

<http://dx.doi.org/10.17703/JCCT.2023.9.1.101>

JCCT 2023-1-14

머신러닝 기반 서울시 로컬브랜드 골목상권 추천시스템 설계

Seoul Local Brand Alley Commercial Area Recommendation System Design Using Machine Learning

김지연*, 장호선**, 박민서***

Jiyeon Kim*, Hyoseon Jang**, Minseo Park***

요약 코로나 19 자영업자비상대책위원회가 발표한 자료에 따르면 지난 2년 동안 코로나19로 인한 소상공인 매출의 95.6%가 감소했으며, 방역을 위한 사회적 거리두기로 인해 피해는 더욱 커졌다. 하지만 최근 사회적 거리두기 지침이 전부 해제되고 상권이 활기를 띠면서 서울시는 코로나19의 장기화로 한계에 부딪혀 폐업하였던 소상공인이나 예비 창업자를 위해 안정적으로 사업을 재기할 수 있도록 로컬브랜드 상권 육성사업을 추진하고 있다. 따라서 본 연구는 서울시 로컬브랜드 상권 육성사업의 대상으로 선정된 골목상권 5곳 중 창업자에게 적합한 골목상권을 추천하는 모델을 설계했다. 이 연구의 서울시 로컬브랜드 골목상권 추천시스템은 Xgboost를 이용한 인구관점 모델과 Decision tree를 이용한 상권특징 모델을 합쳐 해당 상권의 주요 인구 연령대와 주요 업종을 추천한다.

주요어 : 골목상권, 상권분석, 머신러닝, 추천시스템

Abstract According to data released by the Covid 19 Self-Employed Emergency Response Committee, 95.6% of small business sales due to Covid 19 have decreased over the past two years, and the damage has further increased due to social distancing for quarantine. However, as all social distancing guidelines have been lifted, and the commercial district has been revitalized, the Seoul Metropolitan Government is pushing for a project to foster local brand commercial districts so that small business owners or prospective founders who have closed their businesses due to the prolonged COVID-19. Therefore, this study propose the model that recommends alley commercial districts suitable for founders among the five alley commercial districts selected for the project to foster local brand commercial districts in Seoul. The Seoul Metropolitan Government's local brand alley commercial recommendation system recommends major population age groups and major industries in the commercial district by combining the population perspective model using Xgboost and the commercial district characteristic model using Decision Tree.

Key words : Alley Commercial Area, Market Analysis, Machine Learning, Recommendation System

*준회원, 서울여자대학교 디지털미디어학과 학부생 (제1저자)

**준회원, 서울여자대학교 소프트웨어융합과 학부생 (참여저자)

***정회원, 서울여자대학교 테이터사이언스학과 조교수

(교신저자)

접수일: 2022년 11월 27일, 수정완료일: 2023년 1월 4일

게재확정일: 2023년 1월 9일

Received: November 27, 2022 / Revised: January 4, 2023

Accepted: January 9, 2023

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

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

I. 서론

골목상권이란 서울시의 정의에 따르면 대로변이 아닌 거주지 안의 좁은 도로를 따라 형성되는 상업 세력의 범위이다[1]. 이러한 골목상권은 코로나19 확산에 따라 재택근무, 온라인 수업, 온라인 소비와 같은 일상의 변화로 경제가 위축되며 급격한 경제적 타격을 받았다. 특히, 방역을 위한 사회적 거리두기 지침으로 인한 유동 인구 감소로 지역 상권에 상당한 피해가 발생하였다[2][3][4][5]. 한국경제연구원에서 조사한 골목상권 지역업자 현황 조사에 의하면 코로나19 시기인 2021년 상반기의 지역업자의 매출액은 78.5%가 감소하였고, 골목상권 매출액은 평균 21.8% 감소하였다. 그러나, 최근 사회적 거리두기 지침이 전부 해제되면서 골목상권이 다시 활기를 띠고 있다. 서울시는 로컬브랜드 상권 육성 사업을 시행하여 소상공인 경쟁력 제고에 기여하겠다고 밝혔다. 특히, 코로나19 장기화와 온라인 소비의 확산으로 경쟁력이 약화된 소상공인의 위기 극복을 위해 문화자원과 같은 로컬콘텐츠를 보유하여 잠재성은 있으나 활성화가 미흡한 골목상권을 로컬브랜드 상권으로 육성하겠다고 밝혔다[6].

