

Classifying Recipe to increase Site Traffic

A solution for Zesty Bytes by Khaleed Oyeleke

12/04/2023

Business Goals

- To Increase Website Traffic: Utilize data science capabilities to identify and feature recipes on the homepage that are most likely to increase overall website traffic by up to 40%.
- Business Requirement(KPI): Develop a model that can accurately predict high-traffic recipes with at least 80% precision threshold to enhance the effectiveness of homepage content selection and drive more subscriptions.
- Actionable Insights for Strategy: Provide recommendations and strategic next steps based on the model's findings to inform content curation and decision-making processes.

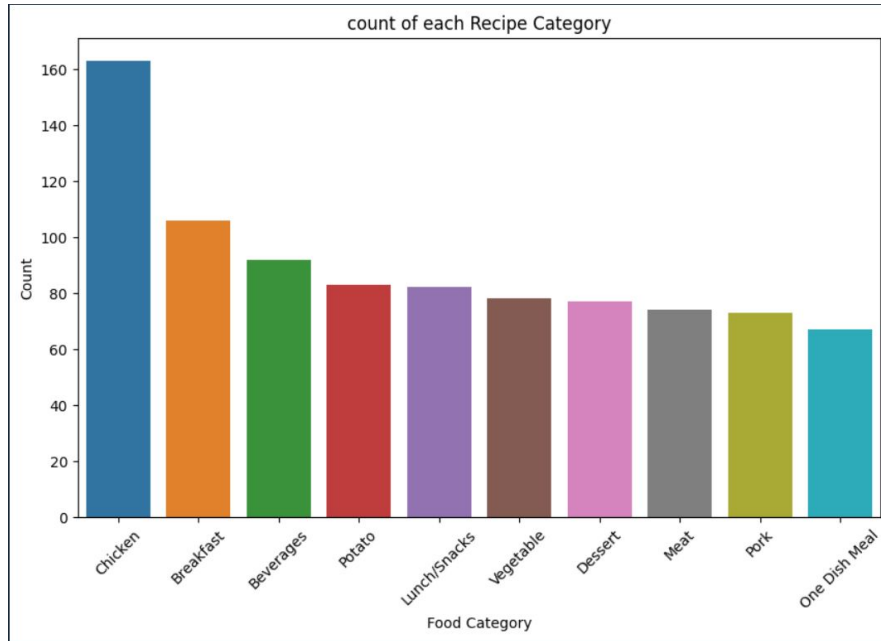
Data

Data for each recipe, as well as whether there was high traffic when the recipe was on the home page.

recipe ▾	calories ▾	carbohydrate ▾	sugar ▾	protein ▾	category ▾	servings ▾	high_traffic ▾
0	1	null	null	null	null Pork	6	High
1	2	35.48	38.56	0.66	0.92 Potato	4	High
2	3	914.28	42.68	3.09	2.88 Breakfast	1	null
3	4	97.03	30.56	38.63	0.02 Beverages	4	High
4	5	27.05	1.85	0.8	0.53 Beverages	4	null

recipe ▾	calories ▾	carbohydrate ▾	sugar ▾	protein ▾	category ▾	servings ▾	high_traffic ▾
1	2	35.48	38.56	0.66	0.92 Potato	4	High
2	3	914.28	42.68	3.09	2.88 Breakfast	1	Not High
3	4	97.03	30.56	38.63	0.02 Beverages	4	High
4	5	27.05	1.85	0.8	0.53 Beverages	4	Not High
5	6	691.15	3.46	1.65	53.93 One Dish Meal	2	High

