DataCamp - December 2024

PREDICTING RECIPE POPULARITY

Leveraging Machine Learning to Optimize Recipe Selection by Classifying High and Low Site Traffic to Drive Engagement and Subscriptions

Presented by **Rolando John R. Aca-ac**

BUSINESS GOALS

- A solution that leverages **machine learning** to enhance decision-making for **recipe selection**.
- **Predict** which recipes will generate **high traffic** to optimize homepage selections.
- **Correctly** classify high-traffic recipes **at least 80% of the time**.
- Minimize the chance of showing **low-traffic recipes** to improve user engagement.
- Increase overall **site traffic and subscriptions**, boosting revenue for **Tasty Bytes**.

TASTY BYTES

DATA

A preview of the of the dataset before preprocessing and validation.

	recipe	calories	carbohydrate	sugar	protein	category	servings	high_traffic
0	1	NaN	NaN	NaN	NaN	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	NaN
3	4	97.03	30.56	38.63	0.02	Beverages	4	High
4	5	27.05	1.85	0.80	0.53	Beverages	4	NaN

DATA

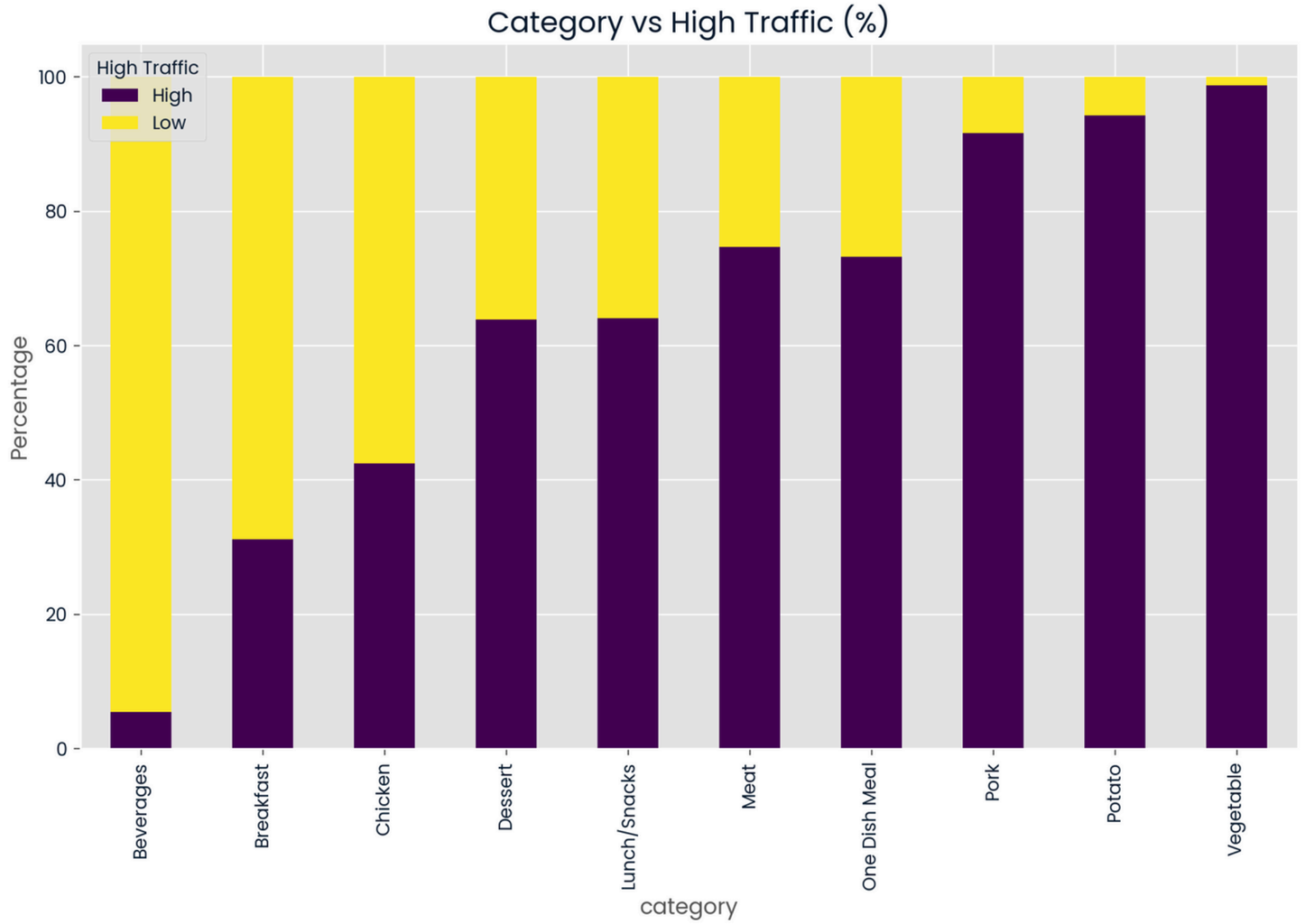
A preview of the dataset after preprocessing and validation.

