

<http://dx.doi.org/10.17703/JCCT.2022.8.4.347>

JCCT 2022-7-43

머신러닝 기반 외식업 프랜차이즈 가맹점 성패 예측

Prediction of Food Franchise Success and Failure Based on Machine Learning

안예린*, 유성민**, 이현희**, 박민서***

Yelyn Ahn*, Sungmin Ryu**, Hyunhee Lee**, Minseo Park***

요약 외식업은 소비자의 수요가 많고 진입장벽이 낮아 창업이 활발하게 일어난다. 하지만 외식업은 폐업률이 높고, 프랜차이즈의 경우 동일 브랜드 내에서도 매출 편차가 크게 나타난다. 따라서 외식업 프랜차이즈의 폐업을 방지하기 위한 연구가 필요하다. 이를 위해, 본 연구에서는 프랜차이즈 가맹점 매출에 영향을 미치는 요인들을 살펴보고, 도출된 요인들에 머신러닝 기법을 활용하여 프랜차이즈의 성패를 예측하고자 한다. 강남구 프랜차이즈 매장의 PoS(Point of Sale) 데이터와 공공데이터를 활용하여 가맹점 매출에 영향을 미치는 여러 요인들을 추출하고, VIF(Variance Inflation Factor)를 활용하여 다중공산성을 제거하여 타당성 있는 변수 선택을 진행한 뒤, 머신러닝 기법 중 분류모델을 활용하여 프랜차이즈 매장의 성패 예측을 진행한다. 이를 통해 최고 정확도 0.92를 가진 프랜차이즈 성패 예측 모델을 제안한다.

주요어 : 외식업 프랜차이즈, 예측 모델링, 성패 예측, 머신러닝, 지도학습

Abstract In the restaurant industry, start-ups are active due to high demand from consumers and low entry barriers. However, the restaurant industry has a high closure rate, and in the case of franchises, there is a large deviation in sales within the same brand. Thus, research is needed to prevent the closure of food franchises. Therefore, this study examines the factors affecting franchise sales and uses machine learning techniques to predict the success and failure of franchises. Various factors that affect franchise sales are extracted by using Point of Sale (PoS) data of food franchise and public data in Gangnam-gu, Seoul. And for more valid variable selection, multicollinearity is removed by using Variance Inflation Factor (VIF). Finally, classification models are used to predict the success and failure of food franchise stores. Through this method, we propose success and failure prediction model for food franchise stores with the accuracy of 0.92.

Key words : Food Franchise, Predictive Modeling, Success and Failure Prediction, Machine Learning, Supervised Learning

*준회원, 서울여자대학교 소프트웨어융합학과 학부생 (제1저자) Received: May 24, 2022 / Revised: June 21, 2022

**준회원, 서울여자대학교 데이터과학전공 학부생 (참여저자) Accepted: July 2, 2022

***정회원, 서울여자대학교 데이터사이언스학과 조교수 (교신저자)

***Corresponding Author: mpark@swu.ac.kr

Dept. of Data Science, Seoul Women's Univ, Korea

접수일: 2022년 5월 24일, 수정완료일: 2022년 6월 21일

게재확정일: 2022년 7월 2일

I. 서론

COVID-19로 인한 경기침체로 청년 구직난이 심화되고 기대수명의 증가와 의도치 않은 조기퇴직으로 인해 청, 장년층의 창업이 증가하였다[1]. 또한 1인 가구가 증가하고, 일명 ‘떡방’과 ‘쿡방’ 같은 음식 관련 미디어가 급증하면서[2] 음식과 관련된 창업 아이템이 주목받고 있다. 실제로 숙박 및 음식점업이 전체 창업기업의 27.3%로 가장 많은 비율을 차지해 외식업 창업이 신규 창업자에게 각광 받고 있다[3]. 하지만 2019 국내외 외식트렌드 조사보고서에 따르면 외식 사업체 개점 후 1년 이내 폐업율이 약 90%에 달한다고 한다[4].

