

## 무슬림 관광객 증대를 위한 머신러닝 기반의 할랄푸드 분류 프레임워크\*

김선아\*\* · 김정원\*\* · 원동연\*\* · 최예림\*\*\*

### 〈목 차〉

- |                      |                  |
|----------------------|------------------|
| I. 연구의 필요성 및 목적      | IV. 실험 수행        |
| II. 이론적 배경           | 4.1 실험 데이터       |
| 2.1 무슬림의 한국 관광       | 4.2 실험 환경        |
| 2.2 할랄푸드             | 4.3 실험 결과        |
| 2.3 머신러닝 적용 식품 분류    | V. 결론            |
| III. 프레임워크           | 5.1 연구 요약        |
| 3.1 데이터 수집 및 전처리     | 5.2 연구 시사점 및 한계점 |
| 3.2 머신러닝을 적용한 데이터 학습 | 참고문헌             |
| 3.3 광학적 문자 인식        | <Abstract>       |
| 3.4 섭취가능여부 예측        |                  |

### I. 연구의 필요성 및 목적

최근 중국의 한국관광규제의 대처방안으로 한국관광공사의 ‘프렌들리 코리아’(한국관광공사, 2017), 문화체육관광부의 ‘인바운드 관광시장의 개선’(조운선, 2017) 등 각 정부기관의 정책에서 살펴볼 수 있듯이 관광업계는 관광다변화를 모색하고 있으며 관광다변화를 위해 동남아, 중동시장의 무슬림 관광객을 주목하고 있다

(이석호, 2016). 무슬림 시장의 가치를 살펴보면, 무슬림 인구는 18억 명으로, 이는 전 세계 인구의 1/4에 해당되며, 계속적으로 늘어날 전망이다(Grim, B. J. & Karim, M. S., 2011). 또한, 무슬림 여행객의 인당 지출규모가 높아 1억 8백만여 명의 무슬림 인구가 1,450억 달러인 해외여행의 12.3%를 차지하여 상당히 매력적인 시장으로 각광받고 있다(MasterCard & Crescent Rating, 2015). 실제로도 10년간 방한

\* 본 연구는 경기도의 경기도 지역협력연구센터 사업의 일환으로 수행하였음.

[GRRC경기2017-B01, 지능형 제조 빅데이터 분석 연구]

\*\* 경기대학교 산업경영공학과 학부생, {sunak1028, jeongwon07}@gmail.com, ckddydcjstk@naver.com

\*\*\* 경기대학교 산업경영공학과 조교수, yrchoi@kgu.ac.kr (교신저자)

무슬림은 지속적인 증가추세(한국관광공사, 2016a)에 있었고, 저가항공사의 등장 또는 지리적으로 근접하다는 이점과 K-Pop 및 드라마로 인한 한류 열풍으로 한국관광에 긍정적인 관심이 상승하고 있는 이유로 방한 무슬림의 유입 증가 예상에 따른 대비책이 필요하다.

무슬림들의 한국관광 시 겪는 문제점에 관한 설문조사 결과를, 한국여행의 개선요인 1위로 음식(한국관광공사, 2016b)이 선정되어 음식부분에 불편을 많이 겪고 있음을 알 수 있다. 또한, 여행 중 음식섭취 비중의 설문항목에서는 이슬람의 고리로 인해 여행 중에도 할랄푸드 섭취를 선호(한국관광공사, 2016b)하고 있었다. 할랄푸드란 이슬람 율법에 부합하는 제조과정을 거친 식품으로 정의되고, 할랄푸드로서 인증을 받기 위해서는 육류의 경우 도축의 과정이 꾸란에 근거한 과정대로 이루어져야 하며, 유통과정 또한 이슬람 율법에 맞는 절차를 거쳐야 한다(채경연 & 이희열, 2015).

위와 같은 까다로운 절차로 인해 국내에 할랄인증 받은 할랄푸드는 매우 적기 때문에(이슬기 & 이미경, 2017) 무슬림들은 여행 전 한국에서의 할랄푸드 구매에 대한 어려움에 관한 후기를 듣고, 자국의 인스턴트 할랄푸드를 준비해와 숙소에서 조리 해먹는 경우가 많았다(채경연 & 이희열, 2015). 숙소 내에서의 조리도 마땅하지 않은 경우, 무슬림은 기존에 존재하는 국내의 식품 원재료 표기부분으로 섭취가능여부를 자가판단 해야 하는 문제가 발생한다. 하지만, 이 과정에서 혼합성분에 돼지고기의 부산물 같은 하람재료가 포함되어 있어 섭취가 불가하나, 원재료 표기부분과 알레르기 유발성분 표기란에도 기재되어 있지 않은 경우가

존재한다. 또한, 재료의 수가 너무 많은 경우와 원재료명이 표준화 되어있지 않은 문제로 섭취가능여부에 대한 자가판단이 어렵다.

따라서 자가판단이 어려운 문제를 해결하기 위해 정보통신기술의 발전에 따라 무슬림들은 스마트 관광을 구성하고 있는 채널중의 하나인 어플리케이션을 이용하고 있었다. 여행 중에 방한 무슬림 관광객들의 모바일 이용률은 96.5%(한국관광공사, 2014)로 대부분의 방한 무슬림 관광객들이 모바일 즉, 소셜미디어를 활용하여 스마트관광의 형태로 관광 중 정보를 검색하며(한국관광공사, 2016b), 관광 중 식품 섭취에 대한 의사결정을 도움 받고 있었다(Ricci, 2002).

주목해야할 문제는 스마트 관광의 어플리케이션 채널을 이용하면서 아쉬운 점으로 정보의 결여에 대한 부분을 꼽았다는 것이다(한국관광공사, 2014). 실제로 기존 어플리케이션의 가장 큰 문제점은 모든 식품을 상대로 데이터베이스로 구축하는 것이 현실적으로 어렵기에 식품검색 시 데이터가 존재하지 않아 검색이 불가능한 경우가 많다는 것이다. 검색이 되지 않아 정보가 제공되지 않을 시 무슬림들은 음식 섭취 가능여부에 대한 판별이 불가능해 어려움을 겪게 된다. 검색 불가 문제로 인해 무슬림들은 어플리케이션 내의 커뮤니티 기능을 이용하여 원재료 표기부분을 촬영한 사진으로 어플리케이션 사용자들끼리 자체적으로 섭취가능여부에 대해서 묻고 답하는 번거로운 식품검색 과정을 거친다.

정보획득 시간의 단축을 위해 기존 보유하고 있는 데이터만 가지고도 데이터 자체를 학습시켜 새로운 데이터의 판별이 가능한 지능시스템(Werthner, 2004)인 머신러닝 기법의 도입을 통

해 관광 중 편의제공이 필요하다. 머신러닝이란 주어진 데이터에 대해서 컴퓨터가 학습할 수 있게 하여 스스로 성능을 향상시킴으로써 입력하지 않은 데이터에 대해서 예측의 수행과 판단을 가능하게 하는 기술로, 주로 패턴인식에 적용이 용이하다(Samuel. Arthur L, 1959). 이러한 머신러닝을 식품의 분류기법에 적용하면 무슬림에게 최소한의 섭취가능여부 판별의 가이드라인을 제공할 수 있다.

따라서 본 논문에서는 기존 시스템으로 제품 검색 시 데이터가 존재하지 않을 경우, 부족한 정보제공문제 해결을 위해 머신러닝의 기법을 적용한 무슬림의 식품 섭취가능여부 판별에 활용할 수 있는 프레임워크를 제안하고자 한다. 제안하는 프레임워크는 크게 네 부분으로 구성된다. 첫째, 기존 식품데이터의 첨가물과 섭취가능여부의 데이터 수집과 전처리 과정이 이루어지는 부분, 둘째, 기존 데이터를 머신러닝 기법을 적용하여 학습시키는 부분, 셋째, 사용자가 식품첨가물 표기부분 촬영 시 OCR(광학적 문자 인식, Optical Character Recognition)의 과정을 거쳐 텍스트 데이터로 받아들이고 기존 DB(Data Base)와 대응시키는 부분, 마지막으로, 학습된 데이터를 기반으로 머신러닝 기법을 적용하여 입력된 식품의 섭취가능여부를 예측하는 부분으로 이루어져 있다. 이 프레임워크를 새로운 시스템으로 접근하기보다 기존 이용하고 있는 어플리케이션에 제안하는 프레임워크의 기능을 더해 전환비용을 줄여 사용자들에게 거부감 줄이고 긍정적인 수용을 할 수 있도록 하는 방안이 필요하다(이응규 & 김효정, 2013).