	recipe	calories	carbohydrate	sugar	protein	category	servings	high_traffic
0	1	288.55	21.48	4.55	10.80	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	Low
3	4	97.03	30.56	38.63	0.02	Beverages	4	High
4	5	27.05	1.85	0.80	0.53	Beverages	4	Low

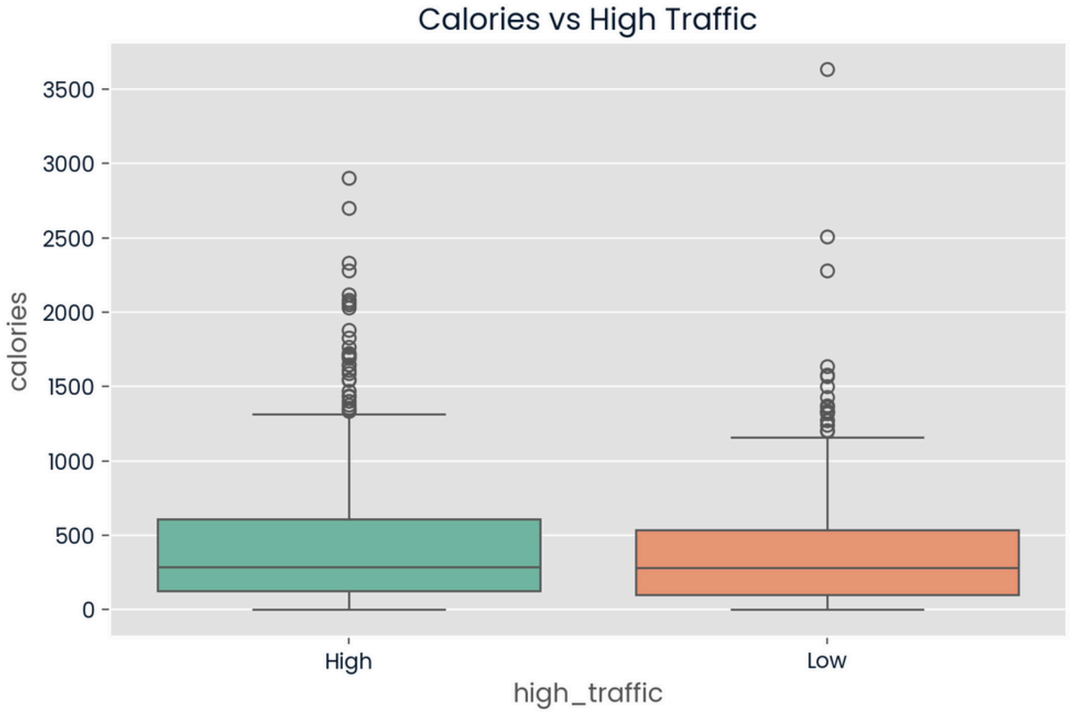
KEY FINDINGS

- Categories like **Vegetable**, **Potato**, and **Pork** have the **highest** likelihood of generating high traffic.

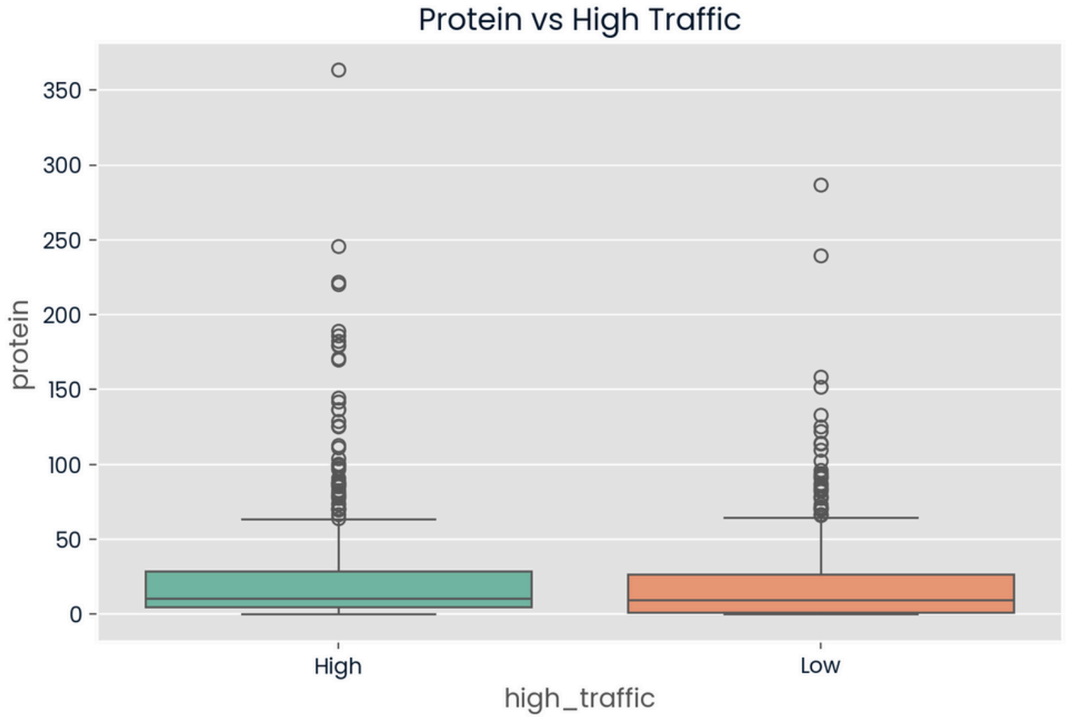
- Conversely, categories like **Beverages**, **Breakfast**, and **Chicken** are the **least** likely to generate high traffic.