서울시의 '우리마을가게 상권분석서비스'가 소상공인들에게 정보 제공에 초점을 두고 있어[7][8][9], 진입하려는 상권의 시시각각 변하는 동태적 변화를 관찰하기에는 한계가 있다. 따라서 본 연구는 상권 분석을 통해 서울시 로컬 브랜드 상권 육성사업 대상인 5곳의 골목상권을 창업자에게 추천해주는 시스템을 구축하였다.

본 논문의 구성은 다음과 같다. 2장에서는 골목상권의 매출에 영향을 미치는 요소를 살펴보고, 3장에서는 상권 추천시스템에 필요한 머신러닝 기법을 설명한다. 4장에서는 본 연구에서 제안하는 머신러닝 모델을 설계하고, 5장에서 모델에 따른 결과를 서술한다. 6장에서는 모델의 결과를 종합적으로 살펴봄에 각 골목상권의 특성에 대한 연구 결론을 기술한다.

II. 골목상권 선정 및 매출에 영향을 미치는 요인

연구를 진행하기에 앞서 골목상권의 매출과 밀접한 연관성이 있는 요인에 대한 선행연구를 조사하였다. Turhan et al.(2013)에서는 상업시설 입지 선정의 결정

요인에 관한 이론적 모델을 제시하였다 ([그림 1] 참조).

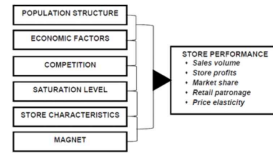


그림 1. 상업 시설 입지 결정에 관한 이론적 모델
Figure 1. A Theoretical Model for Commercial Facility Site Determination

이 연구에서는 상업시설 입지 결정 요인을 크게 인구구조(Population structure), 경제적 요인(Economic factors), 경쟁 관계 (Competition), 포화 수준(Saturation level), 점포 특성(Store characteristics), 집객시설(magnet) 등 6개로 구분한다[10]. 국내에서 진행된 관련 연구로는 김창기-정승영(2013), 이연수 외(2014) 등이 있으며, Turhan et al.과 마찬가지로 주로 인구요인, 입지 요인, 경제적 요인, 경쟁요인 등이 변수로 활용되고 있다[11][12].

III. 상권 추천시스템에 필요한 머신러닝 기법 소개

이번 장에서는 서울시 로컬브랜드 골목상권을 대상으로 상권별 추천시스템을 설계하기 위한 머신러닝 기법을 설명한다.

1. K-means Clustering

K-means Clustering은 k개의 중심점 중 가장 가까운 중심점에 각 항목을 할당하는 과정을 반복해 군집을 형성하는 것으로, 클러스터링의 대상이 되는 객체(Object)들은 각 객체의 특성을 나타내는 속성을 가지고 있다[13].

2. Xgboost

Xgboost는 gradient boosting 알고리즘의 방법 중 하나로, 랜덤 포레스트(Random Forest)의 단점을 개선하기 위해 고안된 의사결정나무 앙상블(decision tree ensemble) 방법이다[14].

3. Decision tree

의사 결정 트리는 지도 학습(Supervised learning)의 일종으로, 관측값과 목표값을 연결하는 데 있어 논리적 명제를 여러 개 생성하는 방식의 의사 결정 트리를 구축하여 이용한다[15].

IV. 골목상권 추천을 위한 머신러닝 모델 설계

본 연구에서는 매출에 영향을 미치는 요인 중 인구구조, 점포 특성, 집객시설에 머신러닝 알고리즘을 적용해 상권을 추출한다. 추출된 각 상권에 대해 상주인구 연령대와 점포 특성을 바탕으로 상권을 추천하는 모델을 제안한다.

모델 설계 전 연령별 인구수와 점포 수, 집객시설 수 등 독립변수 사이의 관별 타당성을 검사하기 위해 VIF(Variance Inflation Factor, 분산팽창요인)를 활용하여 변수들이 서로 영향을 주는지 파악하였다. 이 과정에서 연령별 인구수 변수들이 점포 수와 집객시설 수 변수에 영향을 받아 VIF 수치가 10 이상이 있음을 확인하였다. 따라서 연령별 인구수 변수를 이용한 실험과 점포 특성 및 집객시설 변수를 이용한 실험으로 나누어 진행한다. 여기서 연령별 인구수 변수를 사용한 모델을 인구관점 모델, 그 외 점포 특성 및 집객시설 변수를 사용한 모델을 상권특징 분석 모델로 지칭한다. 모델 전체 프로세스는 [그림2]와 같다.