2장에서는 무슬림의 한국 관광, 할랄푸드 및 식품분류에 머신러닝이 적용된 기존연구를 정

리하며 3장에서는 무슬림의 식품섭취 가능여부 판별을 위한 프레임워크의 전체 구조와 더불어 각 구성 별 상세설명을 한다. 4장에서는 이를 기반으로 실제 식품데이터에 머신러닝 적용 시 모델 별 성능에 대한 실험 결과를 비교하고, 통계적으로 유의성을 검증한다. 마지막으로, 5장에서는 본 연구에 대한 결론과 시사점 및 향후 연구 방안에 대해서 논의한다.

## II. 이론적 배경

### 2.1 무슬림의 한국 관광

이슬람 종교에 대한 교리를 기본으로, 관광은 종교의 일부분이고 여행은 이슬람의 기반이며, 스트레스 없는 삶을 영위하고 무슬림 공동체 구성원들의 연결을 강화하기 위한 목적으로 장려되었다(Hasharina, 2006). 아울러 무슬림은 여행에 중요한 가치를 부여하고 있으며 이 가치는 계속적으로 전승되어 현대 무슬림들도 여행을 중요시 여긴다(Kessler, 1992; Bhardwaj, 1998; Aziz, 2001). 무슬림들의 관광을 이해하기 위해서는 우선적으로 종교와 연관된 삶에 대한 이해가 필요하다. 무슬림 관광은 ‘여행 동안 이슬람의 관습을 지키며, 이슬람의 교리에 부합하는 관광’으로 정의된다(Henderson, 2016).

무슬림들은 생애 한 번은 성지순례를 위한 여행을 하는데 최근에는 종교적 관광 목적 이외에도 휴식과 같은 순수 관광을 위한 여행이 급증하고 있다(서민교, 2017). 무슬림 관광시장의 지속적 성장 전망을 예상하는 다섯 가지 요인으로(Master Card & Crescent Rating, 2016)

첫째, 무슬림 인구의 지속적인 증가이다. 우리나라는 출산율의 저하로 인해 인구가 계속적으로 감소 추세를 예상하는 반면 무슬림 인구는 계속적으로 성장하여 2030년에 약 22억 명까지 증가할 것이라 예상된다. 둘째, 중산층의 확대이다. 터키, 중동 지역은 중산층이 많고, 동남아시아 또한 각 국가의 경제성장으로 무슬림 중산층의 증가가 예상된다. 또한 유럽, 북미 지역의 무슬림들의 높은 교육수준으로 인해 전문직종사자들이 많으므로 무슬림이 창출하는 경제적 가치에 주목해야 한다. 셋째, 젊은 연령층의 인구비중이 높은 점이다. 무슬림은 2010년 기준 평균 중위 값 23세로 비 무슬림에 비해 7살이 낮으며, N세대이자 무슬림 밀레니엄인 1990~2000년대에 태어나 현재 2~30대에 해당하는 이슬람교도, 특히 여성 무슬림들의 교육수준은 매우 높고 현대적 생활방식에 관심이 많아 무슬림 소비시장을 주도할 것이라 예측된다(엄익란, 2014). 넷째, 현 정보화시대에서는 여행정보 수집이 용이한 점이다(양성수 외, 2008). 특히, 중동지역 국가들은 스마트 폰의 이용률이 높으며 이에 따라 인터넷 및 모바일을 통해 여행 전, 중으로 관광정보에 대한 손쉬운 접근이 가능해져 정보획득이 편리 해졌다. 마지막으로, 무슬림에 대한 관심이 높아져 무슬림 친화 서비스가 증가하고 있다. 무슬림 시장의 가치 부상에 따라 무슬림을 위한 할랄 음식점과 기도실과 같은 무슬림 인프라 구축에 대한 관심이 높아지고 있다.

2016년 한국관광공사의 방한 무슬림 관광 실태조사 보고서를 살펴보면, 패키지여행 보다 개별 자유 여행의 비중이 높았으며, 가족과 함께 있는 시간을 중요시하는 무슬림의 특성상 주로

가족단위로 방문하거나, 최근 20대 층은 친구와 함께 여행하는 것을 선호하였다. 방문지역은 명동, 동대문시장, 남산/서울N타워가 있는 서울지역의 방문이 대부분을 차지하였으며, 남이섬이 있는 강원도 지역이 두 번째로 많았다. 그러나 무슬림 인프라가 열악한 지방의 경우 거의 방문하지 않았다(한국관광공사, 2016b). 이들의 주요 방한 목적으로는 여가, 위락, 휴가의 목적이 대부분이었으며, 사업 또는 전문 활동, 쇼핑 순 이었다(한국관광공사, 2016b). 최근에는 의료관광이 주목받고 있는데, 중동지역의 서구적으로 변하는 식습관과 무더운 기후로 인하여 약 성인의 절반이 비만과 고혈압, 당뇨 등을 앓고 있으며, 이 질병을 앓는 환자가 급격히 증가하고 있으나 자국의 의료 인프라가 부족한 실정으로 인해 해외의 의료 서비스에 의존하는 경향이 높아지고 있다(석수선, 2016). 최근에는 중동 지역 부호들의 의료관광의 수요가 늘어나며, 의료 관광의 목적으로 한국을 방문하는 중동 관광객의 수도 급증하고 있다(배영희, 2015). 무슬림 관광객의 수는 중국인 관광객에 비해 낮지만 의료비로 지출한 금액은 아랍에미리트 관광객 기준 1인당 1,771만원에 이르고 카자흐스탄(1인당 256만원), 인도네시아(1인당 193만원)의 지출수준으로 중국(1인당 181만원)보다 높다(한국관광공사, 2014; 서민교, 2017).

무슬림은 이슬람의 다섯 가지 지주로 인해 하루에 다섯 번 기도하여야 한다. 그러나 국내에는 기도실이 제대로 갖춰지지 않아 평소보다 기도를 적게 하며, 여행일정을 고려하여 건너 뛰거나 한 번에 몰아서 행한다. 이로 인해, 관광지에서도 불편함 없이 기도를 할 수 있는 독립적인 공간의 기도실에 대한 니즈가 크다. 설문

조사 결과 기도실 보다 중요하다고 응답한 다섯 가지 지주 중 하나는 할랄에 관한 규정이다(채경연 & 이희열, 2015). 무슬림 여행자의 66.8%가 여행 시 할랄푸드를 선호하고, 식품을 구매 시 90%의 무슬림이 할랄 인증 여부가 구매결정에 영향을 미친다고 응답하였다(Master Card & Crecent Rating, 2016). 마찬가지로 방문 무슬림 관광객의 대부분이 여행 중 할랄푸드를 섭취하고 있었지만, 할랄 인증을 받은 음식점이 서울에 8곳뿐이며, 지방에는 거의 전무한 이유로 하람 성분이 포함된 음식이 아니면 일반 식당을 이용하는 비중이 높았다. 또한, 일부 무슬림들은 사전에 음식에 대한 정보를 입수하고 자국에서 미리 인스턴트 식품을 준비해와 숙소에서 조리해 먹는 경우도 있어 음식에 관한 부분에 많은 어려움을 겪고 있는 것을 알 수 있다. 이로써, 관광 중 무슬림을 위한 음식에 관련된 부분의 미흡한 점이 한국여행 개선요인의 1위로 음식을 응답한 결과에 영향을 미침을 알 수 있다(한국관광공사, 2016b).

무슬림 여행지로서 한국은 치안이 우수하지만, 기도실, 할랄음식 등 무슬림 관광인프라가 부족한 이유로 관광지로서 한국에 대한 평가는 전체 54위로 매우 낮은 순위를 기록(MasterCard & Crescent Rating, 2015)하고 있다. 또한, 현재 무슬림 시장을 주목하고 있는 관광업계의 관심과는 별개로 무슬림들의 관광 관

련 연구는 거의 이루어지지 않고 있다(이희열 & 이선경, 2014). 따라서 관광 중 편의를 제공함으로써 방문 중 우리나라에 대한 긍정적인 경험을 바탕으로 스마트 폰 및 SNS의 발달에 따른 홍보 확산이 필요(채경연 & 이희열, 2015)하다. 따라서 무슬림과 관광 및 그들이 섭취하는 음식 부분의 관심과 그에 따른 심도 깊은 연구가 필요하다.