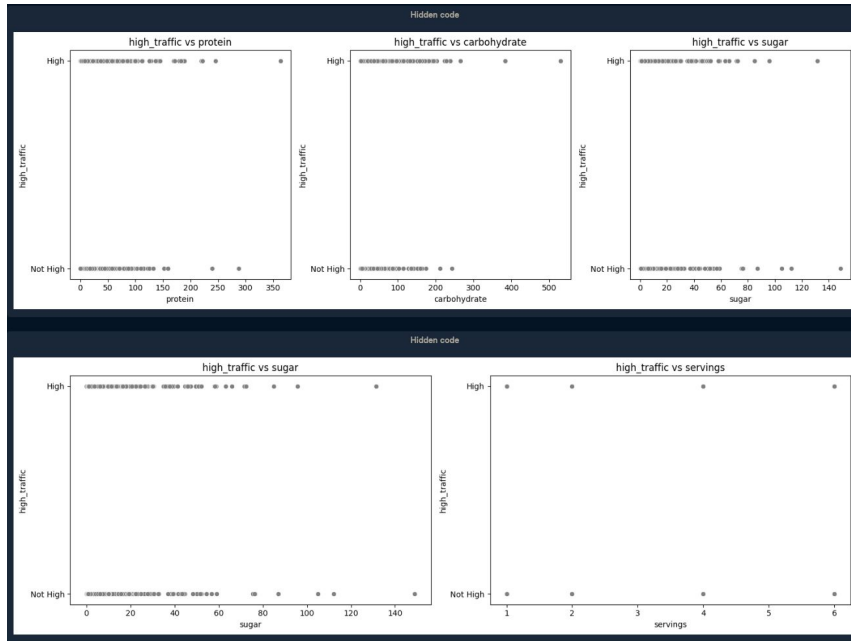
Key Findings



In this dataset, we have 10 categories of Recipes.

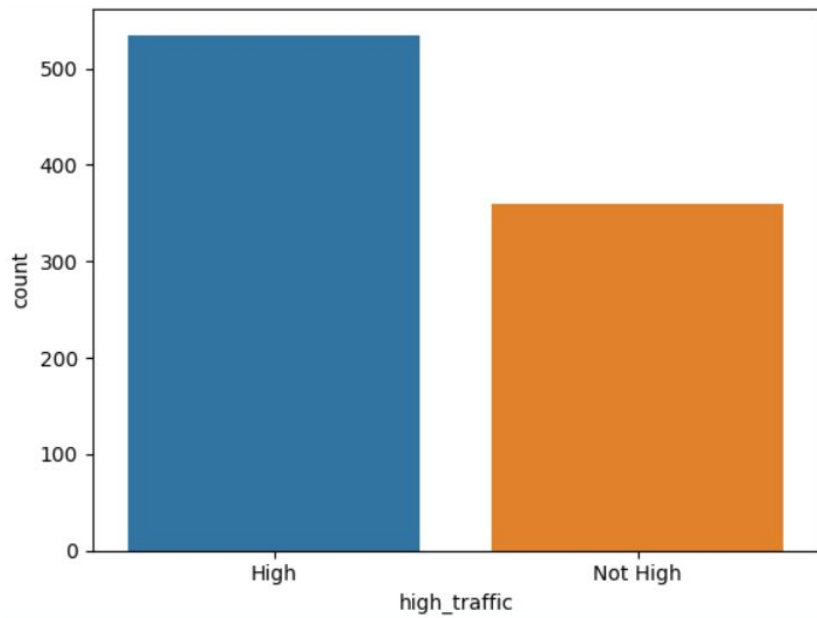
This is the total count for each recipe category. We can see Chicken recipe was served considerably the most.

Key Findings



The relationship between the numeric columns and high traffic is a non-linear relationship.