외식업은 소비자의 수요가 많고 진입장벽이 낮아 창업이 활발하게 일어나지만, 그에 비례하게 폐업률도 높은 것으로 나타났다. 이로 인해 사업을 처음 시작하는 창업자들은 상대적으로 실패할 확률이 적은 프랜차이즈 창업을 선호한다[5]. 하지만 외식업의 가맹점 평균 매출액은 전체 평균에 못 미치고 있다. 프랜차이즈의 경우 본사 산하에 가맹점이 운영되는 형태로 동일한 브랜드 내에서는 비슷한 인테리어, 메뉴 구성으로 운영됨에도 불구하고 매출의 편차가 크게 나타난다.

지금까지는 프랜차이즈 가맹점 매출에 영향을 미치는 요인들을 분석하는 연구가 활발히 진행되고 있다. 그러나, 도출된 요인들을 활용하여 프랜차이즈의 성패를 예측하는 연구는 부족한 실정이다. 따라서, 본 연구는 기존 연구에서 도출된 가맹점 매출에 영향을 미치는 요인들을 활용해 성패를 예측하고자 한다. 이를 위해 2019년 2월 강남구 프랜차이즈 K치킨의 PoS(Point of Sale)데이터를 활용하여, 매출 요인을 분석한다. 월 매출을 기준으로 성공과 실패 매장을 분류하여 매장 PoS 데이터 기반으로 머신러닝 알고리즘을 학습시켜 성패를 예측한다.

본 논문의 구성은 다음과 같다. 2장에서는 매출에 영향을 미치는 요인을 살펴보고, 3장에서는 성패예측에 효과적인 알고리즘을 살펴본다. 4장에서는 본 연구에서 제안하는 알고리즘에 대해 설명하고, 5장과 6장에서는 연구 결과 및 결론을 기술한다.

II. 매출에 영향을 미치는 요인

매출에 영향을 미치는 요인에 대한 선행연구를 검토하였다. 인구특성, 입지특성, 경제특성, 매장특성 등이 매출에 영향을 미치는 것으로 나타났다. 특히 물리적 환경특성인 매장의 입지 환경이 매출 성과에 주된 영향을 끼친다는 결과가 보고되었다. 표 1은 매출 데이터의 수집 및 가공이 용이한 편의점, 프랜차이즈 커피전문점 등의 연구 결과를 보여준다.

1. 편의점

편의점을 주제로 한 연구의 경우, 편의점의 매출과 이익에 영향을 미치는 입지요인을 상관분석과 다중회귀분석을 통하여 실증하였다[6][7]. 편의점 매출에 영향을 미치는 요인으로 경쟁점수를 도출하였으며, 인구특성을 함께 고려할 것을 제안하였다. 또한 인구와 매장의 입점 형태가 매출에 영향을 미친다는 결과를 도출하였다[7].

2. 프랜차이즈 커피전문점

프랜차이즈를 주제로 한 경우, 서울 소재 프랜차이즈 커피전문점 117개의 입지특성과 매출액의 관계를 연구하였다. 선정한 변수의 특성을 고려하여 헤도닉 함수의 선형 모형 중 반로그 모형[8]을 적용하여 분석을 수행하였다. 분석 결과 매장면적, 횡단보도까지의 거리, 교통시설, 유흥주점, 대형마트 변수가 통계적으로 유의미한 것으로 나타났다[9].

3. 소매업종(외식업)

3가지 소매업종을 대상으로 진행한 연구에서는 물리적 입지특성이 소매업 매출에 미치는 영향을 GLS (Generalized least squared) 분석[10]을 활용하여 연구하였다. 세 가지 연구 대상 중 외식업의 경우 경쟁점수, 상업지역 면적, 거주인구, 용적률이 매출에 양(+)의 영향을 미치는 것으로 드러났다[11].

대다수의 연구에서 전통적인 방식인 상관분석과 회귀분석을 활용하여 매출에 영향을 미치는 요인을 분석하였다. 그러나 매출 데이터는 대량의 데이터가 빠르게 축적되고 환경에 따라 변하기 때문에, 기존 분석 방법과 더불어 빅데이터의 특성을 고려한 분석 방법을 도입할 필요성이 있다[12].