## 2.2 할랄푸드

이슬람 경전인 꾸란이나 예언자 무함마드의 언행록인 하디스에 삼라만상의 모든 기본원칙으로 할랄 또는 하람이 구분지어 진다(최영길, 1997). 할랄은 아랍어로 법이 인정 및 허가하는 것이며(Kamali, 2003) 일반적으로 이슬람법(Shari'ah)에 따라 허가된 것들을 의미한다. 이와 반대되는 개념인 하람은 불법적인 또는 허가되지 않는 것을 의미한다. 하람의 종류로는, 돼지, 송곳니를 가진 육식 동물, 양서류 등 이슬람법에 어긋나는 동물을 포함해 알코올 등이 있다(김종도 & 최영길, 2014). 무슬림은 할랄로 인정되는 고기만 소비할 수 있으며, 할랄 동물의 도축방법도 반드시 할랄 방식의 규정을 따라야 한다(홍완수, 2015). 무슬림은 돼지고기 뿐만 아니라 돼지에서 파생되는 모든 부산물에 대한 섭취도 금지되어 있다. 돼지의 부산물은

할랄 도축규정(홍완수, 2015):

- 1) 정신적으로 완전한 성인 무슬림에 의해 도살되어야 함
- 2) 도살되는 동물의 경우는 이슬람법에 허가된 동물이어야 함
- 3) 도살되는 동물은 도살되기 전 살아있는 상태여야 함
- 4) 도살되기 직전에 "Bismillah(알라의 이름으로)"라고 말하고 도살해야 함
- 5) 도살에 사용되는 도구(칼)는 날카로워야 하며 한번 움직임으로 도살해야 함
- 6) 도살하는 자는 동물의 목 부분의 기도, 식도, 경동맥을 끊어야 함

소시지부터 젤리, 아이스크림, 약을 제조하는 다양한 용도로 식품에 사용이 되어 할랄로 판별하기 위해서는 부산물도 엄격히 다뤄져야 한다. 할랄과 하람은 다양한 분야에서 구분 지어질 수 있는데, 대부분 식품에 주로 적용되며 최근에는 화장품, 퍼스널 케어 제품, 식기, 서비스 등의 분야에도 엄격하게 적용되고 있다 (Hanzaee & Ramezani, 2011).

시대의 변화에 따라 유전자조작, 첨가제 같은 새로운 형태의 식재료가 개발되고, 복잡한 가공과정 및 새로운 가공 식품의 출현으로 할랄과 하람을 구분하기 어려워져 할랄인증제가 도입되었다(김철민, 2015). 세계적으로 150~200여개의 할랄 인증기관이 운영되고 있으며, 그 중에서도 공신력 있는 인증기관으로 인정받고 있는 기관은 JAKIM(말레이시아 이슬람 개발국, JabatanKemajuan Islam Malaysia), MUI(인도네시아 율법학자위원회, Majelis Ulama Indonesia) 등이 있다. 그러나, 할랄인증은 한국에서의 적용이 어려운데, 그 이유로는 첫째, 한국의 할랄인증기관인 KMF(한국 이슬람교 중앙회, Korea Muslim Federation)는 이슬람 선교를 목적으로 설립된 종교 재단법인이며, 할랄 전문가를 비롯한 인력의 한계로 급변하는 할랄 산업에 따른 대응력이 저조하다. 또한, 국제 할랄 표준이 없어서, 할랄 수출을 위한 인증을 받기 위해서는 할랄 수입국가가 인정하는 인증을 각각 획득해야 한다는 어려움이 있어(김은미, 2015) 대부분의 식품이 할랄인증을 획득하지 못하고 있다.

최근, 까다로운 인증절차와 검열과정으로 인하여 비 무슬림에게도 할랄푸드는 깨끗하고 안전하다고 인식되어 세계 각국은 할랄을 주목하

고 있다(Bonne & Verbeke, 2006). 식품안전에 대한 관심이 높아짐에 따라 할랄푸드를 성분, 안전, 위생 측면에서 엄격한 검사와 절차를 따를 믿을 수 있는 식품으로 인정하고 있다(김은미, 2015). 할랄푸드 시장은 할랄푸드의 안전성에 따른 선호도 증가로, '13년 전년대비 10.8% 증가한 1조 2,920억 규모를 기록하였으며, 향후 지속적인 성장으로 '19년에는 2조 5,390억 달러 규모에 이를 것으로 전망되고 있다(김은미, 2015). 할랄푸드 시장규모가 확대됨에 따라 세계적 식품회사들이 앞 다투어 할랄푸드에 시장에 뛰어들어 할랄제품의 비중을 늘리고 있다(power, 2009). 우리나라의 식품기업들의 할랄 산업 진출은 아직까지 초기 단계이며(홍완수, 2015), 관련된 연구 또한 할랄 시장 진출과 관련된 전략, 시장성, 소비패턴 등의 연구에 국한되어 있는 척박한 상태이다(김철민, 2015). 따라서, 세계적으로 확대되고 있는 할랄푸드 경향에 맞춰 근본적으로는 할랄인증에 대한 정부의 확대 지원이 필요하며, 음식에 대한 만족이 관광객 전체 만족 및 재방문에 중요한 영향을 미치고 있으므로(Krissoff et al., 2002) 현 시점에서 무슬림들의 할랄 생활을 보장할 수 있도록 하는 시스템적 구비가 필요하다.

## 2.3 머신러닝 적용 식품 분류

제안하는 프레임워크의 학습 부분에 사용되는 머신러닝 기법이 식품을 분류하는 문제에 적용된 연구들이 존재하였다. <표 1>을 살펴보면 이미지 데이터를 기반으로 색상이나 스펙트럼을 이용하여 분류하는 연구(Gomez-Sanchis et al., 2012; Zhu et al., 2010)들이 대부분을 차

<표 1> 머신러닝 적용 식품 분류 기존 연구

데이터	목적	사용모델	사용요인	문헌
이미지	감귤류의 균류 발견	CART, MLP	스펙트럼	Gómez-Sanchis et al., 2012
	꿀의 불순물 검출	SVM, ANN, LDA, KNN	스펙트럼	Zhu et al., 2010
	인삼 묘목의 품질평가	SVM	묘삼 형상, 면적, 길이	오현근 외, 2011
	컴퓨터 비전을 이용한 식품 품질 평가	ANN	Characterization	Du & Sun, 2006
	이미지의 식품 품목 분류	CNN	Color, circular boundary	Kagaya et al., 2014
	음식 이미지 감지 및 음식 양 평가	SVM, NB	색상, 서클 주파수	Aizawa et al., 2013
	토마토 생산 전 자동화된 수율 측정	DTSM	색상, 얼룩, 노출	Yamamoto et al., 2014
	건조 대추의 분류	SVM	프랙탈, 색상	Lou et al., 2012
	망고 열매의 갈변정도 분류	SVM	L, a, b, FD	Zheng & Lu, 2012
	식품 품종 및 원산지 분류	KNN, PLS-DA, ANN, SVM, ELM	스펙트럼	Zheng et al., 2014
식품 섭취 모니터링 프로그램 정확도 향상	SVM	색상, 질감, 크기, 모양, 모두	Pouladzadeh et al., 2013	
비이미지	농업 수확량 예측	J48, RT, NB	밀 작물 수확량, 기상 매개 변수	Alsberg et al, 1997
	당뇨병환자를 위한 식품 추천	SOM, K-means	영양소	Phanich et al, 2010
텍스트	음식 재료, 성분으로 요리 예측	SVM, RF, NB, LR, KNN, DT	요리 재료	Ghewari & Raiyani, 2015

지하고 있었으며, CNN(Convolutional Neural Network)의 모델을 사용한 연구(Kagaya et al., 2014)도 존재하나, CNN은 주로 이미지의 특징을 추출하여 유사점을 찾는 이미지를 인식하여 분류하는 문제에서 사용되므로 텍스트 분류 문제 적용에 어려움이 있다.

비 이미지 기반의 연구로는 영양소와 식품특성을 이용(Phanich et al., 2010)하거나 수확량 및 기상 매개 변수를 이용하여 수율을 예측(Alsberg et al., 1997)하는 연구 등 식품에 관한 속성에 근거하여 분류하는 연구는 있었으나, 텍스트를 이용하여 분류하는 연구는 활발히 이루어지지 않고 있었다. 가장 유사한 텍스트 데이터 기반의 연구로는 음식의 레시피에 포함된

재료 명을 근거로 어떤 식품일 지 예측하는 연구가 존재하였다(Ghewari et al., 2015).

