

http://dx.doi.org/10.17703/JCCT.2024.10.2.459

JCCT 2024-3-53

## 딥러닝 기반 서울시 행정동별 외식업종 상권 변화 예측

### Predicting Changes in Restaurant Business District by Administrative Districts in Seoul using Deep Learning

김지연\*, 오수민\*\*, 박민서\*\*\*

Jiyeon Kim\*, Sumin Oh\*\*, Minseo Park\*\*\*

**요약** 자영업자의 빈번한 폐업은 국가적인 경제 손실을 동반한다. 특히 외식업종이 가장 높은 폐업률을 보이기 때문에 외식업종의 상권 변화를 예측하여 업체의 생존에 도움을 주는 것이 필요하다. 외식업종의 생존율과 폐업률에 영향을 미치는 요인에 대한 연구는 활발하나, 상권의 변화 정도를 예측하는 연구는 부족한 실정이다. 따라서, 본 연구에서는 상권 변화에 초점을 맞추는 연구를 하고자 한다. 이를 위해 서울시 행정동별 상권 변화를 예측하는 딥러닝(Deep Learning) 모델을 설계한다. 첫째, 2023년과 2022년 2분기의 상권 변화와 관련된 변수를 수집한다. 둘째, 1년간의 등락 정도를 백분율로 환산한 후, 증강 단계를 거친다. 셋째, 딥러닝 모델을 활용하여 상권 변화를 예측하는 모델을 제안한다. 향후 본 연구를 고려한 외식업종 지원정책은 상권의 질적 성장 및 경제 성장에 도움이 될 것으로 기대한다.

**주요어 :** 상권, 외식업, 외식업, 머신러닝, 딥러닝

**Abstract** Frequent closures among self-employed individuals lead to national economic losses. Given the high closure rates in the restaurant industry, predicting changes in this sector is crucial for business survival. While research on factors affecting restaurant industry survival is active, studies predicting commercial district changes are lacking. Thus, this study focuses on forecasting such alterations, designing a deep learning model for Seoul's administrative district commercial district changes. It collects 2023 and 2022 second-quarter variables related to these changes, converting yearly fluctuations into percentages for augmentation. The proposed deep learning model aims to predict commercial district changes. Future policies, considering this study, could support restaurant industry growth and economic development.

**Key words :** Business district, Restaurant Business, Restaurant business, Machine learning, Deep learning

#### 1. 서론

2021년 기준 OECD 보고서에 따르면, 한국의 자영업자 비율은 23.9% 수준으로 나타났다[1]. 이는 38개의 전체 회원국 중 8번째로 높은 수준이며, 경제 수준이 높은 주요 G7(Group of Seven)과 비교하면 최고 수준이다. 그러나 높은 자영업자 비율과 달리, 국내 자영업 폐업률은 77.2%로 5년 이내에 절반 이상이 폐업에 직

면한다[2]. 그 중에서도 외식업종의 비율이 20.6%로 가장 높은 폐업률을 보인다[3]. 이와 같은 외식업종의 조기 폐업은 경제 불안, 사회적 안정성 감소, 실업문제와 같은 다양한 사회적 문제를 야기한다[4]. 이에 외식업종의 생존율을 높이기 위한 방안을 제시하고 있으며, 생존율 및 폐업률에 대한 세밀한 연구가 필요하다[5].

외식업종의 생존율 및 폐업률에 대한 연구는 지속적으로 행해져 왔다[5-8]. 외식업종의 폐업률에는 인구수,

\*준회원, 서울여자대학교 디지털미디어학과 학부생  
\*\*준회원, 서울여자대학교 테이터사이언스학과 학부생  
\*\*\*정회원, 서울여자대학교 테이터사이언스학과 조교수 (교신저자)  
접수일: 2024년 1월 5일, 수정완료일: 2024년 1월 22일  
게재확정일: 2024년 2월 1일

Received: January 5, 2024 / Revised: January 22, 2024  
Accepted: February 1, 2024  
\*\*\*Corresponding Author: mpark@swu.ac.kr  
Dept. of Data Science, Seoul Women's Univ, Korea

임대료, 지역내총생산, 소비자물가지수와 같은 다양한 요소들이 영향을 미치며, 외식업종의 장기생존에는 상권의 변화 특성이 가장 큰 영향력을 가진다. 또한, 다양한 용도의 사업체가 밀집해 있는 성장 중인 상권의 경우 생존 확률이 높으며[6-7], 상권의 쇠퇴 수준이 높은 상주인구가 많은 주거 지역일수록 평균 생존 기간이 짧게 나타난다[8].

