Determining if there is Geographic and Cuisine bias in the MICHELIN Guide

ABSTRACT

This project examines whether or not there exists a bias for geographic location and cuisine within the Michelin Guide. To accomplish this, our project analyzes and makes predictions for data about Michelin stars/awards across different geographic locations and cuisines, using a kaggle dataset which extracts restaurant data from the Michelin Guide. Before concluding anything, we made sure to clean and prepare our data using EDA techniques to make sure we could collect insights that were accurate and reflect our goals. Afterwards, we made use of three machine learning algorithms for our predictions and conclusions. These three include: K-Nearest Neighbors (K-NN), Random Forest, and Support Vector Machine/Regression (SVM + SVR

INTRODUCTION

In the culinary world, Michelin stars are universally regarded as the pinnacle of achievement, signifying a restaurant's commitment to quality, innovation, and the art of gastronomy. One star deems a restaurant to be "worth a stop," two stars as "worth a detour," and three stars as "worth a special voyage." Additionally, it's essential to acknowledge that countries can pay Michelin for the privilege of being included in the guide, raising questions about the integrity of the Michelin Guide's stars. The purpose of our project is to explore the potential geographic bias inherent within the Michelin Guide, a French company with a global influence on dining standards and how customers spend their money on restaurants. Michelin stars influence not just the reputation and revenue of certain eateries but also the food industry as a whole. Restaurants may have negative economic effects as a result of the guide's influence over where customers choose to eat, along with affecting tourism revenue for that country or city. For this reason, it is essential to understand any variables affecting the distribution of Michelin stars for both diners looking for outstanding dining experiences and restaurant owners hoping to gain reputation.

METHODOLOGY

Data Acquisition:

Our data source is named Michelin My Maps, from the current quarter of this current year, sourced from Jerry Ng on Kaggle, scraped and curated using Golang with the Colly framework.

Data Preparation:

Some modules we used to accomplish our goal are: pandas, numpy, matplotlib, scikit-learn, and seaborn. With these, we accomplished data cleansing, visualization, analysis, and the creation of our machine learning models. To prepare our data for data analysis, we took out unnecessary columns, removed and cleaned missing values, any existing (and useless) duplicates, and ensured they were the correct data types we needed. With these steps, we made sure that our data was ready for any analysis. We decided to narrow our focus to the correlations between the 'Locations','Cuisine, and 'Price Level' types and the resulting award.

Model Selection:

K-Nearest Neighbors (KNN): We chose this algorithm because it makes no assumptions about the distribution of the data, and we can quickly prototype and explore the data. This model is highly flexible in the number of neighbors which in turn can be used to improve the performance of our model. This model is both easy to interpret and quick to identify patterns within our data.

Random Forest: We chose this algorithm due to its ability to handle overfitting, high-dimensional data well, and its ability to capture relationships with less sensitivity to noisy data and outliers. Due to the need to convert our data to numerical values for a regression model, we opted for a classification model to avoid potential conversion errors.

Support Vector Machines (SVM): We chose this algorithm based on its ability to deal with complex data. This model is ideal in that it can handle a large quantity of information regarding each restaurant.