연구의 목적 측면에서 살펴보면 과채류의 적정 품질을 평가 하거나 식품의 선별과 이 과정을 자동화하기 위한 알고리즘을 개발하는 문제에 머신러닝이 사용되고 있었지만, 문제가 될 수 있는 성분을 검출하여 특정집단에 식품을 추천하는 형태의 연구는 클러스터링 기법을 사용하여 제한적으로 이루어지고 있었다. 하지만 식품성분에 따른 섭취가능여부 데이터에 클러스터링 알고리즘을 적용한다면 비슷한 식품성분을 포함하고 있는 과자, 음료수 등과 같이 식품의 종류별로만 군집이 형성되기 때문에 목적이 맞지 않아 적용이 불가능하다. 따라서 식품

성분에 따라 섭취가능 여부를 분류 할 수 있는 문제에 적용이 가능하고, 좋은 성능을 보이는 머신러닝 모델인 SVM(Support Vector Machine), DT(Decision Tree), NB(Naive Bayes)를 이미지가 아닌 식품의 성분을 이용한 분류 문제에 적용한다면 우수한 성능을 기대할 수 있을 것이다.

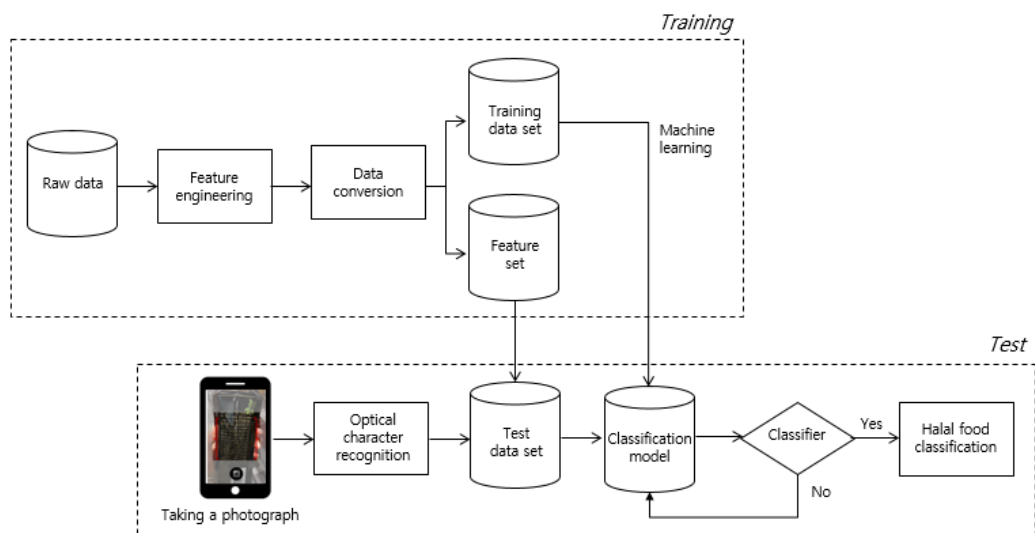
### Ⅲ. 제안 프레임워크

본 연구에서는 기존 어플리케이션을 이용하여 국내에 있는 식품들을 대상으로 검색 시 데이터베이스에 데이터가 존재하지 않아 검색결과가 나오지 않을 경우에 어떠한 정보도 제공하지 않는 문제를 해결하기 위해 프레임워크를 제안한다. 제안하는 프레임워크는 기존의 어플리케이션 시스템에 추가적인 기능의 형태로 머

신러닝 기법을 사용한 지능형 시스템을 이용하여 관광 중 무슬림의 식품의 섭취판별의 의사결정을 도울 수 있는 추천시스템(Ricci, 2002)에 적용이 가능하다.

기존 보유하고 있는 식품데이터에 머신러닝 기법을 적용하여 어떤 속성들이 포함되었을 때 섭취여부에 문제가 되는지 학습시켜 데이터베이스에 존재하지 않는 새로운 데이터가 입력되었을 때 어느 정도의 정확도와 함께 섭취가능 판별에 도움을 줄 수 있는 식품분류결과에 대한 정보를 제공하고자 한다.

<그림 1>은 제안 프레임워크의 개요를 보여준다. 구체적으로 제안 프레임워크는 (a)데이터 수집 및 전처리, (b)머신러닝 기법을 적용한 데이터 학습, (c)광학적 문자 인식, (d)섭취가능여부를 예측하는 부분으로 구성되어 있다.



<그림 1> 제안 프레임워크 개요



### 3.1 데이터 수집 및 전처리

새로운 식품 데이터가 입력되었을 때 섭취가능여부 예측을 수행하기 위해 시중에 판매되고 있는 식품의 포함된 원재료 데이터와 그에 따라 기존의 어플리케이션을 이용하여 검색한 결과로 나온 섭취가능여부의 데이터를 수집한다.

동일한 원재료지만 세부적인 특성이 다를 경우에는 특성을 분류해서 원재료명을 다르게 기재하였다. 예를 들어, 젤라틴의 경우 소의 피혁을 이용해서 만든 젤라틴일 경우와 돼지의 피혁을 이용해서 만든 경우를 구분하여 원재료명을 기재한다. 식품을 구성하는 첨가물을 속성벡터로 만드는 과정에서 너무 많은 속성을 이용할 경우에 분류의 성능에 영향을 줄 수 있기 때문에 속성의 제거와 통합의 과정을 거쳐 차원 축소를 하는 작업이 필요하다(정성욱 외, 2015). 식품 별로 중복된 속성은 한 개만 남기고 나머지를 제거하며, 기본단위의 재료를 가공하여 파생된 성분일 경우에는 기존 원재료로 통합한다. 예를 들어, ‘감자전분, 건 감자, 감자분말을 감자’로 통합한다. 또한, 같은 첨가물이지만 식품 별로 상이하게 기재된 속성은 공통의 속성 명으로 통합한다. 예를 들어, ‘효소제제, 효소제, 효소의 경우 효소’로 통합한다. 속성벡터를 구성한 후 용이한 분석을 위해 한 식품에 대해 첨가물의 포함여부를 포함한 경우에는 1, 포함되지 않은 경우에는 0으로 표현해주는 이진법 형태의 전처리의 과정이 필요하다.

### 3.2 머신러닝 기반의 분류기 학습

데이터의 수집과 전처리과정을 거친 데이터

를 예측수행에 사용하기 위해서는 미리 데이터를 학습시키는 과정이 필요하다. 기존 데이터들이 어떤 첨가물이 포함되었을 때 섭취가능여부에 영향을 미쳤는지에 대해 레이블이 있는 데이터를 이용해 머신러닝 방법을 적용하여 사전 학습을 시킨다. 이 과정에서 식품의 섭취가능여부를 분류하기 위한 방법으로 머신러닝의 대표적인 기법인 SVM, DT, NB의 세 가지 모델을 이용하고자 한다. 기존의 어플리케이션으로 섭취가능여부를 알 수 있는 다량의 식품데이터를 수집하는 것이 어렵기 때문에 위의 모델들을 이용한다면 상대적으로 적은 수의 학습데이터로도 효율적인 분류 성능을 기대(김상연 & 김경범, 2016)할 수 있다.

#### 3.2.1 SVM

SVM은 통계학습 기반의 분류 알고리즘으로, 실제 문제에 적용했을 때 뛰어난 성능을 보인다. 분류, 패턴인식, 회귀 등 다양한 문제해결을 위해 사용되고 있다(Burges & Christopher, 1998). 이차원 데이터의 분류문제에서 집단을 잘 분류하기 위해서 경계의 결정면으로 가장 최적의 초평면의 방정식을 구한다. 최적의 초평면은 두 집단에 있는 데이터들로부터 큰 차이를 가지는 것이 좋으며, 이는 최대마진을 갖는 것을 의미한다. 하지만 실제 문제에 적용하였을 때, 데이터가 선형으로 구분이 안 되는 경우가 발생하는 문제를 해결하기 위해 더 높은 차원으로 대응시켜 분류하기 위해 커널함수를 이용한다. 커널함수의 특징에 따라 분류의 결과가 달라질 수 있으므로, 문제의 목적에 맞게 적절한 커널함수를 선택해야 한다. 대표적인 커널함수로는 Linear, Radial, Sigmoid, Polynomial 등

이 있으며, 이 중에서도 본 연구의 목적에 적합한 Linear와 Radial을 실험에 적용한다.

