

AI 를 활용한 판매 데이터 기반 매출 예측 시스템 개발

전창우, 심준성, 양시온, 윤강훈, 이동윤, 고석주

경북대학교 IT 대학 컴퓨터학부

{wjsckddn505, wnstjd732, mason0110, yghun021007, sub09065, sjkoh}@knu.ac.kr

Development of a Sales Forecast System Based on Sales Data Using Artificial Intelligence

Chang-Woo Jeon, Jun-Seong Shim, Si-On Yang, Kang-Hun Yun, Dong-Yoon Lee, Seok-Joo Koh

School of Computer Science and Engineering, Kyungpook National University

요약

소상공인의 매출은 날씨, 요일, 계절성과 같은 다양한 외부 요인에 영향을 받지만, 이를 반영한 자체 예측 시스템을 구축하기는 어렵다. 이에 본 논문에서는 K-Means 기반의 상권 분류, Prophet 을 활용한 시계열 기반 매출 예측, 그리고 XGBoost 을 이용한 외부 요인 보정을 통합한 양상을 구조를 제안한다. 시스템은 데이터 수집부터 저장, 학습, 예측까지 자동화되며, 실험 결과 Prophet 대비 XGBoost 보정 후의 MAE 가 개선됨을 확인하였다. 제안된 모델은 소상공인의 매출 예측 정확도 향상과 경영 의사결정 지원에 실질적으로 기여할 수 있을 것으로 기대된다.

1. 서론

최근 소상공인을 위한 데이터 기반 경영의 중요성이 부각되면서, 매출 예측 시스템에 대한 수요 또한 증가하고 있다. 소상공인의 매출은 수요, 계절성, 날씨[1]와 같은 외부 요인에 직·간접적인 영향을 받는다. 그러나 소상공인은 대형 프랜차이즈 매장과 달리, 이러한 변수들을 반영한 예측 시스템을 자체적으로 도입하고 활용하는 데에는 기술적, 경제적 한계가 존재한다.

이에 본 논문은 소상공인이 복잡한 예측 시스템을 직접 구축하지 않더라도 손쉽게 활용할 수 있도록, AI 기반의 매출 예측 시스템 구조와 핵심 변수 구성을 제안한다.

2. 예측 수단

2.1. K-means

매출 패턴에 기반한 상권 유형을 분류하기 위해, 각 매장의 매출 데이터를 활용하여 총 네 가지 지표(feature)를 추출하였다. 주요 지표로는 ‘공휴일과 일반 영업일 간의 매출 차이’, ‘주중과 주말 매출 차이’, ‘학기 중과 방학 기간의 매출 차이’, 그리고 ‘계절성의 강도’가 선정되었다. 상권 유형 분류에는 K-Means 클러스터링 알고리즘을 사용하였다. 클러스터링 결과는 매출 패턴의 유사성을 기반으로 주거형, 대학가형, 역세권형, 변화가형,

오피스형으로 유형화 되었으며, 군집 간 유사도 계산에는 유clidean 거리를 사용하였다.

2.2. Prophet

Prophet 은 주, 년 단위의 계절성(seasonality)뿐만 아니라, 추세(trend) 및 이벤트 효과(holiday)와 같은 변화를 반영하여 시계열 데이터를 예측하는 모델이다. 본 연구에서는 개별 매장 단위로 모델을 학습하여, 각 매장의 고유한 특성을 반영한 예측 모델을 구축한다. 또한, 클러스터링으로 구분된 상권별로 연간 및 주간 계절성의 반영 여부를 결정하여 특정 상권이 가지는 계절성을 효과적으로 모델링한다. 마지막으로, 각 매장의 계절성, 이벤트 효과, 그리고 변화점(changepoint)의 영향을 조정하기 위해 Optuna 를 활용한 하이퍼파라미터 튜닝을 수행하여 매장별로 최적의 모델을 도출한다.

2.3. XGBoost

XGBoost 는 Prophet 예측의 한계인 일시적이거나 비시계열적인 외부 요인으로 인한 오차를 보정하기 위해 도입[2]되었다. Prophet 은 시계열의 추세 및 계절성에는 강점을 보이지만, 기상 조건이나 최근 매출 변화율과 같은 외부 요인에는 민감하게 반응하지 못한다. 이를 보완하기 위해, 외부 요인의 영향을 효과적으로 학습할 수 있는 Gradient Boosting 기반의

XGBoost를 오차 보정 모델로 채택하였다.
XGBoost에 입력되는 주요 지표는 다음과 같다:

1. 날씨: 범주형 날씨 데이터를 레이블 인코딩
2. 작일 매출 변동성: 기준일과 전일 간의 매출 증감율

$$lag = \frac{(Sales_{t-1} - Sales_{t-2})}{Sales_{t-2}}$$

3. 동일 요일 간 매출 변동성: 기준 주와 전주 간 동일 요일에 대한 매출 증감율

$$weekly_lag = \frac{(Sales_{t-7} - Sales_{t-14})}{Sales_{t-14}}$$

4. 전날 요일 정보: 레이블 인코딩
5. 매장 군집: K-Means 기반 군집 번호로 인코딩
6. 주말 여부: 이진 플래그 (주말 1, 평일 0)

종속 변수는 Prophet 예측값과 실제 매출의 오차율로, 다음과 같이 정의한다.

