레스토랑 수익 예측을 위한 기계학습 모델 연구

김준석
강남대학교 ICT 공학부

배경 Introduction

새로운 레스토랑을 운영하기 위해서는 시간과 자본의 큰 투자가 필요하다. 잘못된 위치를 선정할 시에는 18개월 이내에 폐쇄되고 영업 손실이 발생한다.

지난 10년간 자영업자수는 감소하고 있지만 음식업종의 종사자수는 오히려 증가하여 음식업의 생존 경쟁은 더 치열해지고 있다. 음식업 및 숙박업은 다른 자영업에 비해 평균 생존기간이 상대적으로 짧은 것으로 나타난다. 생존기간이 가장 짧은 음식업을 가벼운 마음으로 시작하기에는 시간과 자본의 소모가 크다.

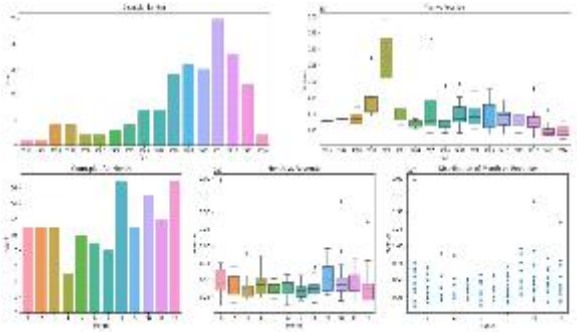
이에 본 연구에서는 요식업에서의 중요한 변수를 찾고, 랜덤 포레스트 기계학습 기법을 사용하여 레스토랑의 수익을 예측하고자 한다.

방법 Methods

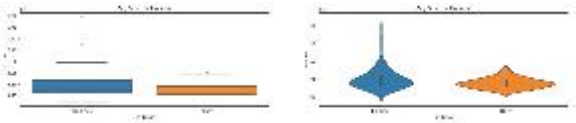
- 개장일(Open Date) 변수를 **to_date()**을 사용하여 레스토랑의 나이(Open Days)를 계산하고, int형 변수로 변경
- 범주형 데이터인 Type 변수 삭제
- 수치형 데이터로 변환시 관계성이 생기는 오류를 방지하기 위해 **get_dummies()**를 사용하여 Big Cities와 Other 변수를 가변수로 변환
- **train_test_split()**을 사용하여 Train set과 Test set의 비율을 7:3으로 분리
- **RandomForsetClassifier()**을 사용하여 변수 중요도에 따른 예측력 높은 변수를 설정
- **StandardScaler()**을 사용하여 표준화를 통해 0의 평균, 1의 표준편차를 갖도록 변환
- **PCA()**를 사용하여 특성 행렬의 차원을 축소
- **KernelPCA()**을 사용하여 선형적으로 구분
- RMSE를 통해 모델 비교 후 모델 선정
- **RandomForestRegerssor()**을 사용하여 학습

결과 Results

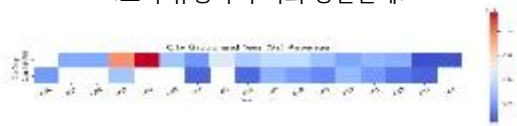
-데이터 시각화를 통해 수익 예측에 영향을 주는 변수 파악



<연월별 시간과 수익의 상관관계>

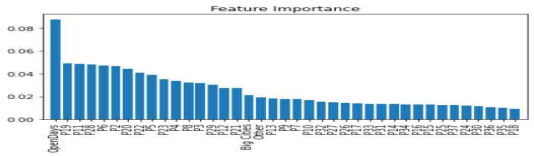


<도시 유형과 수익의 상관관계>



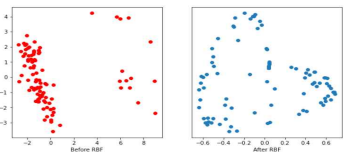
<도시 유형과 년도, 수익의 상관관계>

- 변수 중요도에 따라 상위 19개의 변수를 사용



<변수 중요도>

- 특성 행렬의 차원을 축소한 후 비선형 데이터를 선형 데이터로 반환



<RBF 커널 적용에 따른 데이터 모습>

결과 Results

- RMSE를 통해 모델 비교 후 모델 선정

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

	Models	Train RMSE
0	Lasso Regression - PCA	1.71066e+06
1	Ridge Regression - PCA	1.71067e06
2	Random Forest - PCA	1.6441e+06

- 기계학습 모델의 정확도 및 평균 오차율 계산

예측값	실제값	오차율(%)
3714896.756	2267425.0	63.837691
4486442.688	4952497.0	9.410492
4347229.640	3354383.0	29.714498

정확도(%)	85.01070702666127
평균 오차율(%)	37.6356215854514

결론 및 고찰 Discussion and Conclusion

RandomForest 기법을 통해 변수 중요도를 측정한 결과 수익에 가장 큰 영향을 주는 변수는 레스토랑의 나이이다.

본 연구에서는 수익에 영향을 주는 P-변수의 중요도만을 고려하여 내용을 설명하지 못 했다. P-변수에는 지리적 정보와 인구 정보, 그리고 상업적 정보를 담고 있다는 것 말고는 몇 번째의 P-변수가 어떠한 데이터를 담고 있는지 알 수 없다. 따라서 P-변수와 수익의 상관관계에 대한 설명이 부족하다. P-변수에 대한 자세한 정보를 알 수 있다면 수익 예측의 상관관계에 대해 파악하기 더 수월할 것이라고 생각한다.

본 연구에서는 RadomForest를 사용하여 레스토랑의 수익을 예측했지만 XG Boost나 LightGBM 모델을 사용하면 더 높은 정확도의 수익 예측이 가능할 것이라고 생각한다. 이를 통해 보다 안전한 외식업 투자가 가능할 것으로 기대된다.

ICT 공학부
202104168
김준석