등(2012)의 연구는 콘텐츠 기반 필터링과 협력 필터링을 혼합한 혼합 필터링 방식을 사용하여 고객 맞춤형 건강기능식품 추천 시스템을 개발하여 고객의 구매 만족도를 높이고자 하였다[42].

본 연구에서는 온라인 채널을 통해 해당 기업의 제품을 10년 넘게 판매하고 있으나 그간 수집한 데이터의 분석 및 활용이 전혀 없는 중소 베이커리 기업을 연구대상으로 선정하였다. 기업의 실증 데이터를 가지고 연관규칙분석을 수행한 결과 해당 기업의 고객들이 특정 제품을 높은 확률로 함께 구매한다는 인사이트를 얻을 수 있었다. 기업 측면에서는 B2B와 B2C 채널을 함께 운영하고 있어 COVID-19으로 인한 기업 충주문량 감소, 재고관리 등에 대해 어려움이 있었다. 따라서 원

활한 재고관리를 통한 경영 악화 방지를 위해 연관규칙 분석으로 유의미한 결과를 나타낸 상품 위주의 '소품종 대량생산' 전략으로 COVID-19 위기 상황 속 경영 개선을 도모할 수 있을 것이다. 해당 기업의 실증 데이터 분석을 통해 디지털 역량이 부족한 타 산업 분야에 속한 중소기업에서도 데이터 분석으로 기업의 향후 경영 전략과 변화하는 소비자들의 요구에 대응할 수 있다는 시사점을 제시하고자 하였다.

본 연구의 한계점은 다음과 같다. B2B 중심의 기업이기 때문에 B2C 데이터가 비교적 부족하다. 따라서 도출된 연구 결과를 일반화하기에는 부족할 수 있다. 하지만 기존의 데이터 분석을 통해 데이터 기반 의사결정의 가능성을 파악하고, 재고 효율화 및 위기 대응을 할 수 있음을 시사한다. 또한, 본 연구 대상인 기업은 베이커리 산업에 속한 중소기업으로 모든 산업 분야에서 같은 인사이트를 도출하는 것은 어려울 수 있다. 하지만 골목상권의 소상공인부터 인적, 물적 자원이 부족한 중소기업이 그간 수집된 판매 데이터 분석만으로도 의미 있는 시사점을 도출할 수 있는 가이드라인이 될 수 있다. 본 연구는 이용자의 구매 데이터 기반의 정보 필터링을 이용한 연구로 성별, 연령 등 고객들의 특성이 고려되지 않았다. 향후 연구에서는 본 중소기업의 고객 특성을 포함한 협력 필터링 기반 연구를 진행하여 유사한 성향의 사용자 추천을 통해 중소기업과 소상공인의 경영전략을 제시하는 연구를 진행하고자 한다.

## REFERENCES

- [1] Y. S. Cho. (2020). *Plan to Enhance the Acceptance of SMEs Through Digital Transformation*, Sejong : Korea Institute for Industrial Economics & Trade.
- [2] S. H. Bae. (2021) *Survey on small business owners 2020*. Sejong : Korea SMEs and Startups Agency.
- [3] T. H. Kim. (2020). *Domestic Bakery Market Trends and Change in Consumer Trends*. Seoul : KB Research.
- [4] Boston Consulting Group. (2020). *The Evolving State of Digital Transformation*. Boston. BCG
- [5] Y. H. Goh, Y. R. Yoon, J. N. Park & Y. E. Jo (2020). *Research on how to utilize data for development cooperation-projects*. Seongnam : KOICA.
- [6] H. M. Lee (2018). *Leading innovation in digital*