1. 인구관점 모델

1) 데이터 수집

서울 열린데이터광장 사이트에서 제공하는 서울시 우리마을가게 상권분석서비스 데이터셋 중 2016년부터

2021년까지 6년간의 상권-추정매출, 상권-생활인구, 상권-상주인구, 상권-점포 데이터셋을 활용하였다. [표 1]의 인구관점 변수 정의 표에서 상권-추정매출은 각 상권의 추정 매출을 예측하기 위해 사용되었고, 상권-생활인구와 상권-상주인구는 매출에 영향을 미치는 요인 중 인구구조 요인으로 분석에 사용되었다.

서울시 로컬브랜드 골목상권은 한 곳의 골목만 해당 하는 것이 아닌 여러 곳의 골목상권을 묶어 하나의 로컬브랜드 골목상권으로 지칭하고 있다. 예를 들어 영등포구 양평로의 '선유로운'은 선유로55길, 양평로19길, 양평로22라길 세 곳을 묶어 '선유로운' 하나로 지칭한다. 따라서 본 연구에서는 로컬브랜드 골목상권 5곳에 포함되는 골목상권들의 상권 코드를 알아내기 위해 서울시 '우리마을가게 상권분석서비스'에서 제공하는 상권 위치 정보를 지도와 대조해 상권 코드를 수집하였다. 그 결과 [표 2]와 같이 상권 8곳이 선정되었고, 각 골목이 속하는 로컬브랜드 골목상권 5곳에 대해 각각 모델을 설계하였다.

2) 데이터 전처리

- Data Cleansing

상권-추정매출, 상권-생활인구, 상권-상주인구 모두 로컬브랜드 골목상권에 포함되는 골목 8곳의 데이터만 추출했다. 세 개의 데이터셋을 공통 독립 변수인 골목상권코드와 기준_년_코드를 기준으로 하나의 데이터로 합쳤다. [표 2]는 각각의 데이터셋에서 사용한 변수에 대한 정보를 정리한 것이다.

- Feature Selection

Data Cleansing 과정에서 하나로 합친 데이터셋의 변수들 간에 서로 영향을 미치는 정도를 히트맵으로 분석해 본

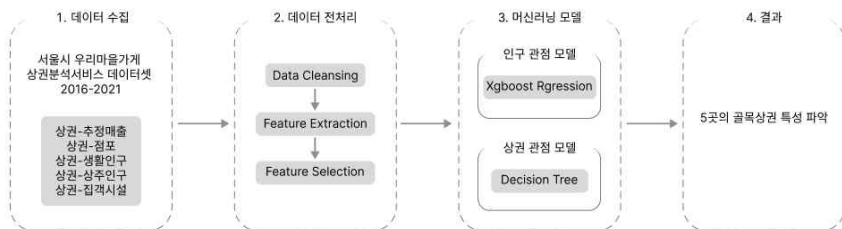


그림 2. 추천시스템 설계를 위한 머신러닝 모델

Figure 2. Machine learning models for recommend system design

표 1. 인구관점 모델의 변수 정의

Table 1. Defining Variables in the Population Perspective Model

변수명		정의	출처
종속 변수	분기당 매출 금액	서울시의 내국인을 대상으로 신용카드 매출과 융합하여 추정매출 산출	서울시 우리마을가게 상권분석서비스 상권-추정매출 2016~2021년도
독립변수	총 상주인구 수	행정구역별 주민등록 통계 데이터를 건물단위별 가구수 및 성별/연령대별 인구수 추정	서울시 우리마을가게 상권분석서비스 상권-상주인구 2016~2021년도
	연령대 10 상주인구 수		
	연령대 20 상주인구 수		
	연령대 30 상주인구 수		
	연령대 40 상주인구 수		
	연령대 50 상주인구 수		
	연령대 60 상주인구 수		
	총 생활인구 수	서울시가 보유한 빅데이터와 KT의 통신데이터로 특정한 '특정 시점'에 '서울의 특정 지역'에 '존재'하는 모든 인구	서울시 우리마을가게 상권분석서비스 상권-생활인구 2016~2021년도
	연령대 10 상주인구 수		
	연령대 20 상주인구 수		
연령대 30 상주인구 수			
연령대 40 상주인구 수			
연령대 50 상주인구 수			
연령대 60 상주인구 수			
패생변수	연령대 2030 상주인구 수	행정구역별 주민등록 통계 데이터를 건물단위별 가구수 및 성별/연령대별 인구수 추정	서울시 우리마을가게 상권분석서비스 상권-상주인구 2016~2021년도
	연령대 4050 상주인구 수		
	연령대 60 이상 상주인구 수		

