Project Proposal: Predicting Restaurant Revenue

David Roh
CSCE 421 200 Spring 2023
Texas A&M University
david.roh@tamu.edu

1 Problem Objective

In the competitive restaurant industry, understanding and predicting revenue is crucial for business sustainability and growth. My objective is to develop a machine learning model that can predict the revenue of restaurants based on features such as demographic data, real estate data, and commercial data. By accurately predicting revenue, restaurateurs can be more well-informed and advised on operations, marketing, and investments to maximize their potential based on other similar, high-performing establishments.

2 Problem Significance

Accurately predicting revenue is important because profit margins are often slim in the restaurant industry and effective financial planning is crucial to success. This ability allows restaurant owners to optimize resource allocation, improve financial forecasts, and identify areas for potential growth. Accurate revenue predictions also help mitigate risks and make informed expansion decisions. As the industry continues to evolve, most notably from the pandemic, data-driven insights are becoming increasingly important for restaurants to survive in a competitive space (National Restaurant Association, 2023).

3 Current State of the Art

Previous research/projects explored using models to predict restaurant success. One notable example is a Kaggle competition that had a dataset with features such as open date, city, restaurant type, and demographic data (Kaggle, 2015). I plan on using this dataset for my project. Some public submissions achieved reasonable accuracy using models like Random Forest and Gradient Boosting. The public leaderboard's accuracy is obfuscated and top-performing models are private. Another is a paper from Stanford students that predicted the success of restaurants based on Yelp sentiment analysis, and various characteristics to predict the star rating. They found that Neural Nets and SVR were best. (Kang & Vo, 2016)

4 Proposed Approach

My approach will mainly focus on feature engineering, as many of the features are obfuscated. I plan to utilize the provided tabular data and extract relevant features from columns P1 to P37. I am considering encoding categorical variables (e.g., City Group, Type) and possibly extracting temporal features from the Open Date column.

Potentially applicable machine learning algorithms include Linear Regression, Elastic Net, KNN Regressor, Lasso/Ridge Regression, Random Forest, Gradient Boosting, and XGBoost. I will evaluate model performance using metrics such as Root Mean Square Error and F-1 score. To establish a baseline, I will compare the models' performance to a simple mean model that predicts the average revenue.

5 Data Description

The dataset includes the following columns:

- Id: Restaurant id.
- Open Date: Opening date for a restaurant.
- City: City that the restaurant is in.
- City Group: Type of the city (Big cities, Other).
- Type: Type of the restaurant (FC: Food Court, IL: Inline, DT: Drive Thru, MB: Mobile).
- P1, P2 - P37: Obfuscated demographic, real estate, and commercial data.
- Revenue: Transformed revenue of the restaurant in a given year (target).

6 Expected Outcomes

I expect to develop a model capable of predicting restaurant revenue with reasonable accuracy. I anticipate challenges in handling obfuscated data and ensuring robustness to diverse restaurant types and locations. Hopefully, some regularization method can aid in these. Moreover, I aim to identify key features that contribute significantly to revenue prediction and provide insights into factors affecting restaurant performance.

7 References

- Kaggle. (2015). Restaurant Revenue Prediction. Retrieved from https://www.kaggle.com/c/restaurant-revenue-prediction .
- Kang, S., & Vo, V. (2016). Predicting Success of Restaurants in Las Vegas. Retrieved from https://cs229.stanford.edu/proj2016/report/VoKang-PredictingSuccessOfRestaurantsInLasVegas-report.pdf .
- National Restaurant Association. (2023). Restaurant Industry Outlook 2023. Retrieved from https://restaurant.org/research-and-media/research/research-reports/state-of-the-industry/ .