표 1. 입지요인과 매출에 관한 선행연구

Table 1. Previous literature on location factors and sales

저자	종속변수	요인		영향
이임동 외 (2010)	편의점의 매출, 이익	인구요인	배후인구, 유동인구	1. 75M 상권 내 인구는 매출과 유의한 관계(+)가 있음
		점포요인	점포면적, 점면수, 전면길이, 전면도로유형	1. 도로유형, 점면수, 길이는 유의적이지 않음 2. 점포면적은 매출과 유의한 관계(+)가 있음
		경쟁요인	경쟁점수	1. 매출 및 이익에 가장 영향을 많이 미치는 요인임
		비용요인	월임대료	1. 유의적이지 않음
황규성 외 (2016)	편의점의 일평균 매출	인구요인	거주인수, 상주인구수, 유동인구	1. 배후인구와 매출은 유의한 관계(+)가 있음 2. 유동인구수와 매출은 유의한 관계(+)가 있음
		점포요인	전면 길이, 측면길이, 점면수, 점포면적	1. 점포면적과 매출은 유의한 관계(-)가 있음 2. 점면수는 매출과 유의한 관계(+)가 있으나 일부 점포에서 통계적으로 유의적이지 않음
		입지요인	측면 도로 유형, 입지유형	1. 입지형태와 매출은 유의한 관계(+)가 있음
		경쟁요인	경쟁점수	1. 매출 및 이익에 가장 영향을 많이 미치는 요인임
		비용요인	임대료, 보증금, 권리금	1. 임대료와 매출은 유의한 관계가 있으나 지역마다 편차를 보임
신우진 외 (2011)	프랜차이즈 커피전문점 월평균 매출	점포요인	점포면적, 전면길이, 무선인터넷, 1층 입점, 독립건물, 테라스, 주차공간	1. 점포면적은 매출과 유의한 관계(+)가 있음 2. 이외 특성은 유의적이지 않음
		입지요인	주변시설(유흥시설, 음식점, 학원 등), 횡단보도까지의 거리, 지하철역까지의 거리	1. 횡단보도까지의 거리는 매출과 유의한 관계(+)가 있음 2. 대형마트, 기차역, 버스터미널, 유흥주점의 개수는 매출과 유의한 관계(+)가 있음
성은영 외 (2017)	외식업, 슈퍼마켓, 휴대폰 판매점의 월평균 매출	경제요인	가처분소득	1. 매출과 유의한 관계(+)가 있음
		인구요인	거주인구수, 종사자수	1. 거주인구 및 종사자 수는 외식업 매출과 유의한 관계(+)가 있음
		점포요인	건폐율, 용적률	1. 건폐율과 외식업 매출은 유의한 관계(-)가 있음 2. 용적률과 외식업 매출은 유의한 관계(+)가 있음
		입지요인	지하철역까지의 거리, 버스정류장까지의 거리, 토지이용특성, 도심까지의 거리	1. 주거용도의 면적과 외식업 매출은 유의한 관계(-)가 있음 2. 상업, 업무, 기타용도 면적과 외식업 매출은 유의한 관계(+)가 있음 3. 도심에 가까울수록 매출과 유의한 관계(+)가 있음
		경쟁요인	동종업종 밀도	1. 동종업종의 밀도와 매출은 유의한 관계(+)가 있음

III. 성패 예측에 효과적인 머신러닝 알고리즘

프랜차이즈 성패예측은 성공과 실패의 두 카테고리에 대한 예측이므로 머신러닝 기법 중 분류 모델이 적절하다. 본 장에서는 머신러닝 기법 중 분류 모델에 대해서 살펴보고자 한다.

1. Logistic Regression

Logistic Regression(로지스틱 회귀) 알고리즘은 종속변수가 0과 1의 값을 가지는 이항변수이고 독립변수가 연속 변수일 경우에 주로 사용되는 확률 모델 알고리즘[13]으로, 데이터가 특정 그룹에 속할 확률을 추정 및 예측하는 효과적인 알고리즘이다[14].

2. Decision Tree

Decision Tree(의사결정나무) 알고리즘은 의사결정 규칙을 도표화하여 데이터를 유사한 그룹끼리 세분화하여 여러 소집단으로 분류 및 예측하는 알고리즘[15]이다. 나무 모양 구조로 시각화되기 때문에 다른 분석 알고리즘에 비해 분석 과정을 이해하기 쉽고, 설명이 가능하다는 장점이 있다[16].

