RUBRIC
Alonzi, DS 4002

DS 4002- Fall 25- Zoe Gates
Due: December 8th
Submission Format: Upload link to Github repository on Canvas

Individual Assignment

Why am I doing this?
This case study allows you to explore new methods of classification to expand upon your understanding of data science in real life applications. It will guide you in understanding how to make smart analytical decisions about how to handle messy data in order to achieve your desired product.

What am I going to do?
This Github repository for this case study can be found at https://github.com/gatesz33/CS3-DS4002. You will obtain our scraped data using this link. Then you will use our Python scripts in order to clean and parse the data. You will then utilize Sci-Kit in order to implement TF-IDF vectorization for the remaining ingredients and cooking methods. You will then build and train a Support Vector Machine with your vectorized variables. You will cross-validate your results and print their accuracies and confusion matrices in order to understand points of confusion for the model. Then, you will use the same model to build and train a multinomial linear regression. Similarly, you will cross validate and print accuracy scores and confusion matrices. These outputs will allow you to assess the strength of both of the models and understand how the model can be adjusted to achieve higher accuracy and reliability in the future. Deliverables Include:

- Well documented, code
- Confusion Matrices for both of the models– include comments on what you see
- A link to a Github Repository with all utilized materials

How will I know I have Succeeded? You will meet expectations on CS3 Create Case Study when you follow the criteria in the rubric below

Spec Category	Spec Details
Formatting	• One Github Repository (submitted via link on Canvas) ○ Create a new Github repository titled 'CS_Asian_Recipe_Classifier_

	4002.’ It should contain ■ README.md ■ LICENSE ■ Script titled ‘asian_recipe_modeling’ ■ Your data (the scraped recipes) ■ Output folder with confusion matrices (images)
README.md	● The README.md show contains a brief description of the steps you took and the outputs you achieved (including accuracy) while working through the case study. You should include a brief description of how to navigate the repository.
Scripts/Supplemental Documents	● Scripts used to build and train the SVM and Multinomial in Jupyter or Google Colab ○ Ensure you’ve commented it out, so that somebody picking up the file can understand the steps you took to achieve your results. ● Confusion Matrices from the models

Acknowledgements: Special thanks to Professor Alonzi for providing the rubric formatting!