그러나 대부분의 연구들은 고전적인 통계 방법을 사용하여 상권을 분석한다는 한계가 존재한다. 주요 방법론으로 생존분석 모형이 사용되었으나 이는 비선형적인 상권 데이터를 활용할 수 없다는 한계를 가지며, 생존함수를 비교하는 것으로는 미래의 상권 변화를 예측하기에 충분하지 않다.

따라서 본 연구에서는 비선형적인 상권 데이터인 지역, 인구, 비용과 같은 다양한 변수를 활용하여 딥러닝(Deep Learning) 기반의 서울시 행정동별 외식업종의 상권 변화 예측 모델을 제안하고자 한다. 딥러닝은 통계에 비해 비선형적인 복잡한 데이터를 분석하고 모델링하는데 효과적이다[9]. 데이터는 서울시 전반에 걸쳐 분포하고 있는 외식업종을 우선 대상으로 한다.

본 논문은 다음과 같이 구성된다. 제 2장에서는 외식업종의 생존율과 폐업률에 영향을 미치는 요인을 분석한 선행 연구에 대해 살펴본다. 제 3장에서는 딥러닝 기반 상권 변화 예측 모델을 제안한다. 제 4장에서는 연구 결과에 대해 설명하고, 제 5장에서는 결론을 언급한다.

## II. 선행연구

상권의 공간적 특성에 따라 외식업종의 생존을 다룬 통계 기반의 연구는 지속적으로 행해져 왔다[5-8]. 외식업종의 생존과 폐업은 상권 변화와 긴밀하게 연관되어 있으며 지역별 상권 변화를 분석하는 것이 주를 이룬다.

최은준 외[4]는 서울시 전체에 걸쳐 생존 및 폐업한 사업체의 영업 기간 변화를 관찰하여 상권을 분석하였다. 서울시 열린데이터광장에서 제공하는 영업 시작일, 폐업 일자를 활용하여 생존 및 폐업 기간을 계산하고, 데이터로부터 특성을 추정하는 커널밀도분석(Kernel Density Estimation)을 활용하여 지역별 상권의 변화를 파악하였다. 이를 바탕으로 서울시 주요 상권을 안정적

인 상권과 빠르게 변화하는 상권으로 분류하는 방법을 제안하였다.

이동현 외[7]는 자영업자 중 폐업률이 가장 높은 업종인 외식업을 대상으로 상권의 성장 및 쇠퇴에 따라 상권 유형을 구별하고 유형에 따라 생존율과 폐업률을 분석하였다. 연구에서 제안한 상권활성화지수를 활용하여 활성화-성장상권, 활성화-쇠퇴상권, 비활성화-성장상권, 비활성화-쇠퇴상권, 현상유지상권 등의 5가지 상권 유형으로 구분하였다. 상권 유형별로 외식업종의 생존율을 비교하기 위해 카플란-마이어법(Kaplan Meier Estimation)을 이용하였다. 카플란-마이어법(Kaplan Meier Estimation)은 시간의 흐름에 따라 어떠한 사건의 발생 확률을 알아보는 통계 분석이다[7]. 활성화-성장상권의 생존율이 가장 높으며, 비활성화-쇠퇴상권의 생존율이 가장 낮게 나타나는 것을 확인하였으며, 활성화 정도가 높은 곳에 위치한 외식업종의 생존율이 더 높음을 증명하였다.

