

Technology Used



Smunch User Segmentation and Meal Preference Analysis

Smunch is a B2B food service based out of Berlin. Subscribing companies are given 4 or 8 meal options per day.

My project is to identify unique user clusters based on ingredient preferences and predict the successfulness of a meal based on each users preferences. This allows Smunch to better understand their customer base and provide a wider reaching, superior product.

Clustering

What Smunch knows about their users:

1. 10% always choose meal with meat.
2. 11% never choose meat (vegetarian).
3. 2% only eat vegan.

What about the **missing 79%**? What do they like?

Kmeans/Hierarchical clustering revealed 3 distinct ingredient preference groups.

Veg/Vegan 39%



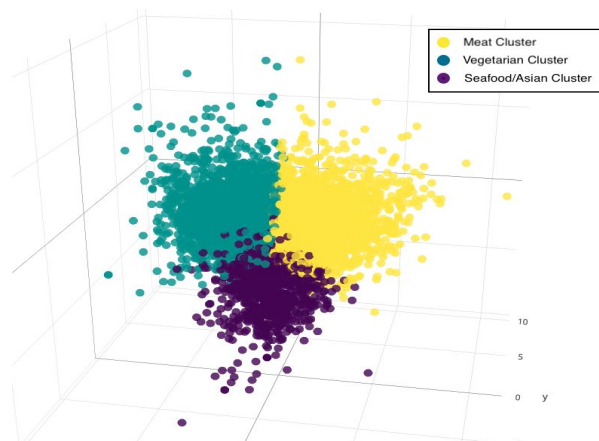
Meat/Dairy 49%



Asian 11%



Using PCA and tSNE Dimensionality Reduction techniques these clusters can be viewed in 3D.



Order Prediction

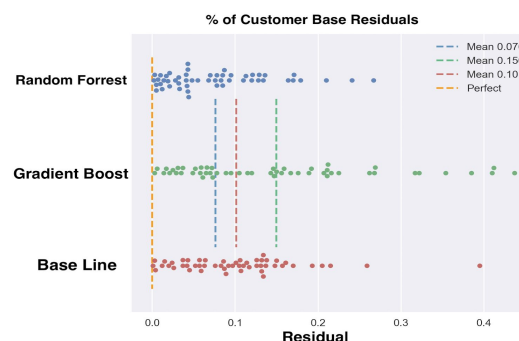
1. Will a new meal perform well?
2. Providing a food option for every group?

Attempt 1: Ratings are biased! Ratings, especially bad ones, often have nothing to do with the actual food. After testing Random Forest, Gradient Boost, and Neural Network, the lack of signal became clear and another target feature would need to be created.

Attempt 2: Every user is different! Every user has unique ingredient preferences and chooses one of the offered meals based on them. Using a created target value called C-Score I was able to utilize a Random Forest model to predict the "successfulness" of a dish.

$$C - Score = \# \text{ times meal ordered} / \# \text{ times meal offered}$$

Model Performance: **25% Better** than base



Model Output:

- % of customer base that will order each meal

Model	Meal1 %	Meal2 %	Meal3 %	Meal4 %
RF Pred	24.5	9.1	17.8	48.5
Base	25.0	25.0	25.0	25.0
Actual	33.3	8.0	18.7	40.0

Noah Schumacher

noahschumacher14@gmail.com | + 1415-250-3166 | Santa Monica, CA

[Github](#), [LinkedIn](#)

Objective

To use my skills in Machine Learning, Data Engineering and Statistics as well as my aptitude for critical thinking, problem solving, and communication, to be an effective and efficient Data Scientist.

Skills & Abilities

- **Tools:** Python, SQL, MongoDB, Spark, Flask, AWS, Git, Linux, Docker, Qt.
- **Analysis:** A/B Testing, Frequentist and Bayesian Statistics, ROC and Profit Curves
- **Machine Learning:** KNN, Logistic Regression, Linear Regression, Decision Trees (RF, Boosting, Ada), SVM's and Kernels, Time-Series Analysis, NLP, Neural Networks, Recommendation Systems, Dimensionality Reduction

Experience

Galvanize | San Francisco CA

January 2019 - Apr 2019

- Capstone:
 - **Smunch User Segmentation and Meal Preference Analysis:** a comprehensive user analysis project with two main functions. First cluster the users based on their ingredient and nutritional preferences and identify the size and key ingredients for each unique cluster. Second, predict the “successfulness” of a meal based on each customers learned ingredient preferences. Prediction is wrapped up in a Web Application that Smunch can use to test meal ideas, explore customer preferences, and understand their quickly expanding customer base. [link](#)
- Case Studies:
 - **Fraud Detection** analyzed event postings features to give a soft classification of the event being fraudulent. This study utilized a Random Forrest model and NLP techniques to make a prediction. Final Product was wrapped into a Web App built with Flask and hosted on AWS. [link](#)
 - **Movie Recommender** takes user and movie information and ranks top predicted movies for a particular user. The final model utilized ALS Matrix factorization for movie rating predictions. Other attempted models included ensemble of recommending popular movies and Random Forrest. [link](#)
 - **Churn Prediction** study utilized a Gradient Boosted model to predict users that are at risk of churning. Validated model with ROC plots and then used the model to create Profit Curves for hypothetical business solutions. [link](#)

Solutions Consultant, FarApp | Santa Monica, CA

October 2018 - December 2018

- Managed client integrations and implementations.
- Assisted customers with technical problems and issues with product. Received a 100% customer satisfaction score with several customer comments on helpfulness and efficiency.
- Collaborated with engineering team to build and deploy features and fixes of platform. One feature allowed customers to automatically set shipping methods based on weight; improved customer satisfaction and ease of use.

Data Science Associate, Smunch GmbH | Berlin, Germany

June 2017 - Sept 2017

- Implemented A Linear Regression model to predict total number of expected meal orders per day. Model was deployed on the backend website for operations team to use with interactive parameter selection.
- Utilized Python, SQL, Ruby on Rails, Docker, Qt, Scipy, and ML techniques.
- Completed various mini projects such as retention rates, company cohort breakdowns, and informed CEO of specific trends.

Summer Intern, Stanford Linear Accelerator Laboratory | Palo Alto, CA

June 2015/16 - Sept 2015/16

- Member of the Photon Control and Data Systems group for two summers completing several projects.
- Designed and installed a Vacuum valve fail-safe switch with pneumatic monitoring capabilities. Was implemented in experimental hutch and designs expected to be installed in several other hutches.
- Wrote serial driver protocol file for sample delivery system in experimental hutch.
- Designed Pneumatic safety system for each of the Experimental Hutches in LCLS. Extensively researched Pneumatic controls and feedback systems. Presented cost-effective solution to group for further review.

Education

Galvanize Inc.

Graduated: Apr 2019

Data Science Immersive

- Intensive 700+ hour Data Science immersive program in San Francisco
- Solidified skills in Programming, Statistics, and Machine Learning

University of California, Santa Barbara

Graduated: June 2018

Bachelors of Science, Physics

GPA: 3.32/4.0