표 2. 로컬브랜드 5곳에 해당하는 상권 8곳

Table 2. 8 commercial districts corresponding to 5 local brands

상권명	골목상권명	골목상권 코드
장충단길	퇴계로 56길	1000047
합마르뜨	성지 3길	1000485
	독막로 8길	1000465
선유로운	선유로 55길	1000710
	양평로 19길	1000719
	양평로 22라길	1000720
오류버들	오류로 8길	1000650
양재천길	논현로 27길	1000852

결과 [그림 3]과 같은 결과를 추출하였다. 히트맵은 열분포 정보 제공하는 시각화 방법[14]으로 상주인구 수는 20대와 30대를 제외한 대부분의 연령대가 수치 0.5 이상으로 매출에 양의 영향을 끼치는 반면, 생활인구 수는 평균 0.2의 수치로 매출과 밀접한 연관이 있다고 보기 어렵다. 따라서 생활인구는 제외하고 상주인구 수만 분석 대상으로 하여 실험을 진행했다.

- Feature Extraction

10대부터 60대 이상 연령대별 상주인구 수 각각의 변수들 간에 독립변수 판별 타당성을 검사하기 위해 VIF(Variance Inflation Factor, 분산팽창요인)를 활용해서 서로 영향을 주는지 파악했다. VIF는 보통 10보다 큰

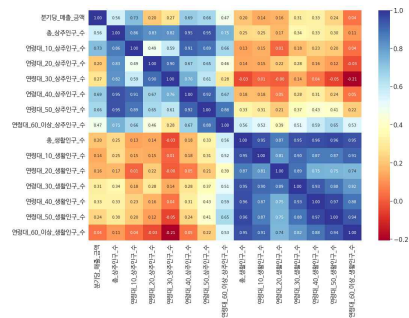


그림 3. 매출과 상주/생활 인구 수에 대한 히트맵

Figure 3. Heat map for the relationship between sales and the number of resident/living population

표 3. 인구관점 모델의 최종 선정된 독립 변수들의 VIF 결과 Table 3. VIF results of the final selected independent variables of the population perspective model

Features	VIF Factor
연령대 2030 상주인구 수	2.7
연령대 4050 상주인구 수	6.9
연령대 60대 이상 상주인구 수	4

경우 다중공산성 문제가 존재하는 것으로 간주한다 [14][15]. 검사 결과 모든 독립변수의 VIF가 10을 넘어 연령대별 변수 각각을 독립변수로 활용할 수 없다고 판단하였다. 따라서, K-means Clustering을 활용하여 서로 비슷한 양상을 가지는 연령대끼리 묶은 범주형 과정

표 4. 상권특징 분석 모델의 변수 정의

Table 4. Variable definition of commercial district characteristics analysis model

변수명	정의	출처	
종속 변수	퇴계로56길	중구 장충단길 상권은 '히스토리컬 시티'로 76년 전통의 과거적인 '태극당'부터 촉발, 냉면까지 장충동의 다양한 자산과 사람들의 이야기가 담긴 상권	서울시 우리마을 가게 상권분석 서비스 2021년도
	성지3길	마포구 합마르프 상권은 '크리에이터 타운'으로 독립서점, 갤러리, 이색 맛집 등 자신만의 맛을 만들어 가는 창작자와 소비자가 모이는 차별화된 상권	
	독막로 8길		
	선유로 55길	영등포구 선유로운 상권은 '선(善), 여유로운'이라는 이름으로 반려동물에게도 친화적인 상권	
	양평로19		
	양평로22라길		
	오유로8길		
논현로27길	서초구 양재천길 상권은 '살롱 in 양재'으로 제프와 클래식 음악이 흐르는 상권		
독립변수	분기당 매출 금액	서울시의 내국인을 대상으로 신용카드 매출과 융합하여 추정매출 산출	
	개업 점포수	100개 생활밀접업종*의 점포수, 개·폐업수(률), 생존율, 점포이력 *사업체수가 많고, 종사자수가 5인 미만인 소규모 사업체 비중이 높으며, 창업 등 진출입이 용이한 업종으로 정의	
	폐업 점포수		
	집객시설 수	인구집중 유발하는 주요/집객시설 구분 및 위치정보 (관공서, 금융기관, 병원, 학교, 유통점, 문화관광 / 영화관, 숙박시설, 교통관련 시설)	

변수를 만들었다. K-means Clustering을 적용한 결과 40대, 50대가 서로 유사한 양상을 보였고, 20대와30대가 서로 유사한 양상을 보였다. 그리고 60대 이상의 경우 비슷한 양상을 보이는 연령대가 없어 다른 연령과 묶어 파생 변수로 만들지 않고 독립적인 변수로 분석에 사용하였다. 10대의 경우 40대, 50대 상주인구 수에 많은 영향을 받아 분석 대상 변수에서 제거했다. 그 결과 최종적으로 완성된 총 3개의 파생 변수들의 VIF는 [표 3]과 같다.

