Would You Like Fries With That?

Predicting Bestselling Food Items at Major Fast Food Resturants

Ian Alvarado, Akshay Arun, Jessica Joy, Marie Vaughan
Georgetown University

Introduction.....$1

Over one in three Americans eat fast food each day. Grabbing a burger and fries can be a fun, tasty treat, but these choices are associated with increased intake of calories, fat, and sodium (CDC). This project aims to investigate what drives consumer preferences when ordering at fast food restaurants, ultimately informing public health initiatives and business decisions. We utilize binary classification techniques to determine what items will become bestsellers and what characteristics contribute to their popularity.

Abstract.....$2

Several binary classification models were built to predict bestseller status, generating feature importance data and comparing accuracies and results amongst the models. Next, we compared the nutritional content of these bestselling items to the FDA's daily reference values, revealing how popular choices align with recommended dietary guidelines. Our top performing models identify several important macronutrients and meal types most common amongst bestsellers, shedding light on the preferences of fast food consumers.

Among the models we tested for our binary classification task, the Support Vector Machine was the top performer (Accuracy = 0.83, F1 Score = 0.71). Using permutation significance to determine which features contributed the most to bestseller status, we identified total fat, protein, and carbohydrates as influential features. Notably, these bestselling items contain higher levels of these nutrients compared to non-bestselling items, in some cases reaching about 50% of the FDA's recommended daily value in one fast food item.

Methodology.....$3

Our nutritional data for each fast food item was collected from MenuStat.org, and price information was scraped from FastFoodMenuPrices.com. To classify which items are bestsellers, we scraped data from FastFoodNutrition.org to collect the 15 most popular items for each restaurant. These files were combined into a single csv which had 18 unique restaurants and 978 menu items (275 of which were bestsellers).

Data Cleaning and Model:
Dropped 'item', 'item description', and 'restaurant' from features
One hot encoded 'food category'
Imputed missing values with median values
Support Vector Machine (SVM) Model:
- Created pipeline with SMOTE, scaler, and backward feature selection
- Grid search for best parameters
- 5 K-fold cross validation

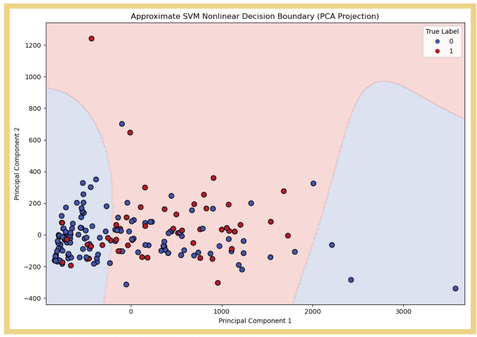


Figure 4: SVM Decision Bounary of PCA Projection Components

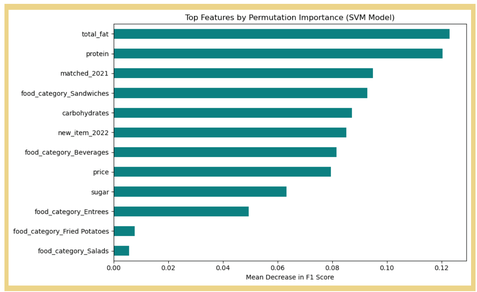


Figure 5: Feature Importances of the SVM Model

Model	Testing Accuracy	Testing F1 Score
Logistic Regression	0.65	0.53
QDA	0.77	0.53
LDA	0.78	0.53
Random Forest	0.82	0.63
SVM	0.83	0.71

