# A Machine Learning Analysis of Recipe Reviews and User Feedback

By: Rowan Amanna

## Introduction

The problem that I am aiming to tackle is doing an in-depth analysis on recipe reviews and user feedback (UCI Machine Learning Repository, 2023). This review will help analyze and predict the quality of the recipes from user reviews, as well as look at other features which may help to predict this. With a wide amount of data present, these models can help understand and interpret the data better, while also giving a new perspective.

## Data

The dataset used in my project is about recipe reviews and user feedback. This specifically focuses on the user's reputation, reply count, thumbs up data, as well as the high score in terms of the reviews. This dataset focuses on giving a statistical perspective on how users interact with certain recipes and their feedback. When it came to Preprocessing the steps, I followed included handling missing data and scaling using standardization.

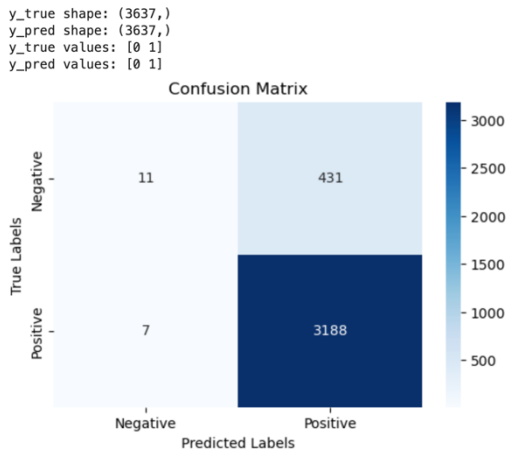
## Methods

For the two methods, I chose the logistic regression and K-Nearest Neighbors. When it came to the logistic Regression, I chose it due to its efficiency and compatibility for the classification problems. This model's goal is to predict the probability of a recipe rated on a Positive or a Negative scale. For my second method, which was KNN the main goal was to analyze the data and classify them based on the majority vote. These two models were trained, validated and evaluated, which also helped in terms of accuracy and efficiency.

**Results**

The two models were trained using x_test and y_test data, with other metrics to help with performance. When it came to testing the predictions were compared to the real values and then checked for accuracy and performance. The results of both models showed a fairly accurate performance by both models and helped in finding a accurate way of predicting the data(UCI Machine Learning Repository, 2023).

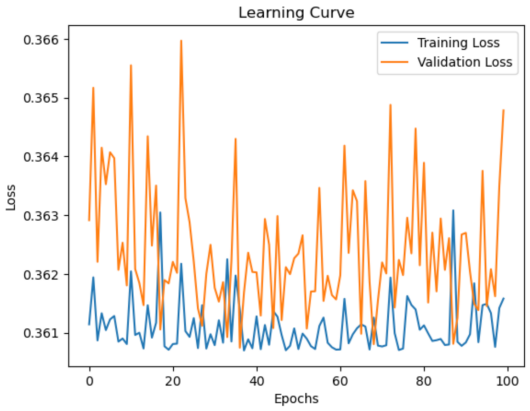
```
Accuracy: 0.88
Classification Report:
              precision    recall  f1-score   support

    Negative       0.61      0.02      0.05       442
    Positive       0.88      1.00      0.94      3195

    accuracy                           0.88      3637
   macro avg       0.75      0.51      0.49      3637
weighted avg       0.85      0.88      0.83      3637
```



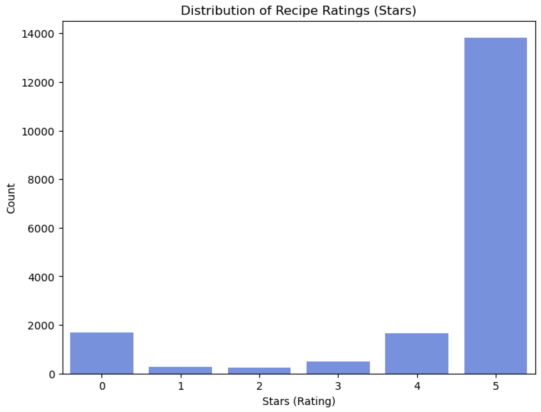
**Discuss observation and analyze them**

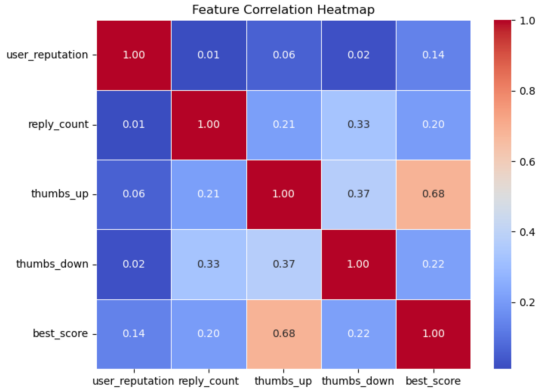
In terms of observations made, the logistic regression model showed good overall accuracy but when it came to analyzing part of the data it would have some discrepancies. This was also the case in the KNN model. Since most of the recipes were labeled as positive, this created some fluctuations in the data.

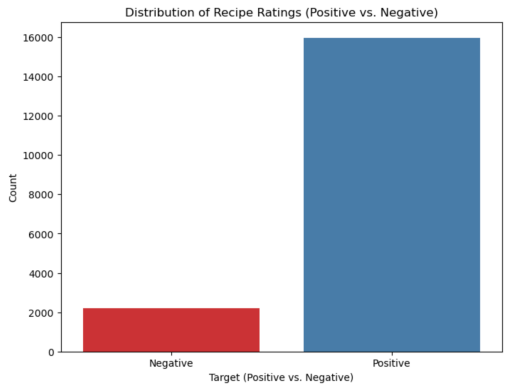
```
Training Accuracy: 0.8803
```




```
Test Accuracy: 0.8793
```






## Conclusion

This project was the culmination of all the things I learned in this class so far. While it was still introductory, it gave me a better grasp of how these models work and how they can be used to help various fields. In this project, I got to put my current knowledge of these models to the test with real world data. In my case, this was through recipe and user feedback data(UCI Machine Learning Repository, 2023). While at first, I found it overwhelming to find a suitable dataset, it was also cool to learn more about the dataset and how the models can be used to predict things within it. Furthermore, through this, I can also implement various other models as well as to my existing ones, to see if I could get a more accurate and efficient model.

## References

Ali, Amir, Stanislaw  Matuszewski, and Jacek Czupyt. "Recipe Reviews and User Feedback." UCI Machine Learning Repository, 2023, https://doi.org/10.24432/C5FG95.

**Acknowledgement**

I also referred to the homework's we did in the class to create the models. I used ChatGPT for help loading my dataset, as I had errors importing it.