- transformation. *Technology Tech Issue*. Daejeon : ITFIND.
- [7] Y. O. Kwon. (2013). Data Analytics in Education : Current and Future Directions. *Journal of Intelligence and Information Systems*, 19(2), 87-100.
- [8] H. G. Jun, G. S. Hyun, K. B. Lim, W. H. Lee & H. J. Kim.(2014). Big Data Preprocessing for Predicting Box Office Success. *KIISE Transactions on Computing Practices*, 20(12), 615-622.
- [9] D. J. Kim, D. I. Park & J. S. Park. (2018). Study on the Change of Marketing Strategy through Data Mining Technique. *Korea Business Review*, 22(2), 177-194.
- [10] J. S. Lee, J. H. Lee, G. J. Kim, S. S. Park & D. S. Jang. (2015). Establishment of Strategy for Management of Technology Using Data Mining Technique. *Journal of Korean Institute of Intelligent Systems*. 25(2), 126-132.
- [11] Y. J. Yi, S. H. Lee & J. S. Yi. (2014). KB Kookmin Card's Marketing Activities and Use of Big Data. *Korea Business Review*, 18(1), 145-176.
- [12] K. C. Ahn, C. B. Moon, B. M. Kim, Y. S. Shin & H. S. Kim. (2012). POS Data Analysis System based on Association Rule Analysis. *Journal of the Korea Industrial Information Systems Research*, 17(5), 9-17.
- [13] Y. H. Cho. (2020). A Study on the Service Marketing Promotion Plans of Hotel Bakery and Perception on Hotel Bakery by Analyzing Unstructured Big Data. *Journal of Tourism and Leisure Research*, 32(2), 285-298.
- [14] A. Dagvadorj, Y. R. Lee & H. S. Kim.(2020).A Study on the Consumer's Perception of Bakery Using Big Data Analytics. *Culinary Science & Hospitality Research*, 26(4), 66-75.
- [15] D. Y. Oh & D. H. Kim. (2017). Effects of Large Business Group's Intra-Firm Transaction on Small and Medium-sized Enterprises. *Ordo Economics Journal*, 20(2), 49-67.
- [16] D. K. An, S. C. Ho & J. W. Choi. (2019). A Study on Efficient Decision-Making Using Big Data by Managers of Small and Medium Size Enterprises. *Korean Review of Corporation Management*, 10(3), 311-326.
- [17] J. Y. Park & K. I. Kim. (2018). ICT Utilization for Optimization of SME Decision Making. *Journal of Convergence for Information Technology*, 8(1), 275-280.
- [18] D. W. Park & K. Y. Kwahk. (2020). The Effects of Information Systems Based Working Environment on the Performance of SMEs. *Korean Management Review*, 49(1), 215-249.
- [19] Y. Liu, H. F. Liu & J. S. Jung. (2020). Strategies for Improving Export and Distribution of SMEs Using the Platform of the 4th Industrial Revolution Technology: Focusing on the Chinese SMEs. *Journal of Distribution and Management Research*, 23(6), 31-48.
- [20] H. Y. Lee, S. K. Lee & C. J. Suh. (2019). Effects of SST(Self-Service Technology) Service Quality and Perceived Value on Customer Engagement. *Journal of Korea Service Management Society*, 20(5), 69-92.
- [21] Korean Rural Economic Institute. (2021), *Report on the status of restrant business management*. Naju : Korean Rural Economic Institute.
- [22] R. Agrawal, T. Imielinski & A. Swami (1993). *Mining association rules between sets of items in large databases*. Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA.
- [23] M. J. A. Berry & S. L. Gordon. (2004). *Data mining techniques: for marketing, sales and customer relationship management*. New York, NY: John Wiley & Sons. Inc.
- [24] C. Westphal & T. Blaxton. (1998). *Data mining solutions: Methods and tools for solving real-world problems*. New York, NY: John Willey & Sons. Inc.
- [25] Y. K. Jeong. (2003). A Study on the Analysis of Restaurant Customers' Behavior Patterns Using Association Rules. *Korean Journal of Hospitality & Tourism*, 12(2), 1-22.
- [26] A. Savasere, E. Omiecinski & S. Navathe. (1995). An efficient algorithm for mining association rules in large databases. *Proceedings of 21st International Conference on Very Large Data Bases*. (pp. 432-444). San Francisco : Morgan Kaufman Publishers Inc..
- [27] A. Paul & M. David. (2015). *Data mining for the social sciences: An introduction*. Oakland : University of California Press.
- [28] K. S. Lee & J. W. Yoon. (2016). Rapid Hybrid Recommender System with Web Log for Outbound Leisure Products. *KIISE Transactions on Computing Practices*, 22(12), 646-65. DOI : 10.5626/KTCP.2016.22.12.6
- [29] E. S. Won & S. Y. Kim. (2020). An Analysis of Consumers Purchasing Patterns for Fresh Food Products Using Association Rules. *Journal of Agriculture & Life Science*, 54(4), 111-122. DOI : 10.14397/jals.2020.54.4.111

