ML4VA Restaurant Recommender

By Ali Ibrahim, Andrew Lin, Kaleb Getachew

Motivation

- The current dire financial situation of many small businesses in Virginia due to COVID-19 inspired this project. By recommending local restaurants, we aim to relieve some of the financial hardship that is being experienced right now.

- In addition, when people visit a new city, they will most likely want to eat at a local restaurant. This project will aim to solve is the universal problem of selecting a compatible local restaurant based on preferences and attributes.

Project Planning

Data Gathering/Cleaning

Gathered and Merged Data

We were able to survey a diverse group of people to get relevant information regarding our prediction model.

Model Selection/Building

Model Building

We chose an array of ML models to conduct our recommendations.

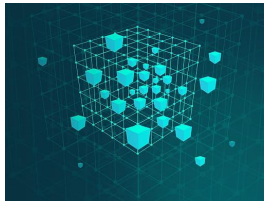
Deployment

Built in Django, Deployed on Heroku

The trained models pickled from Google colab and unpickled into Django where the web application was built and then deployed using Heroku.

Data Gathering/Cleaning

- The original dataset we used came from the UCI Machine Learning Repository
  - This dataset had only 138 observations
  - The data was collected from people based in Mexico
    - For the cuisine preferences, there were over 100 categories



- We grouped the cuisine categories into 10 categories: “Mexican/Southwestern/Latin American”, “American”, “Asian”, “African”, “Austronesian”, “European”, “Modern”, “Desserts/Cafe”, “Italian”, Mediterranean/Middle East”

- After condensing the categories to only 10, we realized it was severely imbalanced towards the category of “Mexican/Southwestern/Latin American” cuisines.
  - **Solution:** Conducted our own survey using the same parameters from the UCI dataset, resulting in 120 new observations to append to the pre-existing dataset.
  - In total, we had 258 observations

Getting More Observations For the Lagging Categories

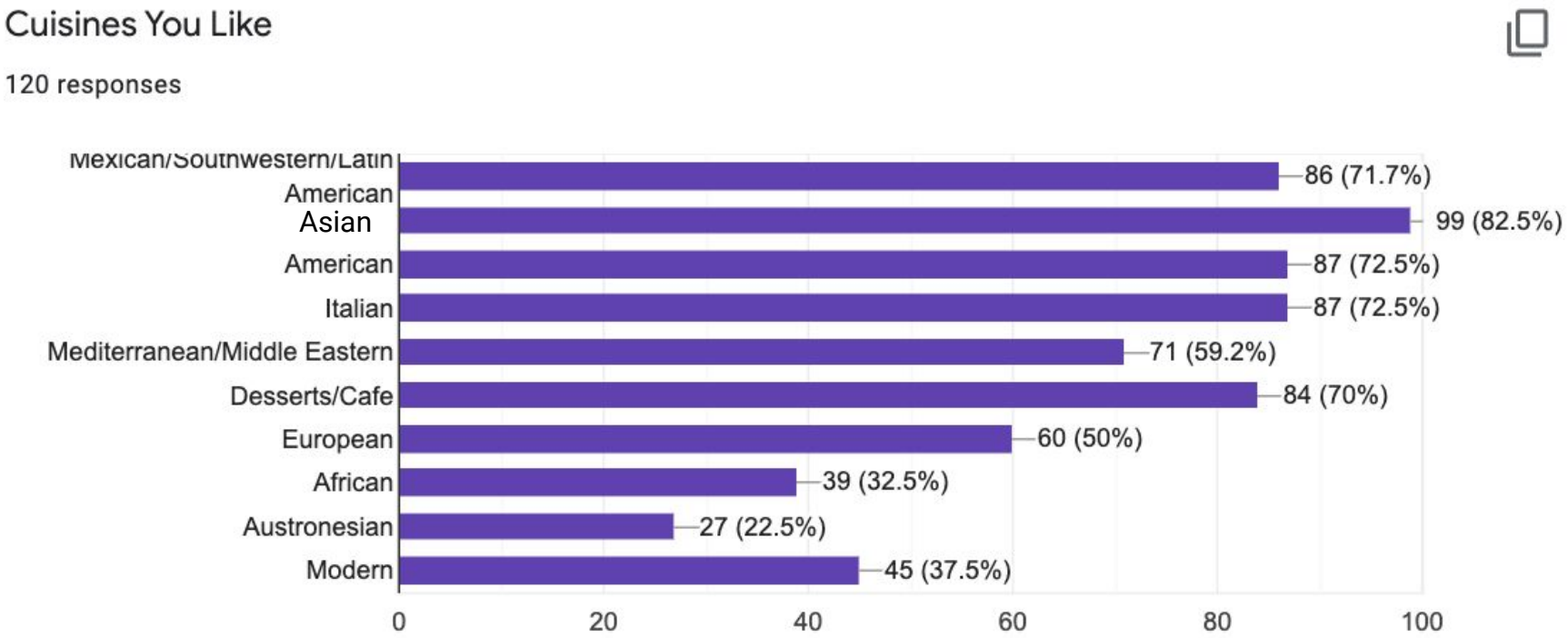
Original Dataset

```
Mexican/Southwestern/Latin American    116
American                                57
Asian                                   33
Desserts/Cafe                           30
European                                24
Modern                                  24
Italian                                 16
Mediterranean/Middle East               14
Austronesian                             9
African                                  7
Name: Rcuisine, dtype: int64
```

Original Dataset combined with Survey

```
Mexican/Southwestern/Latin American    202
American                               144
Asian                                  132
Desserts/Cafe                          114
Italian                                103
Mediterranean/Middle East               85
European                                84
Modern                                  69
African                                 46
Austronesian                            36
Name: Rcuisine, dtype: int64
```

The Answers For Cuisines People Liked On the Survey



Model Selection/Building

- We condensed Asian and Austronesian together and African and Mediterranean/Middle East together to train the model since the models had a difficult time learning African and Austronesian by themselves.
  - In total, we built 8 binary classifiers to classify each the 8 categories.