김성호 외[8]는 행정안전부에서 제공하는 개별 음식점의 공간 데이터를 활용하여 2013년 서울시에 창업한 음식점을 대상으로 2019년까지의 생존율 및 생존요인을 분석하였다. 업체의 위치정보 및 개폐업 일자와 같은 음식점의 개별 속성과 지하철 및 버스 정류장의 개수, 유동 인구와 같은 지역 특성을 반영한 변수들을 사용하였다. 대표적인 생존함수인 카플란-마이어법(Kaplan Meier Estimation)과 콕스 비례위험모형(Cox proportional hazard model)을 활용하였다. 콕스 비례위험모형(Cox proportional hazard model)은 카플란-마이어법(Kaplan Meier Estimation)이 사건의 발생 여부만을 중요하게 다루는 것과 달리, 생존에 영향을 미치는 여러 위험 요소에 대해서도 분석이 가능하다[8]. 서울시 지역별 쇠퇴 수준이 음식점 생존율 및 생존요인에 영향을 미친다는 것을 증명하였다.

이처럼 외식업종의 생존 및 폐업에 관련된 연구들은 통계를 활용하여 상권 분석에 초점을 맞추어 진행되고 있다. 그러나 앞선 연구들에서 활용한 생존함수는 변수가 시간에 따라 일정하다는 가정이 필요하기에 실제 상권 데이터에 활용하는 데에 한계가 존재한다. 따라서 본 논문에서는 딥러닝을 활용하여 실제 상권 데이터에 적용 가능한 상권 변화 예측 모델을 제안하고자 한다.

### III. 연구방법

본 연구의 제안 방법은 데이터 수집, 데이터 전처리, 딥러닝 기반 분류 모델을 활용한 행정동별 상권 변화 예측 모델 개발 세 단계로 구성된다.

#### 1. 데이터 수집

서울 열린데이터 광장과 서울시 상권 분석 서비스에 서 서울특별시의 행정동별로 구분된 2023년 2분기와 2022년 2분기의 데이터를 수집하였다. 최근 1년 동안의 상권 변화를 파악하기 위해 수집할 수 있는 가장 최신 데이터인 2023년 2분기와 1년 전 데이터인 2022년 2분기 데이터를 모두 사용하였다. 상권 변화에 영향을 끼치는 요인으로 상권변화지표, 아파트 단지 수, 아파트 평균 시가, 음식 지출 총 금액, 점포 수, 프랜차이즈 점포 수, 개업 점포 수, 폐업 점포 수, 총 유동인구 수, 총 상주인구 수를 활용하였다. 변수에 대한 설명은 표 1에서 확인할 수 있다.

#### 2. 데이터 전처리

수집한 데이터에서 한 개 이상의 변수값이 누락된 행이 있을 경우 결측치로 판단하여 해당 행을 제거하였다. 상권의 1년 동안의 성장과 하락을 데이터에 반영하여 상권 변화를 예측하기 위해 2023년 2분기 값에서 2022년 2분기 값을 뺀 후 백분율로 변환하여 사용하였다. 이는 상권의 변화를 보다 직관적으로 이해하기 위해 진행하였다. 이상치(Outliers)는 데이터의 분포를 왜곡시켜 모델의 학습시 문제를 유발할 수 있으므로 절댓값이 500 이상의 값을 갖는 행은 제거하였다.

중속 변수로 사용한 상권변화지표는 행정동의 변화를 생존한 사업체의 평균 영업 기간과 폐업한 사업체의 평균 영업 기간을 기준으로 4개의 등급으로 나누어 분류한 지표이다. 4개의 등급은 각각 도시재생 및 신규 개발 상권으로 상권이 크게 변화할 가능성을 보이는 다이나믹, 신규 업체가 경쟁력을 보여 상권이 확장 중인 상권확장, 기존 업체가 경쟁력을 보이며 상권이 축소 중인 상권축소, 신규 업체의 유입이 없어 상권의 변화가 거의 없는 정체로 구성되어 있다. 수집한 데이터에서 상권변화지표는 문자형으로 저장되어 있었기 때문에 0 ~ 3의 숫자로 변환한 결과, 0은 다이나믹, 1은 상권확장, 2는 상권축소, 3은 정체를 의미한다. 전처리 후

데이터의 행의 개수는 411개이다.