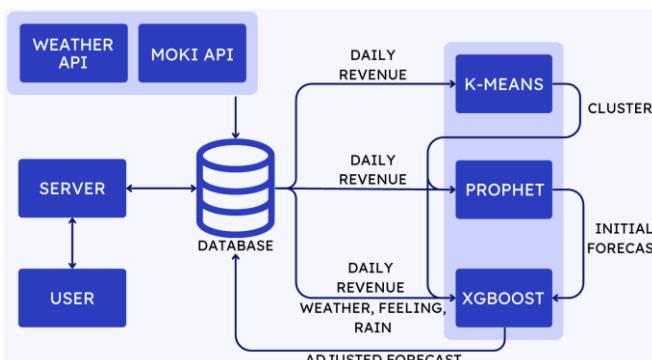
$$y = \frac{(Sales - Prophet Forecast)}{Prophet Forecast}$$

3. 시스템 아키텍처

시스템의 아키텍처는 (그림 1)과 같이 데이터 수집, 저장, 학습, 예측의 4 단계로 구성된다. 먼저, 멘토 기업인 (주)모키에서 제공하는 매장 정보 및 매출 데이터를 수집하고, OpenWeather API를 이용하여 날씨 데이터를 확보한다. 수집된 데이터는 데이터베이스에 저장된다.

이후, 저장된 데이터는 K-Means, Prophet, XGBoost 모델을 이용하여 학습 및 예측이 수행된다. 특히, 본 시스템은 주기적으로 데이터를 학습하고, 최신 정보를 기반으로 예측 값을 도출하는 스케줄링을 수행한다. 이를 통해 주기적으로 추가되는 데이터를 반영하여, 보다 정확한 매출 예측 값을 제공할 수 있도록 한다.

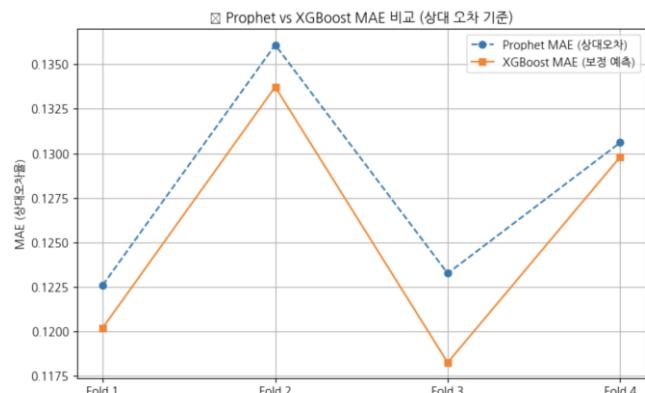
최종적으로, 예측된 매출 데이터는 데이터베이스에 저장되며, 이를 서버를 통해 사용자에게 제공한다.



(그림 1) 시스템 아키텍처

4. 실험 결과

모델의 성능 검증은 최근 4 개월을 대상으로, 각 fold 가 1 개월 단위인 K-fold 검증을 통해 수행되었다. Prophet을 단독으로 사용한 경우와 Prophet의 예측치를 XGBoost로 보정한 경우를 비교한 결과, 모든 fold에서 보정된 모델이 더 낮은 평균 절대 오차(MAE)를 기록하였다. (그림 2) 참고



(그림 2) XGBoost 적용 전후 오차(MAE) 비교

또한 전체 예측 결과의 오차율 분포를 분석한 결과 <표 1>과 같이, 약 53.9%가 10% 이내의 오차율을 보였고, 70% 이상이 15% 이내에 포함되었다. 이는 본 연구에서 제안한 모델이 예측 정확도와 안정성 측면에서 효과적임을 나타낸다.

오차율 구간	0 ~ 10%	10 ~ 15%	15 ~ 20%	20 ~ 25%	25 ~ 30%	30% 이상
비율(%)	53.8	16.7	12.4	5.57	4.02	7.43

<표 1> 구간별 오차율 분포

5. 결론 및 기대 효과

본 논문에서는 소상공인을 위한 매출 예측 시스템을 제안하였다. Prophet으로 개별 매장의 특성을 반영하고, XGBoost를 통해 외부 요인에 의한 예측 오차를 보정하는 방식으로 앙상블 구조를 설계하였으며, 이를 통해 기존 단일 모델 대비 우수한 예측 성능을 달성하였다. 이는 소상공인의 매출 예측 정확도를 제고하여 경영 의사결정의 효율성 향상에 기여할 수 있음을 시사한다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음(2021-0-01082)

참고문헌

- [1] 성은영, 성현곤, 최창규, “기상요인이 식음료업의 매출에 미치는 영향 분석”, 부동산학연구, 제23권, 제1호, 61-72, 2017.
- [2] Asim Shakeel, Daotong Chong, Jinshi Wang, “District heating load forecasting with a hybrid model based on LightGBM and FB-prophet”, Journal of Cleaner Production, Volume 409, 2023, 137130,