### 3.2.2 DT

DT는 귀납적 학습기반의 알고리즘으로, 학습데이터로부터 의사결정과정을 나무형태로 시각화가 가능한 분석방법이다. DT는 여러 예측변수에 근거하여 목표변수의 범주를 몇 개의 소집단으로 분류하는 지도 분류 학습의 문제나, 자료로부터 규칙을 찾아내고 이를 통해 알고자 하는 것을 예측하는 문제 등에 사용된다(최종 후 외, 1998). 분석자가 의사결정과정을 쉽게 이해하고 설명할 수 있다는 장점이 있다. DT는 속성을 구분 짓는 기준이 되는 마디들로 구성 되어 있으며, 뿌리마디로부터 시작하여 가지를 쳐나가며 자식마디를 형성하며 끝마디에 이르러 나무가 완성된다. 뿌리마디로부터 끝마디까지의 분리단계를 깊이라고 한다. DT의 알고리즘으로는 ID3(Iterative Dichotomiser 3), C4.5(successor of ID3), C5.0(successor of ID4), CART(Classification And Regression Tree), CHAID(Chi-squared Automatic Interaction Detector) 등이 있다.

DT의 학습에서 각 노드에서 분기하기 위한 최적의 질문은 정보이득(IG, Information Gain)이 최대가 되도록 만들어주는 것이 핵심이다. IG는 자식노드의 데이터가 제대로 분류되지 않고 섞여 있는 정도를 의미하는 데이터의 불순도가 작으면 작을수록 커지게 된다. 어느 특정 노드에서  $m$ 개의 자식 노드로 분기되는 경우 정보이득은 식 (1)로 정의한다.

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j). \quad (1)$$

이때,  $D_p$ 는 부모노드에 존재하는 데이터 세트이고,  $D_j$ 는  $j$ 번째 자식노드에 존재하는 데이터 세트이다.  $I(D_p)$ 는  $D_p$ 의 불순도이며,  $I(D_j)$ 는  $D_j$ 의 불순도이다.  $N_p$ 는  $D_p$ 의 데이터 세트에 속하는 데이터 개수이고,  $N_j$ 는  $D_j$ 의 데이터 세트에 속하는 데이터의 개수이다.

### 3.2.3 NB

NB는 특성들 간의 독립을 가정하는 조건부 확률을 사용하는 베이지 정리를 적용한 통계적 분류 알고리즘으로, 지도 학습에서 효율적으로 훈련이 가능하다. 텍스트 분류에서 여러 범주 중 하나를 결정하는 문제에서 의사결정나무와 비슷한 성능을 보이며, 적절한 전처리의 과정을 거치면 SVM과 견줄 수 있을 성능을 보인다(Rish & Irina, 2001). 주어진 문서에서 그것이 각 카테고리에 할당된 확률을 계산할 때 베이지 정리를 이용한다.

각 카테고리에 분류될 확률은 식 (2)와 같이 계산한다.

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}. \quad (2)$$

이때,  $x$ 는 입력데이터를  $C_k$ 는 분류되는  $k$ 번째 카테고리를 의미한다.

## 3.3 광학적 문자 인식

기존 어플리케이션의 데이터 입력방식은 바

코드 스캔이나 바코드 번호를 타이핑하는 형태로 검색이 이루어졌다. 하지만 이 방법은 데이터베이스에 식품에 대한 원재료 데이터가 존재할 때 이용이 가능한 검색방법이다. 따라서 새로운 데이터의 입력방식으로 식품 후면에 기재되어 있는 원재료 표기부분을 데이터로 받아들일 수 있는 광학적 문자 인식의 기술을 적용하고자 한다. 광학적 문자 인식이란 빛을 이용하여 문자를 판별하여 자동으로 컴퓨터에 입력되는 방법이다. 현재 한국어의 광학적 문자인식 정도를 향상시키는 연구가 꾸준히 이루어지고 있다(권재욱 외, 1992). 인쇄체 한국문자 인식률은 98.28%로 상당히 높은 인식률(조성배 & 김진형, 1990)을 보이고 있다.

광학적 문자 인식의 과정을 자세하게 살펴보면 식품의 뒷면에 식품을 구성하고 있는 원재료성분들이 표기되어 있는 부분을 사용자가 휴대폰을 이용하여 촬영한다. 촬영한 사진에 기재되어 있는 원재료를 텍스트 데이터로 받아들이는 광학적 문자 인식의 과정을 거친다. 원재료를 쉽표 단위로 구분하여(이경호, 2001) 각각의 원재료를 통합의 과정을 거쳐 기존에 구성해둔 속성벡터와 대응시키는데, 이에 대한 규칙은 쉽표가 출현하기 전 줄 바꿈이 있을 경우 붙여서 인식하는 것을 원칙으로 하고(정영미 & 이재운, 1999), 통합된 속성벡터가 되기 전 원 데이터를 원재료 사전으로 만들어 사전에 의거하여 대응시킨다(송광호 외, 2014). 이후 속성벡터에 따라 식품의 원재료 포함여부를 이진형태로 표현해준다. 위의 과정으로 전처리된 데이터를 바탕으로 머신러닝을 통해 프레임워크의 최종과정이인 섭취 가능여부 예측에 이용한다.

### 3.4 섭취 가능 여부 예측

기존의 식품 데이터를 머신러닝을 적용하여 학습시킨 후 분류모델을 생성한다. 새로운 데이터가 OCR의 과정을 거쳐 들어왔을 때 머신러닝의 기법이 적용된 학습된 분류모델을 통해 할랄푸드 혹은 하람푸드에 속할 정도를 계산하고 이들의 비교를 통해 섭취 가능여부를 판단한다.

## IV. 실험 수행

### 4.1 실험 데이터

제안하는 프레임워크의 예측부분의 성능을 평가하기 위해 실제 국내·해외 식품데이터를 이용하여 실험을 수행하였다. 데이터의 수집은 2017년 4월 10일 부터 2017년 7월 30일까지 약 삼 개월 동안 수행하였다. 국내 식품데이터는 실제 식품 후면의 바코드 스캔을 통해 'HalalKorea'의 어플리케이션으로 검색하여 섭취 가능 여부를 수집하였다. 해외 식품데이터는 'Scanhalal' 어플리케이션으로 해외 사용자들이 실시간으로 검색하는 식품에 대한 원재료와 섭취 가능여부에 대한 데이터를 수집하였다.

<표 2>는 수집한 데이터에 대해 요약한 표이다. 수집한 데이터는 섭취가 가능한 식품이 488개, 섭취가 불가능한 식품이 412개로 비슷한 비율로 수집하여 데이터가 편향되지 않도록 하였다. 또한, 국내를 지원하는 할랄푸드 검색 어플리케이션이 활성화되지 않아 식품 데이터의 수집이 어려운 관계로 해외 식품 데이터를 번역

<표 2> 실험 데이터 요약 (식품의약품안전처, 2016)

	국가 별		섭취가능 여부 별	
	국내 데이터	454	섭취가능	488
해외 데이터	446	섭취불가능	412	
식품유형 별				
과자류	농산가공식품류	당류	면류	
190	38	9	75	
별꽃	빙과류	빵류	수산가공식품류	
1	18	21	9	
식용유지류	식육가공품류	유가공품류	음료류	
7	13	70	12	
잼류	조미식품류	즉석식품류	초콜릿류	
5	40	48	59	
캔디류	특수용도식품류			
169	6			
총	900			

하여 이용하였다.

식품의 원재료 데이터는 식품의 후면에 원재료 표기부분을 이용하였으며, 원재료 표기부분을 통합하여 속성벡터를 구성하였다. 원재료를 구성하고 있는 변수로는 구아검, 구연산나트륨, 소르비톨, 아스파탐 등의 화학 성분과 녹차, 생강, 옥수수, 토마토 등의 천연 성분들로 구성되어 있다. 앞에서 언급했듯이 식품데이터에 비해 너무 많은 차원의 원재료 속성 벡터는 분류 성능에 영향을 미칠 수 있으므로, 속성의 통합과 제거의 과정을 거쳤다. 이 과정에서 한 식품에만 존재하여 섭취가능여부에 영향을 거의 미치지 않는다고 판단되는 원재료 성분은 제거하였다.

식품의 섭취가능 여부 데이터는 지도학습을 이용하기 위해 할랄푸드 여부가 명시된 데이터를 획득해야 하므로 기존 어플리케이션을 통해 식품검색 하여 섭취가능 여부를 확인할 수 있는 데이터를 수집하였다. 데이터의 신뢰성 확보를 위해 오류가 있는 데이터를 필터링 하고자

전문가가 한차례 추가 검증을 거친 데이터만을 이용하였다.