3. Random Forest

Random Forest 알고리즘은 단일 Decision Tree만 사용하는 것이 아닌 여러 개의 Tree 조합으로 확장한 메타학습 형태의 분류 및 예측 알고리즘이다[17]. 단일 Decision Tree는 샘플링된 일부의 독립변수와 표본을

이용하여 학습하므로 예측 성능이 낮지만, Random Forest는 다수의 Decision Tree를 조합하여 잡음이 많은 데이터에서도 좋은 성능을 보이며, Feature Importance(변수 중요도)를 계산할 수 있는 장점을 가지고 있다[18].

IV. 프랜차이즈 성패 예측을 위한 머신러닝 모델설계

본 연구에서는 선행 연구에서 밝혀진 매출에 영향을 미치는 요인을 머신러닝 알고리즘에 적용하여 프랜차이즈 성패를 예측하는 모델을 제안한다. 전체 프로세스는 그림 1과 같다.

1. 데이터 수집

서울특별시 강남구에 위치한 프랜차이즈 K치킨 가맹점

33개를 대상으로 점포의 매출 데이터(2019년 2월 PoS 데이터)를 수집하였다. 또한, 선행연구의 매출 영향 요인을 기반으로 접근성, 경쟁점, 입지특성, 인구특성을 반영하여 관련 공공데이터를 수집하였다. 접근성 중 버스정류장 개수를 파악하고자 서울시 열린데이터광장(<https://www.data.seoul.go.kr>)의 버스정류장 좌표정보(2019)를 활용하였고, 지하철역 출구 개수는 산림빅데이터거래소(<https://www.bigdata-forest.kr>)의 서울시 지하철역 출구 정보(2019)를 사용하였다. 경쟁점 데이터를 도출하기 위해 서울시 열린데이터광장의 서울특별시 강남구 일반음식점 인허가 정보(2019)를 사용하였다. 또한, 입지특성을 파악하기 위해 환경부 환경공간정보서비스(<https://egis.me.go.kr/>)의 세분류 토지피복지도(2019)를 사용하였다. 마지막으로 인구특성을 파악하기 위해 서울시 열린데이터광장의 행정동별 서울생활인구