- For each Cuisine model, we trained:
  - Random Forest Classifier, Logistic Regression, Support Vector Machine, Adaboosted Decision Tree, GradientBoostingClassifier, Bagging Decision Tree Classifier, and Bernoulli Naive Bayes Classifier

- We used GridSearchCV and RandomizedSearchCV where we saw fit to find the optimal hyperparameters for each model

- The loss used was “balanced accuracy”, which helped with the imbalance of the labels

- We then combined a different combination of those tuned models into a VotingClassifier for each of the cuisines

Deploying the Model to a Website

- Although we built and trained the models in Google colab, we built the website using the popular web framework Django.
- We used the python pickle module to serialize and unserialize our models one by one and the pipeline from colab and into Django.
- Once everything was working as expected locally, we deployed the website using Heroku. It's live at ml4va.herokuapp.com/predictor for anyone to visit.
- In the future, we'd like to add more locations throughout Virginia, which will only require updates in Django and not the actual models.

ML4VA Restaurant Predictor

Classify Me

Location*
Charlottesville

Smoker*
True

Drink level*
Abstemious

Dress preference*
Formal

Ambience*
Family

Transport*
Car Owner

Marital status*
Single

Children*
Independent

Interest*
Technology

Personality*
Hard Worker

Religion*
Christian

Activity*
Student

Fav color*
Red

Weight*
Enter your weight in kilograms

Budget*
Low

Height*
Enter your height in meters

Age*
20

Get Recommendations

Here are the cuisines we think you'll like.

American
Italian
Dessert/Cafe
Asian
Modern
Mediterranean/Middle East
European

http://thenewyorkerdeli.com/
https://bellacinos.com/
http://www.alexandersva.com/

https://www.luigisvirginia.com/
http://www.franksatbrambleton.com/Home.htm
http://www.fortunatoroanoke.com/

http://littlegreenhive.com/
https://www.millmountaincoffee.com/
http://sweetdonkeycoffee.com/

http://www.benguisushi.com/
http://redjasmine.net/
https://www.facebook.com/vinacaferoanoke/

https://riverandrailrestaurant.com/
https://petersseafood.com/
https://www.wasenacitytaproom.com/

https://www.facebook.com/pages/category/Mediterranean-Restaurant/Bethlehem-Restaurant-Grocery-671666456276493/
http://www.cedarsofroanoke.com/
https://www.petramedgrill.com/

Challenges

- The first challenge was working with an initially small dataset. The recall was high at first, but the true negative rate was very low. This meant that our model didn't have enough observations where the label was 0 to learn from, as a result it was only able to perform as good as the data it was given.

- The dataset was small, so we decided to conduct a survey to gather more data.
  - Cleaning took a very large amount time

- After trying the models we learned in class, the performance was not less than ideal.
  - We conducted some research into other models that worked well with smaller datasets and one such model we found was the Naïve Bayes Classifier.

Evaluation Metrics

```
[[33  3]
 [ 3 13]]
Precision: 0.81
Recall : 0.81
Accuracy :0.88
F1 Score :0.81
AUC: 0.8645833333333333
```

American Cuisine Model

```
[[32  3]
 [ 5 12]]
Precision: 0.8
Recall : 0.71
Accuracy :0.85
F1 Score :0.75
AUC: 0.8100840336134454
```

Desserts/Cafe Cuisine Model

```
[[40  3]
 [ 4  5]]
Precision: 0.62
Recall : 0.56
Accuracy :0.87
F1 Score :0.59
AUC: 0.7428940568475453
```

European Model

```
[[37  5]
 [ 6  4]]
Precision: 0.44
Recall : 0.4
Accuracy :0.79
F1 Score :0.42
AUC: 0.6404761904761904
```

Modern Model

```
[[ 7  6]
 [ 8 31]]
Precision: 0.84
Recall : 0.79
Accuracy :0.73
F1 Score :0.82
AUC: 0.6666666666666665
```

Mexican/Southwestern/ Latin American Cuisine Model

```
[[29  4]
 [ 6 13]]
Precision: 0.76
Recall : 0.68
Accuracy :0.81
F1 Score :0.72
AUC: 0.781499202551834
```

Asian/Austronesian Cuisine Model

```
[[30  2]
 [ 8 12]]
Precision: 0.86
Recall : 0.6
Accuracy :0.81
F1 Score :0.71
AUC: 0.76875
```

Italian Cuisine Model

```
[[32  4]
 [ 7  9]]
Precision: 0.69
Recall : 0.56
Accuracy :0.79
F1 Score :0.62
AUC: 0.7256944444444444
```

Mediterranean/ Middle Eastern/African Cuisine Model

Next Steps

We noticed that our models performed much better after being trained on the new, larger dataset. This indicates that this a problem that requires a larger dataset. The next steps from here would be to continue to collect data and train that data on the model.

Credits

This work has been a part of the “Machine Learning for Virginia” project at the University of Virginia.

- *Thank you to everyone who filled out the survey!*
- Original Dataset from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data
- Roanoke Star Background Image
  - https://www.roanokeva.gov/ImageRepository/Document?documentID=13604
- Roanoker Restaurant Background Image
  - https://theroanoker.com/downloads/250/download/The-Roanoker-restaurant.jpg?cb=3fe076be7ff6b958e5271d05edf36ddf