#### 4.2 실험 환경

모델 별 실험의 성능을 평가하기 위한 지표로는 Precision, Recall, F1-Score의 값을 이용하여 판단하였다. 분류기의 성능 평가 지표로서 널리 사용되는 Accuracy는 전체 결과 중 정답과 오답을 모두 포함하여 시스템이 예측한 답과 실제 답이 일치한 비율을 나타내어 전체적인 분류에 대한 수행도를 판단하는 지표이다. 데이터에 불균형이 있는 경우 분류기가 전체를 특정 클래스로 예측하였을 때 분류기의 성능이 좋다고 해석이 왜곡될 수 있기 때문에 본 실험의 평가 지표로 이용하지 않았다. <표 3>의 평가지표 구성 요소를 바탕으로 수식으로 정리하였다.

<표 3> 평가지표 구성 요소

		실제 정답	
		Positive	Negative
실험결과	Positive	True Positive( <i>tp</i> )	False Positive( <i>fp</i> )
	Negative	False Negative( <i>fn</i> )	True Negative( <i>tn</i> )

Precision과 Recall은 각각 전체 시스템이 예측한 답과 실제 정답과의 비율, 실제 답과 시스템이 맞춘 답과의 비율을 의미하며 식 (3)과 (4)로 표현할 수 있다.

$$Precision = \frac{tp}{tp + fp} \quad (3)$$

$$Recall = \frac{tp}{tp + fn} \quad (4)$$

F1-Score는 Precision과 Recall의 조화평균으로 Precision과 Recall의 트레이드오프를 반영하여 하나의 값으로 나타낸 정확도이다. 본 실험에서는 Precision과 Recall의 가중치를 동일하게 1을 부여하였으며 그에 관한 수식은 식 (5)이다.

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

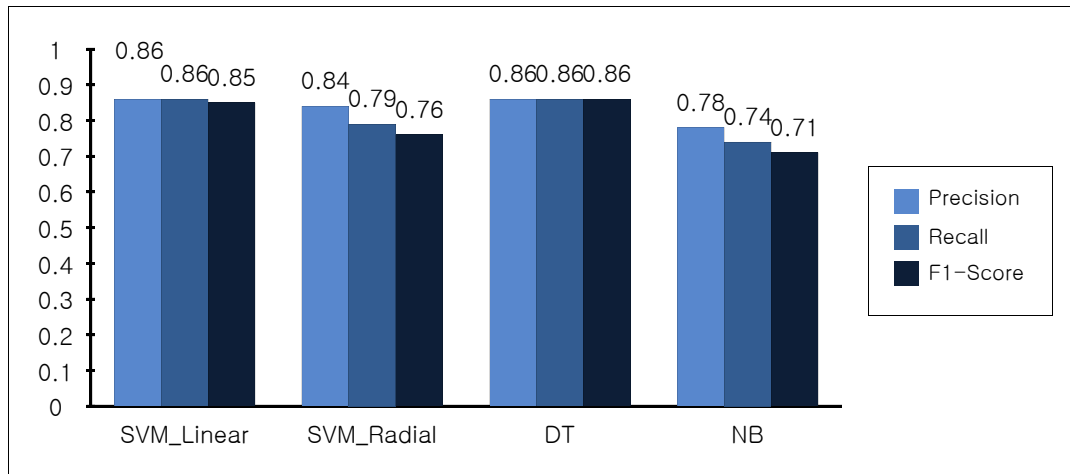
이를 통해 머신러닝 기법을 적용하여 원재료 포함여부에 따라 섭취가능, 불가능을 분류하여 예측하는 과정을 수행하였다. 학습 및 테스트 데이터의 비율을 90%와 10%로 적용하였다. 비교 모델로는 Linear kernel과 Radial kernel을 사용한 SVM(SVM\_Linear, SVM\_Radial), DT, NB를 이용하여 실험을 수행하였다. 제안하는 프레임워크의 구현은 Python을 이용하였으며, 세 종류의 분류 모델의 학습 및 테스트는

Scikit-learn API를 사용하여 이루어졌다.

### 4.3 실험 결과

네 가지의 모델 별로 실험을 수행한 결과를 <그림 2>의 그래프로 도식화 하였다.

SVM\_Linear와 DT 모델의 성능이 각 평가 지표 별로 고르게 0.85 이상의 우수한 분류성능을 보였다. 성능지표에 따른 결과값에 대한 해석으로는 실제 서비스로 제공될 시 분류기가 섭취가능으로 예측한 결과에 대해 사용자가 신뢰를 가지고 대부분 이를 섭취한다면 Precision 기준으로 0.86의 수치를 도출한 SVM\_Linear와 DT가 동일하게 적합하다. Recall 값도 마찬가지로 0.86으로 SVM\_Linear와 DT가 동일하므로 Precision과 Recall의 성능지표 기준으로 두 모델 간의 차이는 없다. 그러나, Precision과 Recall 성능을 동시에 나타내는 지표인 F1-Score 기준으로는 0.01의 차이로 DT의 성능이 미세하게 우수함을 알 수 있었다. 시스템 자체의 성능을 평가식품 개수에 비해 속성이 많은 데이터의 특성상 노이즈가 많이 발생할 수 있지만, SVM\_Linear 모델은 노이즈에 영향을 많이 받지 않기 때문에 식품데이터의 분류기로 적합하고, DT 모델은 특징의 유형에 상관없이 잘 동작하며 이상치에 상대적으로 덜 민감하여 좋은 성능을 보인 것으로 해석이 가능하다. NB는 보통 문헌 분류에 사용이 많이 되는 모델로 목적이 부합하지 않아 좋지 못한 성능을 보인



<그림 2> 모델 별 성능 비교 결과

것으로 추정된다.

추가적으로, Paired *t*-test 검증을 이용하여 모델 별 성능 차이가 통계적으로도 유의한지 확인해보았다. 귀무가설(H0)와 대립가설(H1)은 아래와 같다.

H0 : 모델 간 정확도의 차이가 없다.

H1 : 모델 간 정확도의 차이가 있다.

귀무가설은 ‘모델 간 정확도의 차이가 없다.’, 대립가설은 ‘모델 간 정확도의 차이가 있다.’로 두고 30개의 표본을 랜덤 샘플링하여 각 모델 끼리 Paired *t*-test를 수행하였다. SVM\_Radial은 평가지표 별 변동이 큰 관계로 의미 있는 데이터가 아니므로, SVM은 Linear의 커널만 이용하였고, DT, NB 세 가지의 모델 간의 조합으로 유의성을 검정하였다. 그 결과, SVM\_Linear와 DT를 비교해 보았을 때의 *p*값은 0.591로 유의하지 않았다. 따라서 귀무가설을 기각할 수 있으며, 모델 간 정확도의 차이가 있다고 할 수 없다. DT와 NB, SVM\_Linear와 NB를 비교하였을 때 *p*값은 각각 0.000으로 귀무가설을 채택하며, 결론적으로 모델 간 정확도의 차이가 있

다고 설명할 수 있다. 즉, SVM\_Linear와 DT는 성능이 동일하고 그 다음 NB의 순으로 정확도가 우수하다고 할 수 있다.

## V. 결론

### 5.1 연구 요약

본 연구에서는 무슬림의 관광 중 어플리케이션을 이용하여 식품의 섭취가능여부에 대한 검색 시 검색이 불가능하여 아무런 정보를 획득하지 못해 발생하는 음식부분의 불만족을 개선하기 위해 머신러닝 기법이 적용된 프레임워크로서, 기존 시스템에 추가적인 기능을 더하는 형식으로 적용될 수 있다. 프레임워크 중 학습부분의 머신러닝의 적용이 타당한 지에 대해서 평가하기 위해 머신러닝의 대표적인 모델의 성능비교 실험을 수행하였다. 그 결과 SVM\_Linear과 DT 모델의 성능이 가장 우수하였으며, 통계적인 검정으로도 유의미한 결론을 도출

할 수 있었다. 따라서 텍스트 데이터 기반의 문제가 되는 성분 검출을 통한 할랄푸드 판별문제에 위의 모델이 적합함을 알 수 있었다. 제안하는 프레임워크로 무슬림의 스마트 관광 중 종교 생활의 편의를 제공하여 한국 관광에 대한 인식 제고가 기대되며, 이로 인한 구전효과 확산으로 한국에 대한 관심증진과 재방문율을 높이고, 관광업계의 경제 및 다양한 부가가치 창출에 따른 발전을 기대한다.