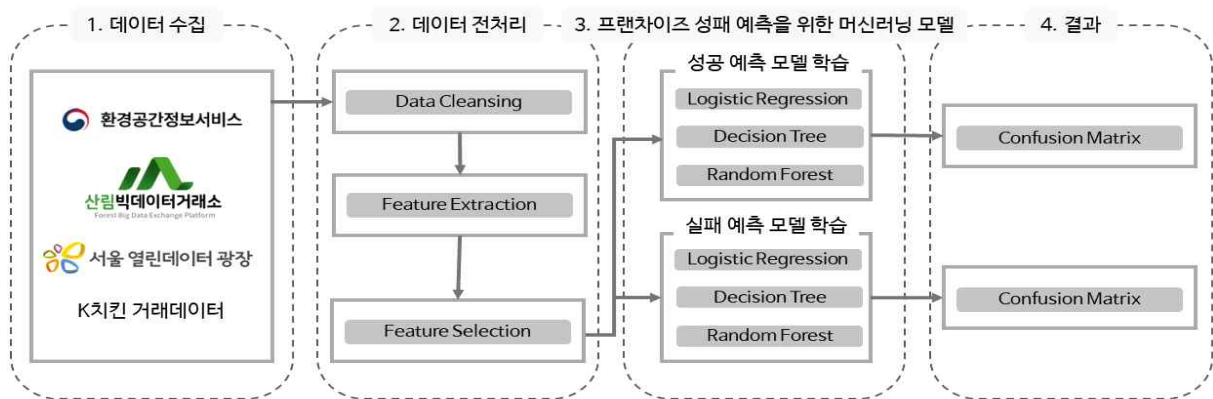


그림 1. 프랜차이즈 성패 예측을 위한 머신러닝 모델

Figure 1. Machine learning model for prediction of franchise success and failure

표 2. 변수의 정의

Table 2. Definition of Variables

변수명		정의	단위	출처(2019년 기준)	
종속 변수	매장의 성공 및 실패	성공매장: (일 매출합계)/(행정동별 일평균 거주인구) > 2500만원 실패매장: (일 매출합계)/(행정동별 일평균 거주인구) < 500만원	성공, 실패	PoS 데이터	
독립 변수	접근성	버스정류장	매장 반경 500m의 버스정류장 개수	개	서울시 열린데이터광장
		지하철역 출구	매장 반경 500m의 지하철역 출구 개수	개	산림빅데이터거래소
	경쟁점	매장 반경 500m의 동종업종 개수	개	서울시 열린데이터광장	
	입지 특성	주거지역 면적	매장 반경 500m의 주거지역 면적 합	km ²	환경부 환경공간정보서비스
		상업지역 면적	매장 반경 500m의 상업지역 면적 합	km ²	
		주거지역 비율	(매장 반경 500m의 주거지역 면적 합계) / (매장 반경 500m의 주거지역 면적, 상업지역 면적의 합계)	-	
	상업지역 비율	(매장 반경 500m의 상업지역 면적 합계) / (매장 반경 500m의 주거지역 면적, 상업지역 면적의 합계)	-		
인구 특성	유동인구 비율	(행정동별 일평균 활동인구) / (행정동별 일평균 거주인구)	-	서울시 열린데이터광장	

(2019)에서 활동인구수를, 서울시 주민등록인구(2019)에서 거주인구수를 구독하였다 (표 2 참조).

2. 데이터 전처리

1) Data Cleansing

다양한 경로를 통해 수집한 좌표정보 데이터들을 EPSG:5179 - Korea 2000 / Unified CS 좌표계 기준으로 모두 변환하였다. 또한, 매장 반경 500m 외부의 데이터를 제거하여 데이터 정제 과정을 진행하였다.

2) Feature Extraction

행정동별 거주인구의 불균형이 매출에 부정적인 영향을 미칠 것을 고려하여 PoS 데이터의 일 매출을 매장별로 합산하여 행정동별 거주인구로 나누어 종속변수로 사용하였다. 1만 인구당 월 매출이 2500만원 이상인 성공 매장(1)과 그 이외의 매장(0), 인구당 월 매출이 500만원 이하인 실패 매장(1)과 그 이외의 매장(0)으로 정의하였다.

독립변수인 접근성, 경쟁점, 입지특성, 인구특성은 모두 QGIS(ver 3.22 Biatowiza)를 사용하여 도출하였다. 매장 반경 500m 이내의 버스정류장 좌표의 개수와 지하철역 출구 좌표의 개수를 계산하여 접근성 변수로 사용하였다. 경쟁점 변수의 경우 강남구 일반음식점 인허가 정보(2019)에서 2019년 2월 당시 강남구 내에서 운영된 동종업종을 추출하여 매장 반경 500m 이내의 동종업종 좌표의 개수를 계산하여 사용하였다. 입지특성은 2019 세분류 토지피복지도에서 중분류를 활용하여 매장 반경 500m 이내의 주거지역 및 상업지역의 면적과 비율을 계산하여 사용하였다. 인구특성은 일 평균 활동인구를 거주인구로 나누어 유동인구 비율로 계산하여 변수로 활용하였다 (표 2 참조).

3) Feature Selection

추출된 독립변수의 판별 타당성을 위해 VIF (Variance Inflation Factor, 분산팽창요인)를 활용하여 변수 간의 영향요인을 제거하였다. VIF는 독립 변수간 상관관계를 측정하는 척도로 일반적으로 10이 넘으면 다중공선성(독립변수 간 상관관계)이 존재하는 것으로 간주한다 [19]. 변수 선택 과정을 통해 최종 선정된 총 7개의 독립변수 간 VIF는 표 3과 같다.