3) 인구관점 모델

전처리한 데이터셋을 8:2의 비율로 Train set과 Test set으로 나누어 학습시켰다. Xgboost 알고리즘을 사용하여 매출 예측 모델을 학습시켜 특성 중요도 (Feature Importance)를 추출했으며 모델의 성능을 평가하기 위해 R-squared 지표로 정확도를 검증하였다.

2. 상권특징 분석 모델

1) 데이터 수집

서울 열린데이터광장 사이트에서 제공하는 서울시 우리마을가게 상권분석서비스 데이터셋 중 상권-추정매출, 상권-점포, 상권-생활인구, 상권-상주인구, 상권-집객시설 데이터셋을 활용하였다.

2) 데이터 전처리

- Data Cleansing

수집한 데이터셋들을 결합하고, 중복되는 Feature들은 제거하였다. 분기당 매출 금액이 4분기에 따라 나누어져 있어 불필요한 행들을 줄이기 위해 평균값으로 변환하였다. 서울시 브랜드 육성사업의 5가지 골목상권에 해당하는 8곳의 골목을 추출하였고 이는 [표 4] 상권특징 분석 모델 변수 정의 표의 종속변수와 같다. 골목상권의 특성상 외식업에 대한 창업의 수요가 높기 때문에 50가지의 업종 중 외식업 업종 11가지를 추출하였고 이는 [표 5] 외식업종 정리와 같다.

- Feature Extraction

독립변수 중 분기당 매출금액과 시간대별 및 요일별 매출 금액은 중복되는 데이터로 판단하여 제거하였다.

- Feature Selection

추출된 독립변수를 VIF(Variance Inflation Factor, 분산팽창요인)를 활용하여 타당성을 판별하였다. VIF 10 이하로 최종 추출된 변수들은 [표 6]과 같다.

3) 상권특징 분석 모델

전처리한 데이터셋을 7:3의 비율로 Train set과 Test set으로 나누어 Decision Tree Classifier 알고리즘을 사용하여 모델을 학습시켰다. 입력변수들의 중요도와 VIF를 확인하면서 불필요한 Feature를 제외하며 학습을 진행하였으며 상권특징 분석 모델의 Decision Tree는

표 5. 외식업종 정리

Table 5. The food service industry list

대분류	100대 업종명(코드)	세분류명(9차)	예시
외식	한식음식점 (CS100001)	한식 음식점업 (56111)	설렁탕집 · 해물탕집 · 해장국집 · 보쌈집 · 냉면집
	중식음식점 (CS100002)	중식 음식점업 (56112)	대중음식점(중식) · 중국음식점 · 탕수육전문점
	일식음식점 (CS100003)	일식 음식점업 (56113)	초밥집(일식전문점) · 일식 핫집 · 로마다야기 · 일식 우동집
	양식음식점 (CS100004)	서양식 음식점업 (56114) 기타 외국식 음식점업 (56119)	레스토랑(서양식) · 패밀리 레스토랑 · 이탈리아 음식점 · 프랑스 음식점 · 베트남 음식점 · 인도 음식점 · 러시아 음식점 · 태국 음식점
	제과점 (CS100005)	제과점업 (56191)	제과점(추석식) · 떡집(음식점형태) · 도너츠 전문점 · 아이스크림 전문점
	패스트푸드점 (CS100006)	피자, 햄버거, 샌드위치 및 유사음식점업 (56192)	피자전문점 · 샌드위치 전문점 · 햄버거 전문점 · 토스트 전문점
	치킨전문점 (CS100007)	치킨 전문점 (56193)	닭강정 · 봉닭집 · 양념(프라이드) 치킨 전문점
	분식전문점 (CS100008)	분식 및 김밥 전문점 (56194) 그 외 기타 음식점업 (56199)	만두 전문점 · 라면 전문점 · 김밥 전문점 · 떡볶이 전문점 · 간이휴게식당 · 봉어빵 · 호떡집
	호프-간이주점 (CS100009)	기타 주점업 (56219)	소주방 · 호프집 · 막걸리집 · 토속주점
	커피-음료 (CS100010)	비알콜 음료점업 (56220)	주스 전문점 · 커피숍 · 찻집 · 다방