## 5.2 연구 시사점 및 한계점

데이터 수집 및 전처리 부분에 언급한 동일한 원재료지만 세부적인 특성이 다른 경우에 정확도가 떨어진 결과를 제공한다는 한계가 존재한다. 특히, 육류가 포함된 식품일 경우 꾸란에 근거한 도축과정을 따르지 않은 육류일 가능성이 존재하지만 제품이 만들어지는 원재료부터 소비자에게 전달되는 전체적인 Supply chain까지 반영된 할랄푸드의 정보는 획득할 수 없어 세부적인 위험요소를 고려 할 수 없었다는 점이 한계이다. 머신러닝 기법은 데이터가 많을수록 학습이 잘 이루어져 더 높은 정확도가 기대되기 때문에 추가 데이터 수집을 통해 빅데이터를 구축하여 예측 정확도의 향상을 도모해야 한다. 또한, 현재는 국내식품 수집의 어려움으로 인하여 해외식품 데이터와 비슷한 비율로 수집하였으나, 제안하는 프레임워크를 국내에서 적용하기 위해서는 국내 식품 위주의 데이터 수집이 추가적으로 필요하다. 할랄푸드에 대해서 무슬림 별로 엄격정도가 다양하기 때문에 '바닐라' 성분의 경우 알코올 성분이 극소량 함유되어 있어 음식에 대해 엄격한 기준

을 적용하는 무슬림에게 바닐라성분이 섭취가능여부에 큰 영향을 미친다. 따라서 제안하는 프레임워크를 다양한 무슬림에게 범용으로 적용하기 위해서는 현재 '먹어도 될 것이다 혹은 아니다.' 의 두 가지 제안의 이진 분류 문제에서 그치지 않고 다분류 문제로 발전시켜 세분화된 섭취가이드라인 제공을 위한 연구가 요구된다. 성능 극대화를 위해 식품의 분류와 같은 식품의 다른 정보를 모델에 반영할 수 있을 것이다. 다분류 문제는 고도로 숙련된 지식을 바탕으로 반지도학습을 통한 딥러닝을 이용하여 식품 분류에 대한 연구를 확장할 필요가 있다.

또한, 할랄은 최근에는 식품 부문에만 국한되는 것이 아니라 다양한 제조 및 서비스 부문에 적용되고 있다. 따라서 머신러닝 기법을 사용한 프레임워크를 식품분야에만 적용하는 것이 아니라 할랄산업에 포함될 수 있는 화장품, 개인 위생용품, 의약품 등과 같은 공산품들을 대상으로 적용범위를 확대하고자 한다.

궁극적으로 무슬림을 위한 스마트 관광 생태계를 조성하기 위해 1차 공급업체인 식음료 제조업체는 안전한 식품으로 인식되고 있어 비 무슬림들도 주목하고 있는 할랄푸드에 대한 연구가 필요하며, 정부차원에서는 방한 무슬림 관광객 증대에 대한 할랄인증 식품 확대를 위한 기업체에 대한 지원과 기도실, 할랄음식점 등의 전국적인 인프라 구축 사업추진과 같은 제도적 대응이 필요하다. 또한, 우리나라는 종교의 자유가 헌법에 보장되어 있다. 각 지역사회와 종교단체가 정부의 무슬림을 위한 정책을 종교의 지원의 차원에서 보는 일부 시각이 존재하지만, 종교자체를 지원하는 것이 아닌 방한 무슬림들의 종교적 신념을 유지할 수 있도록 존중하고,

편의를 제공하여 더 많은 관광에 대한 부가가치를 창출하고자 하는 것이 목적이다. 따라서 실리적 판단을 통해 종교 배척 보다는 관광 다변화의 일환으로 무슬림을 수용해야 할 필요가 있다. 이에 대해 여행 유통업체에서는 무슬림 시장에 대한 이해를 바탕으로 적극적인 홍보가 필요하다. 최종적으로, 4차 산업혁명에서 주목 받고 있는 기술인 머신러닝과 관광분야와의 융합을 통해 스마트 관광 생태계 조성에 긍정적인 기여를 시사한다.

## 참고문헌

- 권재욱, 조성배, 김진형, “계층적 신경망을 이용한 다중 크기의 다중활자체 한글문서 인식,” (구) 정보과학회논문지, 제19권, 제1호, 1992, pp.69-79.
- 김상연, 김경범, “SVM 기반 실리콘 웨이퍼 마이크로크랙의 분류성능 분석,” 한국정밀공학회지, 제33권, 제9호, 2016, pp.715-721.
- 김은미, “할랄 식품을 둘러싼 국내외 시장현황,” 식품과학과 산업, 제48권, 제2호, 2015, pp.12-24.
- 김종도, 최영길, “신앙과 음식: 이슬람 음식법에 관한 연구-꾸란을 중심으로-,” 한국중동학회논문총, 제34권, 제4호, 2014, pp.183-204.
- 김철민, “할랄 식품시장의 의의와 동향,” 세계농업, 제175호, 2015.
- 배영희, “글로벌헬스케어산업 활성화에 따른 방한 무슬림환자의 의료서비스에 관한 개관-영양서비스를 중심으로,” 대한영양사협회 학술지, 제21권, 제4호, 2015, pp.333-341.
- 석수선, “한국 의료관광산업의 환자 유치 활성화를 위한 디자인 전략,” 디지털디자인학연구, 제16권, 제1호, 2016, pp.1-10.
- 서민교, “우리나라 무슬림 관광시장 개발에 관한 탐색적 연구,” 동북아관광연구, 제13권, 제1호, 2017, pp.1-18.
- 송광호, 민지홍, 이가영, 김유성, “유어의 사진을 이용한 문서 내 표절 구간 탐색 시스템 개발,” 한국정보과학회, 2014년 동계학술회발표논문집, 2014, pp.1385-1387.
- 식품의약품안전처, “식품공전 전부개정 유형분류표”, 2016
- 엄익란, “이슬람 마케팅과 할랄 비즈니스”, 도서출판 한울, 2014, pp.183-190.
- 양성수, 허향진, 최병길, “온라인 구전 관광정보의 구전효과,” 관광학연구, 제32권, 제2호, 2008, pp.109-130.
- 이경호, “자연 언어 처리와 문장 부호,” 국제어문, 제24호, 2001, pp.1-19.
- 이석호, “한국 인바운드 시장의 다변화 전략,” 한국관광정책, 제63호, 2016, pp.85-89.
- 이슬기, 이미경, “블루오션으로서의 무슬림시장과 할랄에 대한 국내 인식 연구 : 빅데이터를 이용한 ‘할랄’ 네트워크 분석,” 관광연구, 제32권 제2호, 2017, pp.493-513.
- 이용규, 김효정, “기존 시스템 사용 습관과 새로운 시스템 수용과의 관계,” 정보시스템연구, 제22권, 제1호, 2013, pp.143-161.



- 이희열, 이선경, “이슬람 관광시장 확대에 따른 우리 관광산업 발전방안에 관한 연구,” 한국이슬람학회 논총, 제24권, 제3호, 2014, pp.27-60.
- 정성욱, 김준우, “단체 건강검진 운영 관리를 위한 통합관리 시스템의 설계 및 구현,” 정보시스템연구, 제24권, 제2호, 2015, pp.163-188.
- 정영미, 이재운, “한국어 텍스트 처리를 위한 줄 경계 띄어쓰기 복원,” 한국정보관리학회, 제6회 학술대회 논문집, 1999, pp.21-24.
- 조성배, 김진형, “한글 문자인식을 위한 신경망 기법의 개선에 관한 연구,” 한국통신학회 워크샵, 1990, pp.56-62.
- 조운선, “2017년 관광인 신년인사회 축사”, 2017.1.
- 채경연, 이희열, “방한 이슬람 관광시장 증진을 위한 할랄 투어리즘 도입방안에 관한 연구,” 한국외식산업학회지, 제11권, 제2호, 2015, pp.95-103.
- 최영길, “이슬람에서 허용된 음식과 금기된 음식,” 인문과학연구논총, 제16호, 1997, pp.299-317.
- 최종후, 한상태, 강현철, 김은석, “Answer Tree 를 이용한 데이터마이닝 의사결정나무 분석,” SPSS 아카데미, 1998, pp.17-23.
- 한국관광공사, “무슬림 관광객 유치의 경제적 효과 추정에 관한 연구”, 2015.
- 한국관광공사, “무슬림 관광객 유치 안내서”, 2016a.
- 한국관광공사, “외래관광객 모바일인터넷 이용 실태조사”, 2014.
- 한국관광공사, “한국관광공사사업계획발표”, 2017.02.
- 한국관광공사, “2016년 방한 무슬림 관광실태 조사”, 2016b.
- 홍완수, “할랄 식품 산업과 할랄 인증,” 식품과 학과 산업, 제48권, 제2호, 2015, pp.2-11.
- Aizawa, K., Maruyama, Y., Li, H., and Morikawa, C., “Food balance estimation by using personal dietary tendencies in a multimedia food log,” *IEEE Transactions on Multimedia*, Vol.15, No.8, 2013, pp.2176-2185.
- Alsberg, B. K., Goodacre, R., Rowland, J. J., Kell, D. B., “Classification of pyrolysis mass spectra by fuzzy multivariate rule induction-comparison with regression, K-nearest neighbor, neural and decision-tree methods,” *Analytica Chimica Acta*, Vol.348, No.1, 1997, pp.389-407.
- Aziz, H., “The journey: an overview of tourism and travel in the Arab/Islamic context,” *Tourism and the less developed world: Issues and Case Studies*, 2001, pp.151-159.
- Bhardwaj, S., “Non-hajj pilgrimage in Islam: A neglected dimension of religious circulation,” *Journal of Cultural Geography*, Vol.17, No.2, 1998, pp.69-87.
- Bonne, K., Verbeke, W., “Muslim consumer’s