표 3. 최종 선정된 독립변수 간 VIF

Table 3. Variance inflation factors in selected features

변수명	Variance Inflation Factor
버스정류장 개수	1.388
지하철역 개수	1.358
경쟁점 개수	1.608
인구수	1.239
유동인구 비율	1.563
상업지역 면적	1.756
주거지역 면적	1.253

3. 프랜차이즈 성패 예측을 위한 머신러닝 모델

수집된 데이터를 전처리 후, 3:5:6.5 비율로 Train set 과 Test set으로 나누어 학습하였다. 특히, 데이터 탐색을 통해, 성공 매장과 실패 매장을 구분하여, 각각 학습을 진행하였다. 모델 학습에는 이진 분류에 효과적인 Logistic Regression, Decision Tree, Random Forest가 사용되었다. 모델의 성능 평가를 위해, 혼동행렬 사용하여 정확도를 검증하였다. 모델 평가를 통해 최적의 모델을 선택한다.

IV. 결 과

본 연구에서 적용했던 머신러닝 알고리즘(Logistic Regression, Decision Tree, Random Forest)에 대한 예측 성능을 비교 분석하였다. 예측 성능 척도로서 Confusion matrix(혼동 행렬)을 활용하여 정확도를 구하였다[20] (표 4 참조).

표 4. Confusion matrix 요소

Table 4. Elements of confusion matrix

요소	설명
<i>TP(True Positive)</i>	성공(실패) 매장을 성공(실패)할 것이라고 맞게 예측한 것
<i>TN(True Negative)</i>	일반 매장을 일반 매장으로 맞게 예측한 것
<i>FP(False Positive)</i>	성공(실패) 매장을 실패(성공)할 것이라고 잘못 예측한 것
<i>FN(False Negative)</i>	일반 매장을 성공(실패)할 것이라고 잘못 예측한 것

정확도는 전체 매장 중 각 매장의 성패를 정확히 예측한 경우를 의미한다. 수식(1)을 통해 구할 수 있다.

$$Accuracy = \frac{TP+ TN}{TP+ FP+ TN+ FN} \quad \dots (1)$$

표 5. 프랜차이즈 성패 예측 모델별 정확도 비교

Table 5. Comparison of accuracy in franchise success and failure predictive models

알고리즘	예측 변수	Accuracy
Decision Tree	성공	0.67
	실패	0.92
Logistic Regression	성공	0.83
	실패	0.83
Random Forest	성공	0.92
	실패	0.83

표 5는 Logistic Regression, Decision Tree, Random Forest 모델의 예측 정확도를 보여준다. 성공 매장 경우, Random Forest 모델이 정확도 0.92로 가장 우수한 성능을 보였으며 실패 매장의 경우 Decision Tree 모델이 정확도 0.92로 가장 좋은 성능을 보였다.

성공 매장 예측에서 가장 좋은 성능을 보인 Random Forest 모델의 Feature Importance를 확인한 결과는 주거지역 면적이 0.21로 예측 성능에 영향력이 가장 큰 것으로 나타났으며, 다음으로 상업지역 면적이 0.19, 유동인구비율이 0.15, 인구수가 0.12, 지하철역 개수가 0.12, 버스정류장 개수가 0.11, 경쟁점 개수가 0.09 순으로 나타났다 (그림 3 참조).

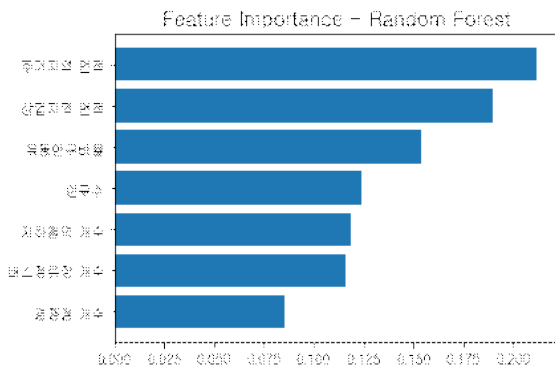


그림 3. 성공 매장 예측 모델(Random Forest)의 Feature Importance
Figure 3. Feature importance of franchise success predictive model by random forest

실패 매장 예측에서 가장 좋은 성능을 보인 Decision Tree 모델의 Feature Importance를 확인한 결과는 지하철역 개수가 0.52로 예측 성능에 영향력이 가장 큰 것으로 나타났으며, 다음으로 인구수가 0.26, 상업지역 면적이 0.22로 나타났다 (그림 4 참조).