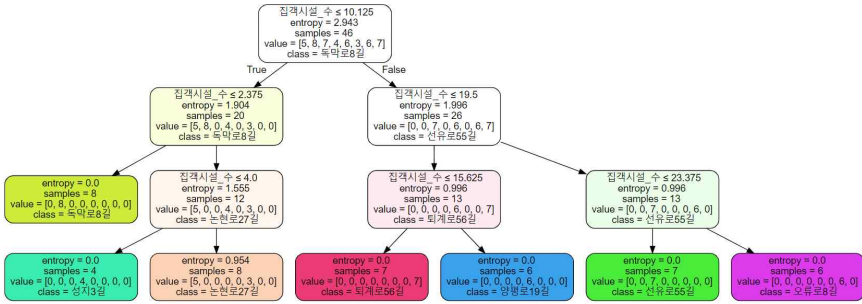


그림 4. 상권특징 분석 모델의 Decision Tree

Figure 4. Decision Tree of Commercial District Characteristics Analysis Model

표 6. 상권분석 모델의 최종 선정된 독립 변수들의 VIF

Table 6. The final selected independent variables' VIF of the commercial district analysis model

Features	VIF Factor
분기당 매출 금액	3.8
개업 점포 수	5.4
폐업 점포 수	4.9
집객시설 수	1.5

[그림 4]와 같다. 모델의 성능을 평가하기 위해 정확도를 검증하였다.

V. 결 과

서울시 로컬브랜드 골목상권 추천을 위한 인구관점 모델의 결과는 [그림 5]과 같다. [그림 5]은 인구관점 모델의 각 상권별 특성중요도 추출 결과 그래프이며,

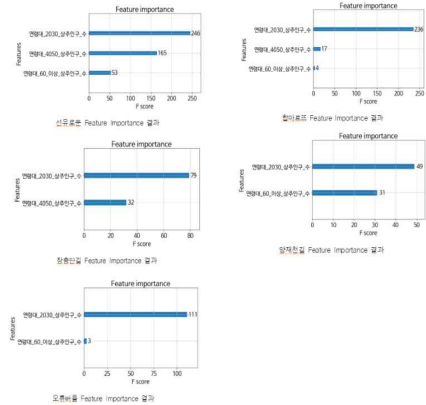


그림 5. 인구관점 모델의 특성중요도 그래프

Figure 5. Features Importance Graph of Population Perspective Model

표 7. 인구관점 모델 및 상권 분석 모델 최종 결과

Table 7. Final result of population perspective model and commercial area analysis model

골목상권		상권특징 분석 모델 결과				인구관점 모델 결과
		집객시설 수	개업률	폐업률	점유율순	상주인구 점유율 순
장충단길	퇴계로56길	15.6 이하	2%	2%	한식음식점(53%) > 분식전문점(11%) > 반찬가게(10%)	2030 (79) > 4050(32)
합마르뜨	성지3길	4.0 이하	7%	3%	커피-음료(41%) > 양식음식점(19%) > 한식음식점(18%)	2030 (236) > 4050 (17) > 60이상(4)
	독막로 8길	2.3 이하	7%	5%	커피-음료(29%) > 한식음식점(19%) > 분식전문점(13%)	
선유로운	선유로 55길	23.3 이하	4%	4%	한식음식점(34%) > 커피-음료(24%) > 분식전문점(14%)	2030 (246) > 4050(165) > 60이상(53)
	양평로19길	15.6 이하	5%	4%	한식음식점(47%) > 분식전문점(11%) > 커피-음료(10%)	
	양평로22라길	4.0 이하	3%	3%	한식음식점(37%) > 커피-음료(31%) > 분식전문점(16%)	
오류버들	오류로8길	23.3 이하	4%	2%	한식음식점(45%) > 분식전문점(13%) > 반찬가게(8%)	2030 (111) > 60이상(3)
양재천길	논현로27길	4.0 이하	2%	2%	한식음식점(40%) > 커피-음료(26%) > 호프-간이주점(11%)	2030 (49) > 60 이상 (31)

연령대별 상주인구가 매출에 영향을 미치는 정도를 수치로 나타낸 것이다.