Figure 3: Model Performance Comparison | Optimal Model (SVM) in Red

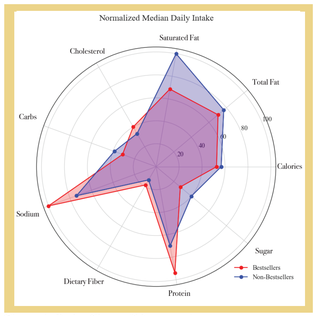


Figure 1: Radar Chart of Macronutrients

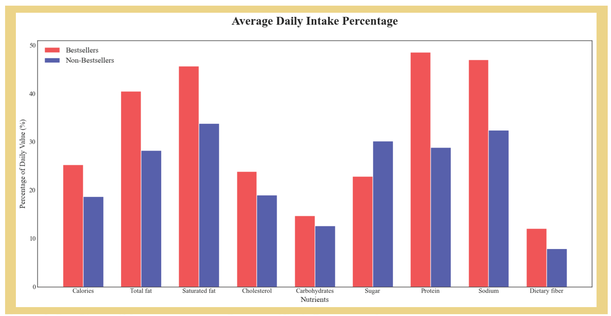


Figure 2: Bar Chart of Macronutrients as a Percent of Daily Recommendations

Results.....$4

Our methods employed several different machine learning models to classify food items as bestsellers and non-bestsellers. The performance of each of these models varied significantly, and is shown in the table above the methodology section. Ultimately, we were fairly successful in developing an accurate model.

Top Performing Model: To determine which of our models performed the best we used two metrics, accuracy and F1 score. Of all models used in our methodology, the Support Vector Machine was the best with an accuracy of 0.83 and an F1 score of 0.71 on a test data set. The associated confusion matrix is shown on this poster. The ROC and precision-recall curves associated with this model are also on display.

Feature Importance: We also display a bar chart showing the top features by permutation importance as determined by the SVM model. These features have the most significant impact on model performance when their values are shuffled around. The top two numeric features by permutation importance are total fat and protein. The top categorical feature by this metric is "matched_2021" which indicates whether a food item was available prior to 2022.

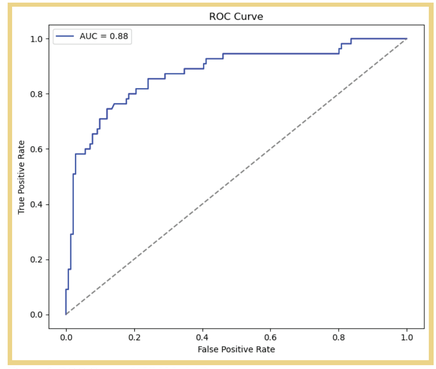
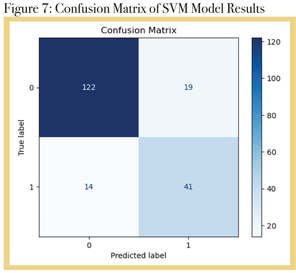


Figure 6: ROC Curve of SVM Model Results

Figure 7: Confusion Matrix of SVM Model Results



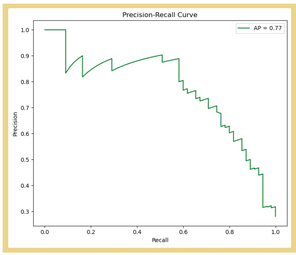


Figure 8: Precision vs. Recall of SVM Model Results

Discussion/Conclusion.....$5

We show that fast food consumers tend to prefer items with higher protein, total fat, saturated fat, and sodium. Bestselling items also tend to have less sugar. Fast food establishments can use these findings to inform new item releases, focusing on curating a menu with higher nutrient dense foods that contain more protein and less sugar. This can nudge customers toward making healthier food choices without hurting sales. Since some bestseller items are also high in both total and saturated fat and sodium (many items making up between 40 - 50% of the daily nutritional value), fast food restaurants should also be liable to transparency in food marketing. More awareness needs to be made of how fast food restaurants market indulgent items and accountability taken in how foods are promoted to the public.

References.....$6

1. "Fast Food Restaurants Nutrition Facts." Fast Food Nutrition, fastfoodnutrition.org/fast-food-restaurants.
2. Fryar, Cheryl, et al. "Fast Food Consumption among Adults in the United States, 2013–2016." CDC, Oct. 2018, www.cdc.gov/nchs/products/databriefs/db322.htm.
3. "Restaurant Nutrition Data: 2022 Annual Data." MenuStat, 2015, www.menustat.org/data.html. Accessed 24 Apr. 2025.
4. "THE TOP 50 FAST FOOD ITEMS - the RINGER." The Ringer, fastfood.theringer.com/.
5. NFL DRVs Food Components, Office of the. U.S. Food and Drug Administration, FDA, www.fda.gov/. Accessed 24 Apr. 2025.

Contact: iam35@georgetown.edu, aa2627@georgetown.edu, jj1129@georgetown.edu, mev71@georgetown.edu