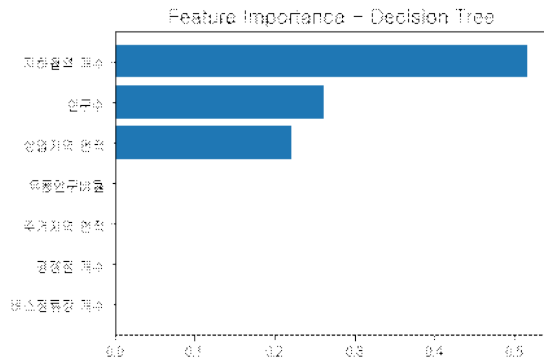


그림 4. 실패 매장 예측 모델(Decision Tree)의 Feature Importance
Figure 4. Feature importance of franchise failure predictive model by decision tree

V. 결론

본 연구는 머신러닝 기법을 활용해 강남구 내 K 프랜차이즈 매장의 성공과 실패를 예측하는 모델을 구축하였다. 예측에 사용된 변수는 입지특성, 접근성, 인구특성을 반영한 다양한 파생 변수를 생성하여 사용했으며 변수 간의 미치는 영향을 줄이기 위해 VIF를 통해 다중공선성을 제거하여 변수 선택을 진행했다. Logistic Regression, Decision Tree, Random Forest를 적용하여 모델을 구축하였으며, 모델의 정확도를 혼동 행렬로 평가하였다. 성공 예측에는 Random Forest가 정확도 0.92로 가장 우수한 성능을 보였고, 실패 예측에는 Decision Tree가 정확도 0.92로 가장 우수한 성능을 보였다. 또한, 예측 성능에 가장 큰 영향을 주는 요인으로는 Random Forest의 경우 주거지역 면적, 상업지역 면적, 유동인구 비율 등의 순으로 나타났으며 Decision Tree의 경우 지하철역 개수, 인구수, 상업지역 면적의 순으로 나타났다. 본 연구를 통해 머신러닝 알고리즘이 외식업의 성패를 예측하는데 효과적임을 알 수 있었고, 성패에 영향을 미치는 요인의 중요도를 파악할 수 있다는 점에서 큰 의미를 가진다.

외식업종의 실패는 단순히 외식업의 실패에 그치지 않고 연관 산업에도 영향을 미치기 때문에 사회적, 경제적으로 파급효과가 크다. 특히, 프랜차이즈의 경우 본사 산하에 가맹점이 운영되는 형태이기 때문에 가맹점의 성패가 본사에 큰 영향을 미칠 확률이 높다. 따라서 본 연구를 통해 가맹점의 성패에 어떤 요인이 중점적으로 영향을 미치는지 알 수 있다. 특히, 접근성 및 경쟁점 등과 같은 다양한 입지 특성을 활용하였기 때문에

신규 가맹점 창업 시 위치를 선정하는데 도움이 될 것으로 예상된다.

그러나, 이 연구에는 몇 가지 한계가 있다. 첫째, 연구에 사용된 데이터가 2019년 2월 한 달 동안의 데이터이기 때문에 시간의 흐름에 따른 입지 특성의 변화와 예측의 영향을 밝혀낼 수 없는 한계가 있다. 또한, 강남구 내 프랜차이즈인 K치킨 대상으로만 진행되었기 때문에 다른 지역의 특성을 반영하지 못하였다. 향후 연구에서 다양한 지역의 데이터와 통합하여 분석을 진행하면 전국 프랜차이즈를 반영하는 더 보편적인 연구가 될 수 있을 것이라고 생각된다.