Key Findings

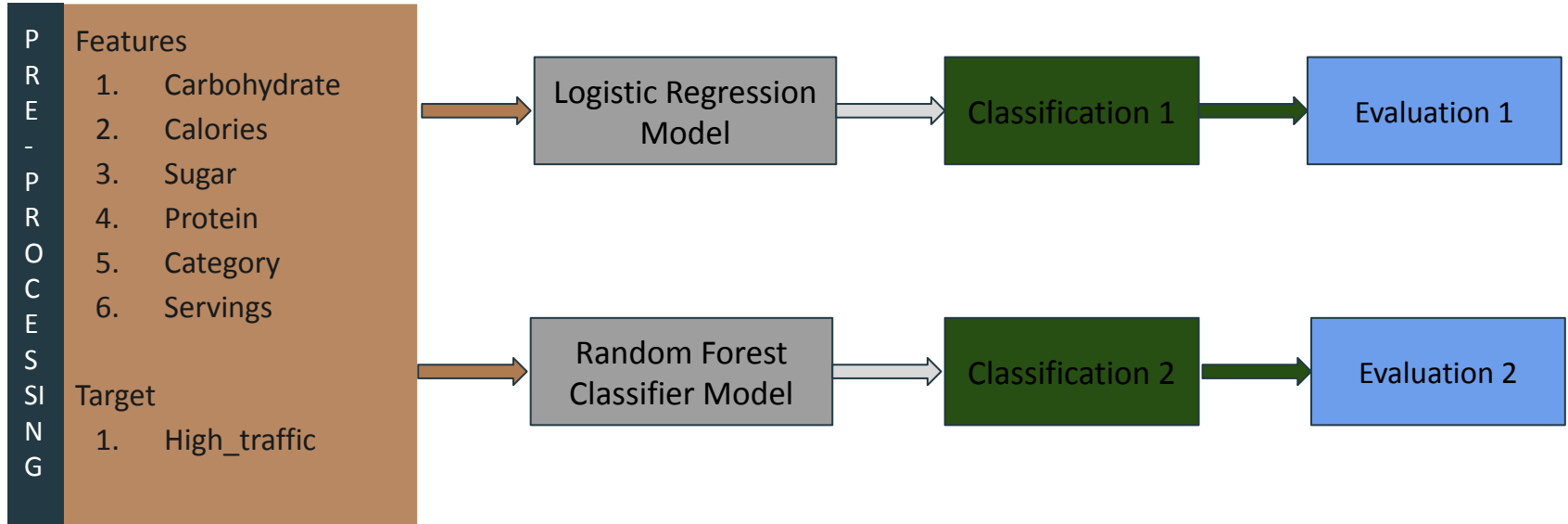


The count of High traffic to Not-high traffic.

High traffic recipes were considerable more than Not-high traffic recipes.

Outcomes

Two Models - Logistic Regression and Random Forest Classifier to classify recipe into two groups as high traffic or Not and evaluate the results.



Outcomes

Two evaluation metrics - Accuracy and Precision

Metrics	Accuracy	Precision
Purpose	Accuracy shows how often our recipe predictions are correct. The higher the better.	Precision tells us how many of the recipes we predicted as high traffic hits were actual hits. The higher the better.
Range	1-100%	1-100%

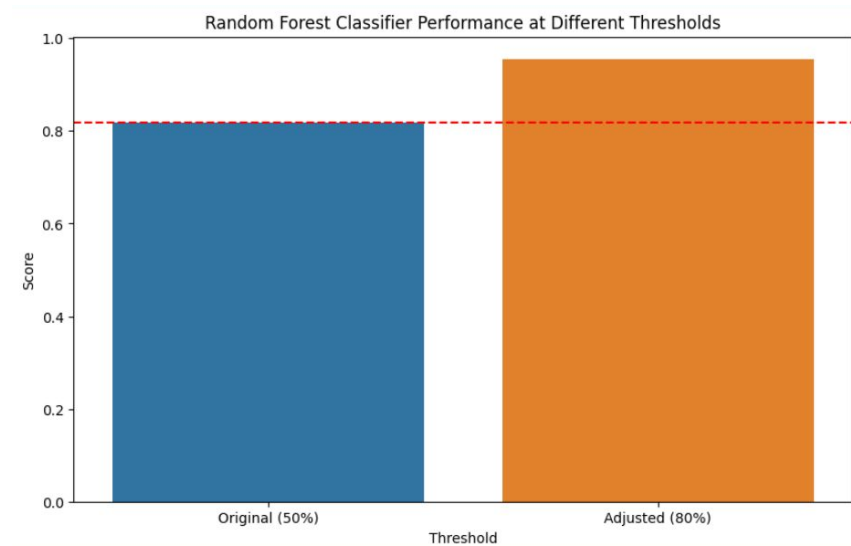
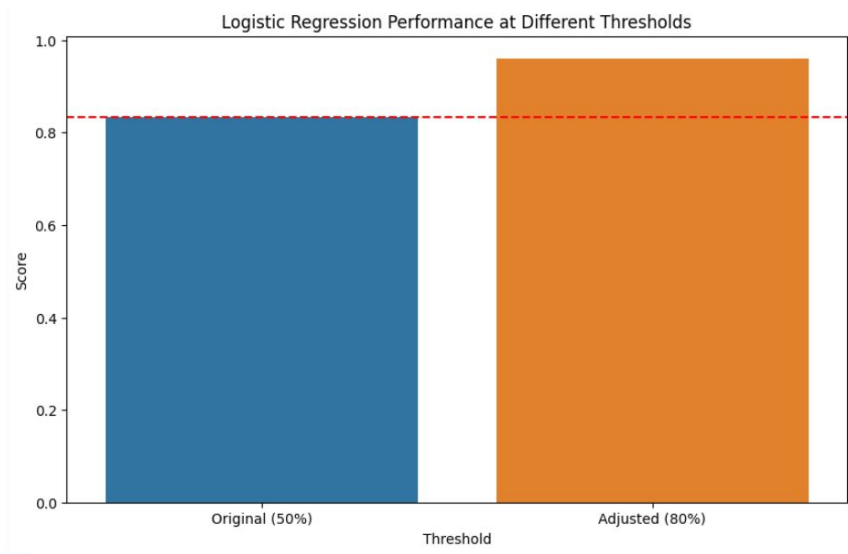
Outcomes