데이터의 다양성을 늘림으로 모델의 일반화 성능을 향상시키고, 중속 변수를 기준으로 하는 클래스의 데이터의 불균형 문제를 해결하기 위해 훈련 데이터 셋의 증강(Data Augmentation) 단계를 거쳤다. 데이터 증강을 위해 사용한 생성 모델은 기존 데이터의 특성을 고려해 새로운 데이터를 생성하는 VAE(Variational Auto encoder)이다. 상권변화지표를 기준으로 데이터를 4개의 그룹으로 나눈 후, 각 그룹의 데이터 개수가 300개가 되도록 증강시켰다. 데이터 증강을 거친 1,200개의 행으로 구성된 데이터는 모델 학습에 사용된다.

전처리한 독립변수의 분산팽창지수(Variance Inflation Factor, VIF)를 확인하여 다중공선성을 평가하였다. VIF가 10 이상의 값을 가지는 경우 해당 변수가 다른 변수들 간의 높은 상관관계를 가진다고 판단한다. 독립변수 9개의 VIF를 모두 확인한 결과 10 이하의 값을 가져 유의함을 확인하였다. 표 1은 본 연구에서 사용한 독립변수의 설명과 VIF이다.

표 1. 독립변수의 분산팽창지수

Table 1. Variance Inflation Factor of Independent Variables

독립변수	설명	VIF
아파트 단지 수	행정동별 동 단위 가구 수	1.1
아파트 평균 시가	행정동별 동 단위 아파트 평균 시가	1.1
음식 지출 총 금액	행정동별 소비자의 주정 소비(음식분야)	3.0
점포 수	행정동별 점포 수	1.9
프랜차이즈 점포 수	행정동별 프랜차이즈 점포 수	2.5
개업 점포 수	행정동별 개업 점포 수	1.2
폐업 점포 수	행정동별 폐업 점포 수	3.4
총 유동인구 수	서울시 생활인구 그리드(50×50) 배분 길 단위 유동인구 생성 후 행정동별 합	1.5
총 상주인구 수	행정동별 건물 단위별 인구수	1.3

#### 3. 모델링

9개의 유의 변수로 구성된 증강 데이터를 상권 변화 예측을 위해 딥러닝 모델에 적용하였다. 머신러닝 모델에도 동시에 적용하여 두 접근법 간의 성능 차이를 비교하였다.

모델링을 위하여 데이터 세트에서 훈련 데이터(Training Sets)와 테스트 데이터(Test Sets)를 무작위로 각각 80%, 20%의 비율로 나누어 구성하였다.

딥러닝 모델은 입력층, 은닉층, 출력층으로 구성된다. 입력층으로 입력된 데이터는 128개, 64개, 32개, 16개의 노드를 가지는 은닉층을 거치고, 출력층을 통해 4개의 값으로 분류된다. 활성화 함수로 각 은닉층은

ReLU(Rectified Linear Unit)를 사용하며 출력층에서는 분류를 위한 Softmax 함수를 사용한다. 네트워크에 사용된 하이퍼파라미터(Hyper-parameter)인 배치(Batch)는 32로 설정하였다. 테스트 데이터(Test Sets)에 과적합(Over-fitting)되는 것을 방지하기 위해 첫 번째, 두 번째, 세 번째 은닉층의 20%의 노드들을 랜덤으로 제외하여 102개, 51개, 25개의 노드의 정보만이 전달되도록 Drop-out 레이어를 추가하였다. 손실 함수는 Sparse Categorical Crossentropy이며, 최적화 함수는 Adam Optimizer이다. 총 200번의 반복 학습을 수행하였다. 분류 모델의 정확도를 검증하기 위하여 4-Fold 교차 검증(Cross Validation)을 수행한다[10]. 표 2는 딥러닝 모델에 사용된 하이퍼파라미터를 정리한 표이다.

표 2. 딥러닝 모델의 하이퍼파라미터  
Table 2. Hyper-parameter of Deep Learning Model

Hyper-parameter	Value
배치 크기	32
Drop-out	20%
손실 함수	Sparse Categorical Crossentropy
최적화 함수	Adam Optimizer
Epoch	200

#### IV. 결과 및 검증

성능 평가 지표로 정밀도(Precision), 재현율(Recall), F1-Score를 측정하였다. 정밀도는 모델이 양성으로 예측한 것 중에서 실제로 양성인 데이터를 나타내며, 예측한 데이터 중에 정답의 비율을 의미한다. 재현율은 실제로 양성인 데이터를 모델이 양성으로 예측한 데이터를 나타내며, 실제 양성 데이터를 얼마나 정확하게 예측했는지를 의미한다. F1-Score는 정밀도와 재현율 사이의 조화 평균 나타내며 모델의 종합적인 성능을 평가한다.