- attitude towards meat consumption in Belgium: insights from a means-end chain approach,” *Anthropology of Food*, Vol.5, 2006, pp.1724.
- Burges, Christopher JC. “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, Vol. 2, No.2, 1998, pp.121-167.
- Du, C. J., Sun, D. W., “Learning techniques used in computer vision for food quality evaluation: a review,” *Journal of Food Engineering*, Vol.72, No.1, 2006, pp.39-55.
- Gomez-Sanchis, J., Martin-Guerrero, J. D., Soria-Olivas, E., Martinez-Sober, M., Magdalena-Benedito, R., Blasco, J., “Detecting rottenness caused by *Penicillium* genus fungi in citrus fruits using machine learning techniques,” *Expert Systems with Applications*, Vol.39, No.1, 2012, pp.780-785.
- Hanzaee, K. H., Ramezani, M. R., “Intention to Halal Products In The World Markets,” *Interdisciplinary Journal of Research in Business*, Vol.1, No.5, 2011, pp.1-7.
- Hasharina, N., Hashim, N., Murphy, J., Muhammad, N., “Tourism and Islam, understanding and embracing the opportunity,” *Inside Out*, 2006, pp.11-15.
- Henderson, J. C., “Muslim Travellers, Tourism Industry Responses and the Case of Japan,” *Tourism Recreation Research*, Vol.41, No.3, 2016, pp.339-347.
- Kagaya, H., Aizawa, K., Ogawa, M., “Food detection and recognition using convolutional neural network,” *Proceedings of the ACM international conference on Multimedia*, 2014, pp.1085-1088.
- Kamali, M. H., “Principles of Islamic Jurisprudence,” *Islamic Texts Society*, 2003.
- Kessler, C. S., “Pilgrims’ progress: The travelers of Islam: Muslim Travellers: Pilgrimage, Migration, and the Religious Imagination,” *Annals of Tourism Research*, Vol.19, No.1, 1992, pp.147-153.
- Krissoff, B., Barry, K., Bohman, M., Mary, Caswell, J. A., Julie, A., “Global Food Trade and Consumer Demand for Quality,” *Kluwer Academic*, 2002.
- Lou, H., Hu, Y., Wang, B., Lu, H., “Dried jujube classification using support vector machine based on fractal parameters and red, green and blue intensity,” *International Journal of Food Science & Technology*, Vol.47, No.9, 2012, pp.1951-1957.
- Phanich, M., Pholkul, P., Phimoltares, S., “Food recommendation system using clustering analysis for diabetic patients,” *Proceedings of the International Conference on IEEE Information*

- Science and Applications*, 2010, pp. 1-8.
- Pouladzadeh, P., Shirmohammadi, S., Arici, T., "Intelligent SVM based food intake measurement system," *Proceedings of the International Conference on IEEE Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, 2013, pp. 87-92.
- Ricci, F., "Travel recommender systems," *IEEE Intelligent Systems*, Vol.17, No.6, 2002, pp.55-57.
- Rish, I., "An empirical study of the naive Bayes classifier," *IBM, IJCAI workshop on empirical methods in artificial intelligence*, Vol. 3. No.22, 2001, pp.41-46.
- Samuel, A. L., "Some Studies in Machine Learning Using the Game of Checkers," *Journal of Research and Development*, Vol.3, No.3, 1959, pp.210.
- Werthner, H., Ricci, F., "E-commerce and tourism," *Communications of the ACM*, Vol.47, No.12, 2004, pp.101-105.
- Yamamoto, K., Guo, W., Yoshioka, Y., Ninomiya, S., "On plant detection of intact tomato fruits using image analysis and machine learning methods," *Sensors*, Vol.14, No.7, 2014, pp.12191-12206.
- Zheng, H., Lu, H., "A least-squares support vector machine (LS-SVM) based on fractal analysis and CIELab parameters for the detection of browning degree on mango (*Mangifera indica* L.)," *Computers and Electronics in Agriculture*, Vol.83, 2012, pp.47-51.
- Zheng, W., Fu, X., Ying, Y., "Spectroscopy-based food classification with extreme learning machine," *Chemometrics and Intelligent Laboratory Systems*, Vol.139, 2014, pp.42-47.
- Zhu, X., Li, S., Shan, Y., Zhang, Z., Li, G., Su, D., Liu, F., "Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics.," *Journal of Food Engineering*, Vol.101, No.1, 2010, pp.92-97.
- Ghewari, Rishikesh, Sunil Raiyani, "Predicting cuisine from ingredients.," 2015.
- Grim, B. J., and Karim, M. S., "The future of the global Muslim population: projections for 2010-2030." *Washington DC: Pew Research Center*, 2011.
- MasterCard & Crescent Rating, Global Muslim Travel Index(GMTI), 2015.
- MasterCard & Crescent Rating, Global Muslim Travel Index (GMTI), 2016.
- Power, C., "Halal: Buying Muslim.," *Time Magazine*, 2009.

**김 선 아 (Kim, Sun-A)**



경기대학교 산업경영공학과 학사 과정에 있으며, 주요 관심 분야는 인공지능, 자연어 처리, 추천시스템 등이다.

**김 정 원 (Kim, Jeong-Won)**



경기대학교 산업경영공학과 학사 과정에 있으며, 주요 관심 분야는 텍스트 마이닝과 의사결정분석 등이다.

**원 동 연 (Won, Dong-Yeon)**



경기대학교 산업경영공학과 학사를 취득하였다. 현재 경기대학교 산업경영공학과 석사 과정에 있으며, 주요 관심 분야는 신뢰성공학, PHM, 베이 지안 방법론 등이다.

**최 예 림 (Choi, Yerim)**



서울대학교 공학사와 박사 학위를 취득하였다. 현재 경기대학교 산업경영공학과 교수로 재직하고 있으며, 주요 관심 분야는 human modeling, data driven AI, text analysis 등이다.

<Abstract>

## **A Halal Food Classification Framework Using Machine Learning Method for Enhancing Muslim Tourists**

Kim, Sun-A · Kim, Jeong-Won · Won, Dong-Yeon · Choi, Yerim

### **Purpose**

The purpose of this study is to introduce a framework that helps Muslims to determine whether a food can be consumed. It can complement existing Halal food classification services having a difficulty of constructing Halal food database.

### **Design/methodology/approach**

The proposed framework includes two components. First, OCR(Optical Character Recognition) technique is utilized to read the food additive information. Second, machine learning methods were used to trained and predicted to determine whether a food can be consumed using the provided information.

### **Findings**

Among the compared machine learning methods, SVM(Support Vector Machine), DT(Decision Tree), and NB(Naive Bayes), SVM with linear kernel and DT had excellent performance in the Halal food classification. The framework which adopting the proposed framework will enhance the tourism experiences of Muslim tourists who consider keeping the Islamic law most importantly. Furthermore, it can eventually contribute to the enhancement of smart tourism ecosystem.

**Keyword:** Machine learning, Halal food, food classification, Muslim, smart tourism

\* 이 논문은 2017년 8월 29일 접수, 2017년 9월 15일 1차 심사, 2017년 9월 28일 게재 확정되었습니다.