Two evaluation metrics - Accuracy and Precision

Model	Accuracy	Precision
Logistic Regression	78%	83%
Random Forest Classifier	78%	82%

Outcomes

Business Requirement(KPI) - Correctly predict high traffic recipes 80% of the time



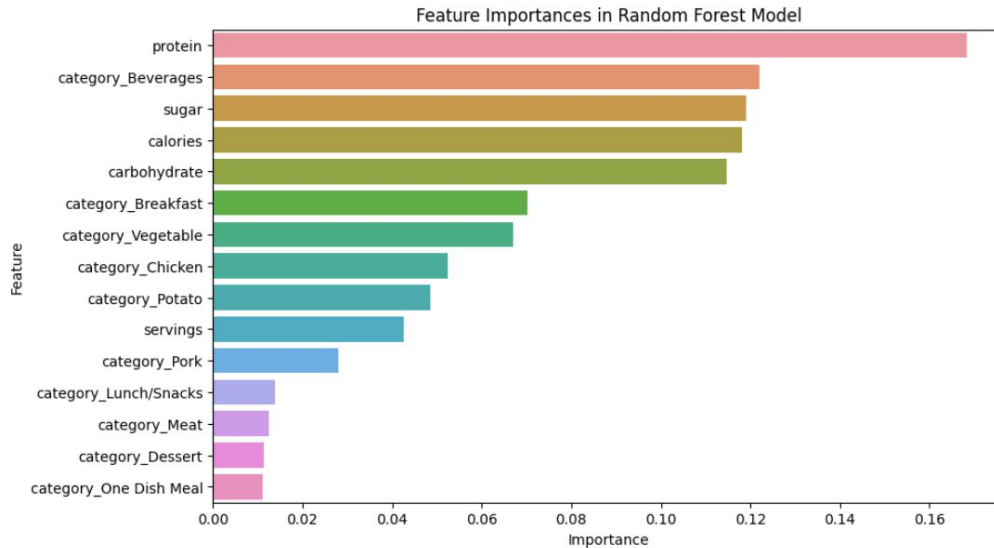
Outcomes

Business Requirement(KPI) - Correctly predict high traffic recipes 80% of the time

Model	Precision at threshold 50%(default)	Precision at threshold 80%(KPI)
Logistic regression	83%	96%
Random Forest Classifier	82%	96%

Outcomes

Feature importance in the Random Forest Classifier Model



These are the major indicators as to whether a Recipe would lead to high traffic.

- Protein
- Beverages Recipe
- Sugar
- Calories

Recommendation

1. I suggest A/B testing the model's recommendations in a real-world setting to further evaluate its effectiveness before full implementation. This can help fine-tune the model and adapt to the actual user behavior patterns.
2. Continuously improve the model by collecting more data, features engineering and fine tuning parameter. This would help us get over the limitation of small dataset, the more data and features we have the better the models we can build.
3. A higher threshold will increase our precision but will slightly reduce our accuracy score. The threshold of 80% should be monitored and review periodically.

THANK YOU