References

- [1] J.-H. Kim, K.-J. Kang, and K.-M. Ryoo, "A study on the effects of catering consultant capability on consulting satisfaction and business performance of small business owners", *International Journal of Tourism and Hospitality Research*, vol. 34, no. 11. Korea Tourism Research Association, pp. 165 - 177, 2020. <https://doi.org/10.21298/IJTHR.2020.11.34.11.165>
- [2] Kim, Sun Jung, "Effects of the characteristics of TV and Internet Food-related Programs on Dietary Self-efficacy of Regular Viewers - focused on single household", *The Journal of the Convergence on Culture Technology (JCCT)*, Vol. 5, No. 4, pp 307-313, 2019, DOI: <https://doi.org/10.17703/JCCT.2019.5.4.307>
- [3] Korea Institute of Startup and Entrepreneurship Development, 『2020 Startup Business Status Survey (as of 2018)』, (2021).
- [4] aT, 『2019 Domestic and International Restaurant Trend Survey』, (2020).
- [5] Korea Fair Trade Commission, 『Franchise Business Status Report』, (2022).
- [6] Im Dong Lee, Chan Ho Lee and Sang Mog Kang. "The Study of Site Factors Affecting the Sales of Convenience Stores", *Journal of KREAA*, vol. 16, pp.53-77, 2010. DOI: <http://dx.doi.org/10.19172/KREAA.23.4.6>
- [7] Hwang Kyu Sung, Jang Hyeong Jin. "A Study on the Effects of Location Factors on Performance of Convenience Stores -Comparative Analysis on Ulsan and Gyeong Nam Regions-", *Korea Real Estate Academy Review*, vol. 64, pp.199-211, 2016.
- [8] Young Man Lee. "A Review of the Hedonic Price Model". *Journal of KREAA*, vol. 14, pp.81-87, 2008.
- [9] Woo Jin Shin, So Youn Moon. "A Study on the Effects of Locational Characteristics on the Sales of a Coffee Shop Franchise", *Journal of KREAA*, vol. 17, pp.111-123, 2011.
- [10] Aitken, A. C. "IV.-On Least Squares and Linear Combination of Observations", *Proceedings of the Royal Society of Edinburgh*, pp. 42 - 48, 1936. DOI: <https://doi.org/10.1017/S0370164600014346>
- [11] Seong Eun Yeong, Choi Chang Gyu. "The Effects of Disposable Income and Built Environment Characteristics on Retail Business Sales", *Journal of KREAA*, vol. 23, pp.77-91, December 2017. DOI: <http://dx.doi.org/10.19172/KREAA.23.4.6>
- [12] Soowook Lee, Manyong Han, "Utilization and Analysis of Big-data.", *International Journal of Advanced Culture Technology(IJACT)*, 7(4), pp. 255-259, 2019. DOI: <https://doi.org/10.17703/IJACT.2019.7.4.255>
- [13] Cox, D.R., Snell, E.J., *Analysis of Binary Data*, 2nd ed, New York, U.S.A: Routledge, 1970.
- [14] SooYoung Kim. "Prediction of Hotel Bankruptcy Using Multivariate Discriminant Analysis, Logistic Regression and Artificial Neural Network", *Journal of Tourism Sciences*, vol. 30, no. 2, pp.53-75, 2006.
- [15] Breiman, L., et al, *Classification and Regression Trees*, Belmont, California: Wadsworth, 1984.
- [16] Jonghoo Choi, Doosung Seo. "Decision Trees and Its Applications", *Journal of The Korean Official Statistics*, vol. 4, no. 1, pp.61-83, March 1999.
- [17] Breiman, Leo., "Random forests", *Machine Learning*, Vol. 45, No. 1, pp. 5-32, 2001, DOI: <https://doi.org/10.1023/A:1010933404324>
- [18] Younhak Oh, Han Kim, Jaesub Yun, Jong-Seok Lee. "Using Data Mining Techniques to Predict Win-Loss in Korean Professional Baseball Games", *Journal of the Korean Institute of Industrial Engineers*, vol. 40, no. 1, pp.8-17, 2014. DOI: <http://dx.doi.org/10.7232/JKIIE.2014.40.1.008>
- [19] D. E. Farrar and R. R. Glauber, "Multi collinearity in regression analysis: The problem revisited", *Review of Economics and Statistics*, Vol.49, No.1 pp.92-107, 1967.
- [20] A. Swets, "Measuring the Accuracy of Diagnostic Systems", *Science*, vol. 240, no. 4857. American Association for the Advancement of Science (AAAS), pp. 1285 - 1293, 1988.