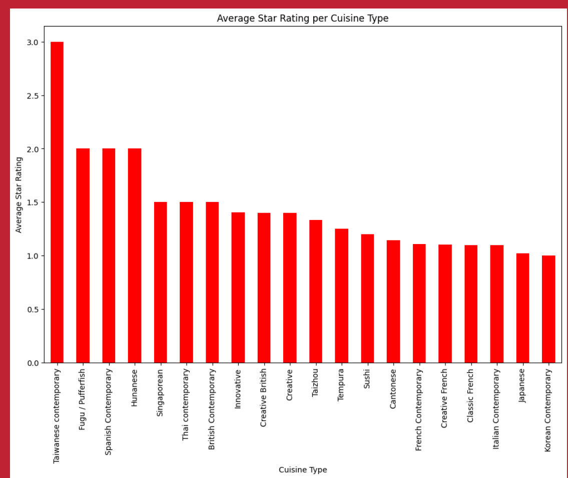
RESULTS AND EVALUATION

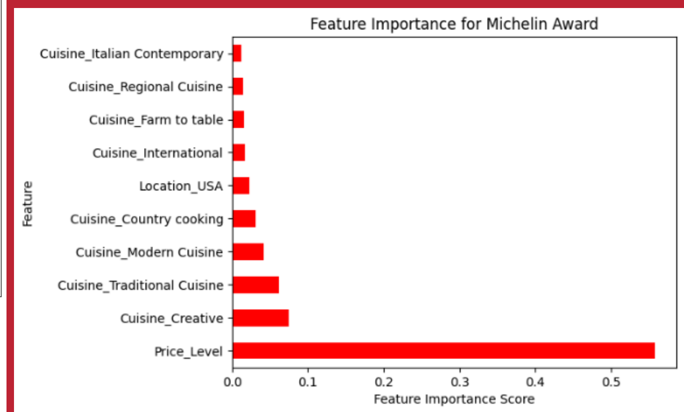
K-Nearest Neighbors: Our K-Nearest Neighbors model had a high accuracy value of 0.86. The model exhibited high recall and precision metrics for both '0 star' and '1 star' ratings. However, the performance faltered notably for '2 star' and '3 star' ratings, with both recall and precision values were significantly lower. Particularly concerning was the 0 values of both categories for '3 star' restaurants.

Random Forest: Our Random Forest classification model was accurate with a value of 0.875 post-tuning with GridSearchCV. While the overall accuracy was high, the recall value for '3 stars' was much lower compared to other award categories. This discrepancy suggests the model is missing more '3-star' restaurants.

Support Vector Machines: Our Support Vector Machine classifier demonstrated a high level of accuracy, achieving a score of 0.86. The classifier showed excellent recall and precision for '0 stars' and moderate recall and precision for '1 stars'. However, it is noteworthy that both recall and precision for the '2 star' and '3 star' categories were nonexistent, indicating that the model was unable to correctly identify any samples within these ratings.

VISUALIZATIONS





IMPACT

Michelin stars undeniably influence not just the reputation and revenue of restaurants, but also steer diners' choices and potentially impact tourism revenue. This focus on specific styles can create a two-tiered system within the restaurant industry, favoring high-end establishments while potentially hindering smaller, more casual eateries. While the Michelin guide is a valuable tool for discovering exceptional dining experiences, it might overlook hidden gems or diverse cuisines that don't fit the mold.

Furthermore, the food industry is constantly evolving. New culinary techniques and regional specialties emerge all the time, potentially creating a bias towards established traditions. For instance, Michelin's emphasis on classic French techniques might undervalue innovative approaches using local ingredients or fusion cuisine. It's important to acknowledge that Michelin has made efforts to include more diverse styles in recent years, but concerns about potential bias due to secretive selection criteria remain.

Using the predictions from our model, the food industry and restaurant owners can gain a greater understanding of the secretive criteria used by the Michelin Guide, and can demand a more transparent and equitable system that promotes and praises restaurants of all price categories, cuisines, and locations.

CONCLUSION

Our project's goals were achieved effectively as we built three machine learning models, which all produced promising accuracy rates around 85% in predicting Michelin star awards. Our Random Forest model demonstrated the highest accuracy at 87.5%. Initial values also displayed that price level had importance on predicting the award. We were unable to solve the issue with our models generally overlooking restaurants with higher ratings.

Our models can definitely still be improved upon. The dataset we used was relatively limited, constrained by the small number of Michelin Star restaurants. The Michelin Star grading criteria is also vague, so combining the dataset with qualitative consumer reviews is a path we could pursue in the future to gain an expanded understanding of the bias in restaurant quality and customer preferences.

RELATED WORK

Project/Analysis 1: Michelin Guide Dataset Analysis
Project/Analysis 2: Michelin Star Data Project
Project/Analysis 3: Michelin Star Data Collection + Analysis

BIBLIOGRAPHY

- Gill, A. A. (2012, October 3). *What's wrong with the Michelin Guide? everything, says A. A. Gill*. Vanity Fair. https://www.vanityfair.com/culture/2012/11/whats-wrong-with-the-michelin-guide
- Hayward, T. (n.d.). *https://www.ft.com/content/e622ec53-ea9f-487a-a434-747f13f5ffa0*. What's wrong with the Michelin Guide? https://www.ft.com/content/e622ec53-ea9f-487a-a434-747f13f5ffa0
- Houck, B. (2018, July 18). *The high price of a michelin guide*. Eater. https://www.eater.com/2018/7/18/17540672/michelin-guides-restaurants-tourism-bangkok-thailand-south-korea-singapore-funding
- Michelin. (2024, February 13). *Countries with the most Michelin-starred restaurants worldwide 2023*. Statista. https://www.statista.com/statistics/1400971/countries-most-michelin-starred-restaurants-worldwide/
- Michelin. (n.d.). *Starred restaurants - all starred restaurants in the michelin guide*. MICHELIN Guide. https://guide.michelin.com/us/en/restaurants/all-starred
- Rickman, C. (2021, May 14). *Is the michelin guide prejudiced?*. Frenchly. https://frenchly.us/is-the-michelin-guide-prejudiced/

Group 14: Anthony Chen, Daniel Kim, Brian Da Silva, Sarah Xie