우선 모든 상권에서 가장 큰 영향을 미친 연령대는 2030 세대였다. 다음으로 영향을 미친 연령대는 4050이였으며, 가장 적은 영향을 미친 연령대는 60대 이상

이었다. 선유로운과 합마르뜨는 모든 연령대의 수치가 그래프 상에 나타나고 있으며, 이 중 2030의 영향력이 압도적임을 알 수 있다. 반면 모든 연령대가 반영되지 않은 골목상권도 있다. 장충단길의 경우 60대의 특성중요도가 누락되었으며, 양재천길과 오류버들의 경우엔 4050대의 특성중요도가 누락된 것을 확인할 수 있었다. 또한 합마르뜨와 선유로운처럼 오류버들의 경우에도 2030의 영향력이 압도적임을 확인할 수 있었다.

표 8. 모델 정확도
 Table 8. Model Accuracy

알고리즘	Accuracy	
Decision Tree	Train set	0.93
	Test set	0.85
Xgboost Regression	선유로운	0.99
	양재천길	0.95
	합마르뜨	0.99
	장충단길	0.95
	오류버들	0.98

상권특징 분석 모델의 결과에서는 집객시설수가 각 골목상권의 특성을 가장 뚜렷하게 반영하고 있다. 각

상권을 분석하기 위하여 인구관점 모델의 결과, 개업률, 폐업률, 외식업의 점유율순을 정리하였고 이는 [표 7]과 같다. 한식음식점이 모든 상권의 상위 3위 안에 포함되어 있고, 커피-음료와 분식 전문점도 높은 편에 속해 있다. 각 상권의 Decision Tree와 Xgboost Regression 모델의 정확도는 [표 8]과 같다.

VI. 결론

본 연구는 머신러닝 기법을 활용해 서울시 로컬브랜드 골목상권의 매출에 영향을 미치는 인구 연령대와 상권 특징을 분석하고, 상권 추천시스템을 설계하였다.

인구관점 모델에서는 연령대별 상주인구 수 변수를 Clustering을 통해 비슷한 연령대끼리 묶어 범주형 과생변수로 만들었다. 이후 XGBoost Regression을 통해 각 상권에 대한 특성 중요도 비교를 진행했다. 그 결과 모든 상권에서 압도적으로 많은 영향을 미친 연령대는 2030 세대였고, 그 다음은 4050, 60대 이상 순이었다.

상권특징 분석 모델에서는 변수에 대한 VIF 검사를 통해 분기당 매출금액, 개업 점포수, 폐업점포수, 집객 시설 수를 독립변수로 추출하였다. 독립변수들의 특성중요도 결과 집객시설 수가 가장 많은 영향을 미치는 것으로 나타났다. 인구관점 모델의 결과와 상권특징 모델의 결과를 연관 지어 살펴보고, 5곳의 상권들을 비교

해보면 각 골목상권의 특성을 파악할 수 있다.

첫째로 장충단길 퇴계로 56길은 비교적 개업률과 폐업률이 낮아 신규 가게의 생존율이 높을 것으로 보인다. 또한 한식음식점의 점유율이 압도적으로 높고, 2030부터 4050까지의 상주인구수가 많아 여러 세대가 즐길 수 있는 한식음식점의 창업을 추천할 수 있다.

둘째로 합마르트의 성지3길과 독박로 8길은 집객시설 수가 낮게 분류되어 통행량을 낮출 것이다. 커피-음료와 2030 상주인구의 점유율이 압도적으로 높을 것으로 보아 커피-음료 업종의 창업과 2030 세대를 타겟팅한 마케팅에 초점을 맞출 것을 추천한다.

셋째로 선유로운의 선유로 55길, 양평로19길, 양평로 22라길은 집객시설수가 높게 분류된것으로 보아 통행량이 많을 것이다. 또한 상주인구가 전 연령대에 비교적 균일하게 분포되어 있다. 따라서 전 연령대가 즐길 수 있는 한식음식점과 커피-음료의 창업을 추천한다.

넷째로 오류버들의 오류로8길은 집객시설수가 높게 분류되어 통행량이 많을 것이다. 업종별 점유율순은 한식음식점이 압도적으로 높으며, 반찬가게가 20% 정도 차지하고 있다. 또한 2030 상주인구의 점유율이 상당히 높은 것으로 볼 때 자취하고 있는 2030 인구가 많을 것이다. 따라서 자취하고 있는 2030을 위한 반찬가게 혹은 1인 전용 한식음식점 창업을 추천한다.