표 3은 본 연구에서 제안한 딥러닝(Deep Learning) 기반의 행정동별 상권 변화 예측 모델의 정량적 평가 결과를 나타낸다.

표 3. 행정동별 상권 변화 예측 모델의 정량적 평가  
Table 3. Quantitative Evaluation of Commercial District Change Prediction Model by Administrative District

Data Sets	Evaluation		
	Precision	Recall	F1-Score
Training Sets	91.81%	88.73%	89.13%
4-Fold Cross Validation	91.94%	88.61%	89.10%
Test Sets	91.84%	87.09%	88.21%

Precision, Recall, F1-Score는 100%에 가까울수록 우수한 성능을 가진다. 제안된 딥러닝 모델의 Precision, Recall, F1-Score는 테스트 셋 기준 각각 91.84%, 87.09%, 88.21%의 우수한 성능을 보이는 것을 알 수 있다.

제안한 딥러닝 모델의 성능을 평가하기 위해 머신러닝 모델과도 비교 분석하였다. 사용된 머신러닝 모델은 Decision Tree[11], Random Forest[12], SVM(Support Vector Machine)[13]이다. Decision Tree는 트리 구조의 모델로 데이터의 특성을 통해 여러 가지 규칙에 따라 분류를 수행한다. 이해하기 쉽고 해석이 용이하지만 과적합에 취약한 경향이 있다. Random Forest는 여러 개의 Decision Tree를 사용하여 분류를 수행하는 앙상블(Ensemble) 모델이다. 과적합에 강하며 데이터가 많은 경우에 좋은 성능을 보인다. SVM은 데이터를 고차원 특성 공간으로 매핑하여 클래스 간의 최적의 결정 경계를 찾는 모델이다. 고차원의 데이터에서 좋은 성능을 보인다. 테스트 셋에 대해 딥러닝과 머신러닝 모델의 성능을 정밀도, 재현율, F1 Score로 비교 분석하였다. 비교한 내용은 표 4와 같다.

표 4. 딥러닝과 머신러닝 모델의 성능 지표  
Table 4. Performance Metrics for Deep Learning and Machine Learning Models

Model	Evaluation		
	Precision	Recall	F1-Score
Decision Tree	86.26%	86.33%	86.22%
Random Forest	89.51%	88.32%	88.62%
SVM	84.37%	84.19%	84.14%
<b>Deep Learning</b>	91.84%	87.09%	88.21%

Decision Tree의 F1-Score는 86.22%, Random Forest의 F1-Score는 88.62%로 나타났다. 이는 데이터를 증강하여 사용하였기에 하나의 트리만을 만드는 Decision Tree보다 여러 트리를 만들어 내는 Random Forest에 우수한 성능을 보이는 것을 알 수 있다. SVM의 F1-Score 84.14%로 Random Forest보다 낮은 성능을 보이는데, 이는 9개의 독립변수가 동일한 중요도를 가지는 것이 아닌 다양한 가중치를 가진다는 것으로 해

석할 수 있다. 마지막으로 Deep Learning 모델의 Precision은 91.84%이며 Recall은 87.09%, F1-Score는 88.21%로 머신러닝 모델의 성능 평가 지표에 비해 전반적으로 성능이 우수한 것을 확인할 수 있다. 이는 딥러닝에서 더 많은 파라미터를 활용하여 고차원의 복잡한 특징을 학습할 수 있고, 대량의 데이터 활용이 가능하기 때문이다.