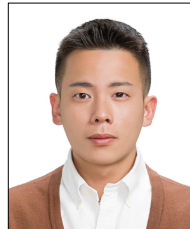
- [30] J. Silva, N. Varela, L. López & R. Millán. (2019). Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm. *Procedia Computer Science*, 151, 1207-1212.  
DOI : <https://doi.org/10.1016/j.procs.2019.04.173>
- [31] W. Y. Chiang. (2018). Identifying high-value airlines customers for strategies of online marketing systems: An empirical case in Taiwan. *Kybernetes*, 47(3), 525-538.  
DOI : 10.1108/K-12-2016-0348
- [32] F. Guillet & H. Hamilton. (2007). *Quality Measures in Data Mining*. Heidelberg : Springer.  
DOI : 10.1007/978-3-540-44918-8\_11
- [33] D. W. Lee. (2017). A Regression-Model-based Method for Combining Interestingness Measures of Association Rule Mining. *Journal of Intelligence and Information Systems*, 23(1), 127-141.
- [34] S. S. Kim. (2021). Association Rules of Comorbidities in Patients Hospitalized for Lung Cancer. *Journal of Health Informatics and Statistics*, 46(1), 100-109.
- [35] M. U. An, E. S. Won, S. Y. Kim & D. H. Yoo. (2019). Development of Sales Strategies for Agricultural Products Using Lift-based Association Rules Network: A Focus on Large Supermarkets and Traditional Markets. *The Journal of Internet Electronic Commerce Research*, 19(3), 105-127.
- [36] S. S. Oh, H. S. Jung & J. Y. Lee. (2019). A Study on the Customer Profile based Item Recommendation System using Association Rules Analysis for Online Duty Free Stores. *Korean Journal of Logistics*, 27(2), 1-16.
- [37] G. Kane, D. Palmer, A. Phillips, D. Kiron & N. Buckley. (2018). *Coming of Age Digitally*. MIT(Online).  
<https://sloanreview.mit.edu/projects/coming-of-age-digitally/>
- [38] S. H. Jin & J. H. Choi. (2008). Data Mining Application for Effective Cross Sell Marketing. *Journal of Korea Data Analysis Society*, 10(5), 2629-2638.
- [39] Y. G. Jung, J. K. Park, J. C. Lee & E. Y. Choi. (2012). An Study on the Product Purchase Patterns using Association Rule. *Journal of Service Research and Studies*, 2(1), 39-46.
- [40] H. S. Kim, S. H. Kim & M. C. Ma. (2021). A Recommendation System Based on PNA Collaborative Filtering for Reinforcing Customer

Loyalty. *Information Systems Review*, 23(2), 21-35.

- [41] S. J. Park, Y. M. Kim & J. J. Ahn. (2019). Development of Product Recommender System using Collaborative Filtering and Stacking Model. *Journal of Convergence for Information Technology*, 9(6), 83-90.
- [42] J. H. Kim, B. H. Ahn & D. Y. Jeong. (2012). A Recommender System using Mixed Filtering for Health Products. *The Journal of Internet Electronic Commerce Research*, 12(2), 109-124.
- [43] D. H. Kim. (2021). *Machine learning-based purchase data analysis for implementing digital transformation strategy*. Masters dissertation Yonsei University, Seoul.

#### 김도훈(Do Hoon Kim)

[정회원]



- 2018년 6월 : University of California Santa Cruz Economics (B.S)
- 2021년 8월 : 연세대학교 기술경영 학협동과정 (경영학석사)
- 2021년 9월 ~ 현재 : 연세대학교 기술경영학협동과정 박사과정

· 관심분야 : 경영, UI/UX, IT정책

· E-Mail : [dkim940627@gmail.com](mailto:dkim940627@gmail.com)

#### 이현준(Hyeon June Lee)

[학생회원]



- 2018년 2월 : 중앙대학교 간호대학 (간호학사)
- 2020년 9월 ~ 현재 : 연세대학교 정보대학원 석사과정
- 관심분야 : 데이터 전략, ICT-콘텐츠
- E-Mail : [shouldbe28@yonsei.ac.kr](mailto:shouldbe28@yonsei.ac.kr)

#### 이봉규(Bong Gyou Lee)

[정회원]



- 1988년 2월 : 연세대학교 상경대학 (경제학사)
- 1992년 5월 : Cornell University, City and Regional Planning (M.S)
- 1994년 1월 : Cornell University, City and Regional Planning (Ph.D)

· 2005년 3월 ~ 현재 : 연세대학교 정보대학원 교수

· 2018년 3월 ~ 2020년 2월 : 연세대학교 학술정보원 원장

· 관심분야 : 플랫폼 비즈니스 전략, ICT-콘텐츠, IT정책

· E-Mail : [bglee@yonsei.ac.kr](mailto:bglee@yonsei.ac.kr)