다섯째로 양재천길 논현로 27길은 모든 상권 중 유일하게 호프-간이주점이 11%를 차지하고 있다. 개업률과 폐업률도 낮아 상권이 안정화 되어 있어 신생 가게의 생존률도 높을 것이다. 따라서 논현로 27에서는 2030 직장인을 위한 호프-간이주점을 추천한다.

모델의 종합적 분석결과로 볼 때 서울시 골목상권 진입을 위해서는 업종별, 연령대별 사업 전략이 필요하며 이를 위해서는 정확한 상권분석이 이루어져야 한다는 것을 확인할 수 있다. 따라서 이 연구에서 도출한 추천시스템 모델을 통해 소상공인의 창업 활동을 도울 수 있을 것이라 기대한다.

References

- [1] Seoul Our Village Shop commercial area analysis service (<https://golgok.seoul.go.kr/>)
- [2] F. M. F., Thomas, S. G. Charlton, I. Lewis and S. Nandavar, "Commuting before and after COVID-19", *Transportation Research Interdisciplinary Perspectives*, Vol.11, No.4, 2021, 100423. (<https://doi.org/10.1016/j.trip.2021.100423>)
- [3] Keumsook Lee, Sohyun Park, Yuhee Ham, "Changes in Subway Traffic in Seoul during Social Distancing due to the Spread of COVID-19 and Geographic Characteristics of the Area Behind the Station," *Journal of the Economic Geographical Society of Korea*, Vol.24, No.2, 2021, pp.127-142. (<https://doi.org/10.23841/egsk.2021.24.2.127>)
- [4] Jungwon Ha, Sunjae Kim, Sugie Lee, "Analysis of Spatio-Temporal Characteristics of Small Business Sales by the Spread of COVID-19 in Seoul, Korea : Using Space Time Cube Model", *Journal of Korea Planning Association*, Vol.56, No.2, 2021, pp.218-234.
- [5] Muhammad Sohail Ahmad, Peter C. Johnson, Seul Lee, Sujin Lee, "A Study on Resilience and Survival Characteristics based on the Types for Seoul's Side Street Trade Areas during the COVID-19 Pandemic". *Korea Real Estate Analysts Association Conference Proceedings*, Vol.1, 2021, pp.35-60.
- [6] Announcement of the 2022 local brand commercial development project in Seoul (https://www.seoul.go.kr/news/news_notice.do#view/359268?tr_code=snews)
- [7] Young Gab Kim. "Overview and Comparison of Domestic Major Commercial District Information System", *Journal of Food service Management*, Vol.18, 2015, pp.275-288.
- [8] Hyeon Cheo Kim, Young Soo An, "A Monitoring the Dynamic Change of Seoul's Side Street Trade Areas Using Density and Diversity of Stores", *Seoul Studies* Vol.20, No.4, 2019, pp. 149-162.
- [9] Hyeoncheol Kim, Seungil Lee, "A Study on the Factors Affecting the Revenue in Seoul's Side Street Trade Areas", *Seoul Studies* Vol.20, No.1, 2019, pp. 117-134
- [10] Glden Turhan, Mehmet Akaln, Cemal Zehir, *Literature Review on Selection Criteria of Store Location Based on Performance Measures*, *Procedia-Social and Behavioral Sciences*, Vol. 99, 2013, pp.391-402, ISSN 1877-0428. (<https://doi.org/10.1016/j.sbspro.2013.10.507>)
- [11] Changki Kim, Seung Young Jeong, *A Study on the Factors Affecting Hotel & Restaurant Location in Seoul*. *Korean Management Consulting Review*, Vol.13, No.4, 2013, pp.297-314.

- [12]Youn Soo Lee, Hyun Shin Park, Seung Hwan Lew, Jun Mo Kang, An Analysis of the Location Factors that Affects the Sales of Campus Commercial District.Seoul Studies, Vol.15, No.1, 2014, pp.17-34.
- [13]Wha-Beum Park, Young-Sung Cho, Hyung - Hwa Ko, Clustering Method of Weighted Preference Using K-means Algorithm and Bayesian Network for Recommender System. Journal of Information Technology Applications and Management, Vol.20, No.3, 2013, pp.219-230. (https://doi.org/10.21219/jitam.2013.20.3_spc.219)
- [14]Sangwhan Kim, Empirical studies on the stock price prediction performance of Xgboost models. Journal of Social Science, Vol.39, No.1, 2022, pp. 29-55.
- [15]Kim Jae-Hwan, Lee, Min-Gu, Analyzing Decision Tree & Random Forest Algorithm, Pro ceedings of KIIT Conference, 2018, pp. 132-133.