## V. 결론

본 논문에서는 딥러닝(Deep Learning)을 활용하여 행정동별 상권 변화를 예측하는 모델을 제안하였다. 2022년부터 2023년까지의 변화 정도를 수치에 반영하기 위해 백분율로 변환하였으며, 모델의 예측력에 문제가 될 수 있는 결측치와 이상치(Outliers)를 제거하는 전처리 과정을 거쳤다. 종속 변수인 상권변화지표의 클래스별 데이터 불균형 문제를 해결하기 위해 VAE (Variational Autoencoder) 모델을 활용하여 클래스별로 동일한 데이터 개수를 가질 수 있도록 증강시켜 주었다. 또한 유의한 독립변수를 탐색하기 위해 분산팽창지수(Variance Inflation Factor, VIF)를 기준으로 유의성 평가를 진행하였다. 그 결과 유의 변수는 아파트 단지 수, 아파트 평균 시가, 음식 지출 총 금액, 점포 수, 프랜차이즈 점포 수, 개업 점포 수, 폐업 점포 수, 총 유동 인구 수, 총 상주인구 수임을 알 수 있었다. 도출된 변수로 모델링 후 실험을 통해 검증한 결과, 상권 변화 예측에 머신러닝 모델보다 본 연구에서 제안한 딥러닝 모델이 보다 우수한 성능을 가짐을 알 수 있었다. 본 연구에서 제안한 딥러닝을 활용한 상권의 동적인 변화를 고려한 외식업종 지원정책은 상권의 질적 성장을 촉진하여 일자리 창출과 경제 성장에 도움을 줄 수 있을 것이라 기대한다.

향후 연구에서는 행정동보다 좁은 범위인 상권으로 수집된 데이터를 활용하여 상권별 상권 변화를 예측함과 동시에 전국으로 넓혀서 분석하고자 한다.

## References

- [1] Organization for Economic Co-operation and Development. <https://www.oecd.org/>
- [2] Korean Statistical information Service <https://kosis.kr/index/index.do>
- [3] National Tax Service (<https://www.nts.go.kr/>)
- [4] E. Choi, S. Cheon, and Su. Lee, "An Analysis of Spatial Changes in Commercial Districts using Survival-Exit Dynamics of Commercial Businesses in Seoul, Korea," *Journal of the Korean Regional Science Association*, Vol. 37, No. 4, pp. 3-19, 2021. DOI:10.22669/krsa.2021.37.4.003
- [5] Y. Nam, "Analysis on the Determinants of Exit of Self-Employed Businesses in Korea," *Bank of Korea Economic Research*, Vol. 2017, No. 5, 2017.
- [6] D. Kim, K. Kim, and Y. An, "A Study on the Spatial Characteristics of the Long-term Survival Commercial Facilities Location-Focused on Ordinary Restaurants in Gangnam-gu, Seoul," *Journal of Korea Planning Association*, Vol. 53, No. 2, pp. 161-181, 2018. DOI: 10.17208/jkpa.2018.04.53.2.161
- [7] D. Lee, J. Lee, and S. Cheon, "Comparison of Survival Rate in Food Service Industry between Growing Commercial Districts and Declining Commercial Districts of Seoul," *Korea Spatial Planning Review*, Vol. 105, pp. 65-84, June 2020. DOI:10.15793/kspr.2020.105.005
- [8] S. Kim, C. Yi, and J. Nam, "Survival Rate and Survival Factors of the Restaurants according to the Decline Level in Seoul," *Journal of Korea Planning Association*, Vol. 58, No. 3, pp. 68-81, 2023. DOI:10.17208/jkpa.2023.06.58.3.68
- [9] S. Lee, "Deep Structured Learning: Architectures and Applications," *The International Journal of Advanced Culture Technology(IJACT)*, Vol. 6, No. 4, pp. 262-265, 2018. DOI:10.17703/IJACT2018.6.4.262
- [10] S. Oh, and M. Park, "Deep Learning-based Happiness Index Model Considering Social Variables and Individual Emotional Index," *The Journal of the Convergence on Culture Technology (JCCT)*, Vol. 7, No. 3, August 2021
- [11] Y. Song, Y. Cho, Y. Seo, and S. Ahn, "Development and its Application of Computer Program for Slope Hazards Prediction using Decision Tree Model," *Journal of the Korean Society of Civil Engineers*, Vol. 29, No. 2C, pp. 59-69, March 2009
- [12] L. Breiman, "Random Forests," *Machine Learning*, Vol. 45, pp. 5-32, 2001
- [13] C. Cortes, and V. Vapnik, "Support-vector networks", *Machine Learning*, Vol. 20, pp. 273-297, 1995