KEY FINDINGS



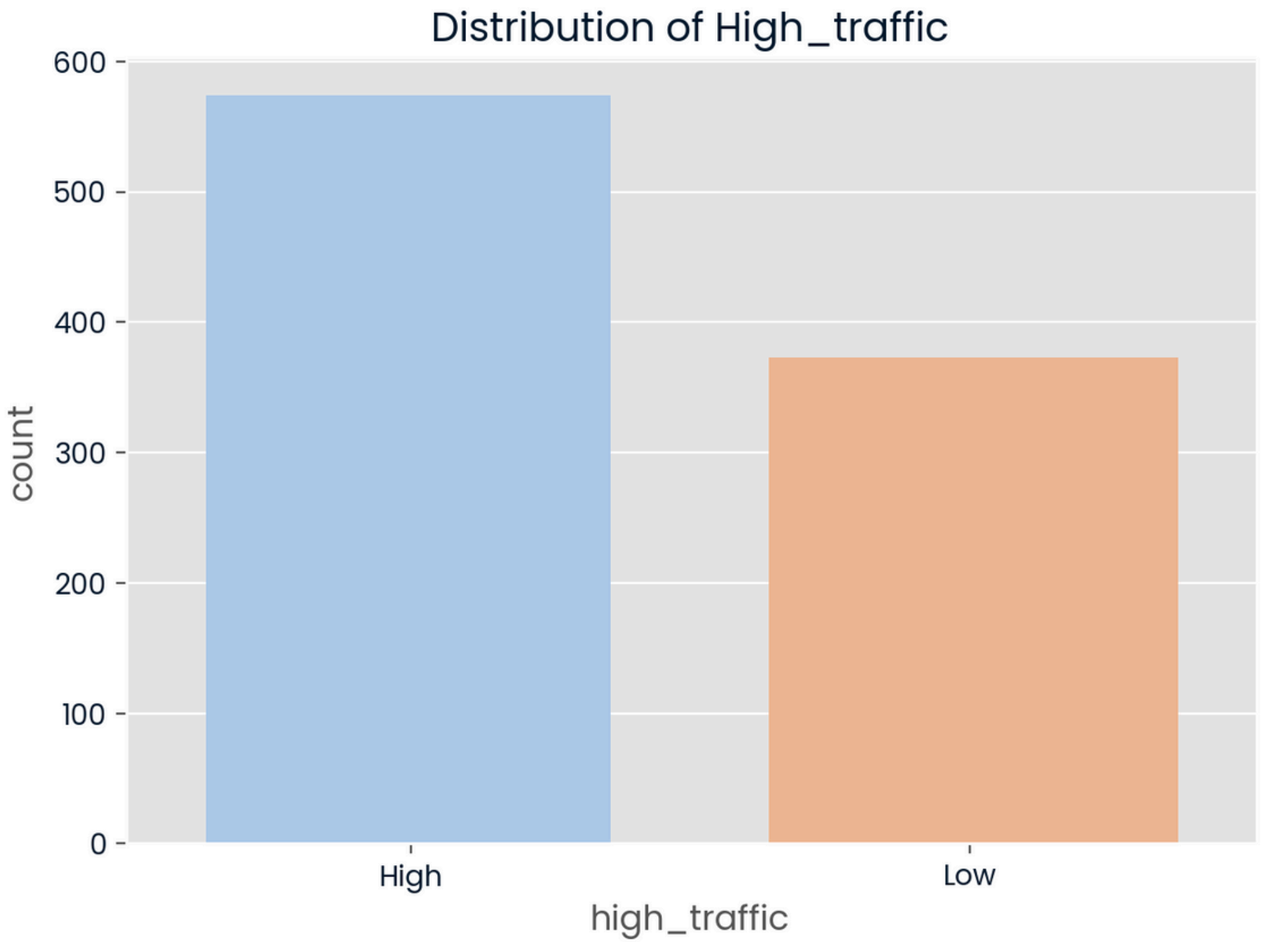
- Calorie content shows a similar median for both high and low-traffic recipes, with high-traffic recipes having slightly broader range



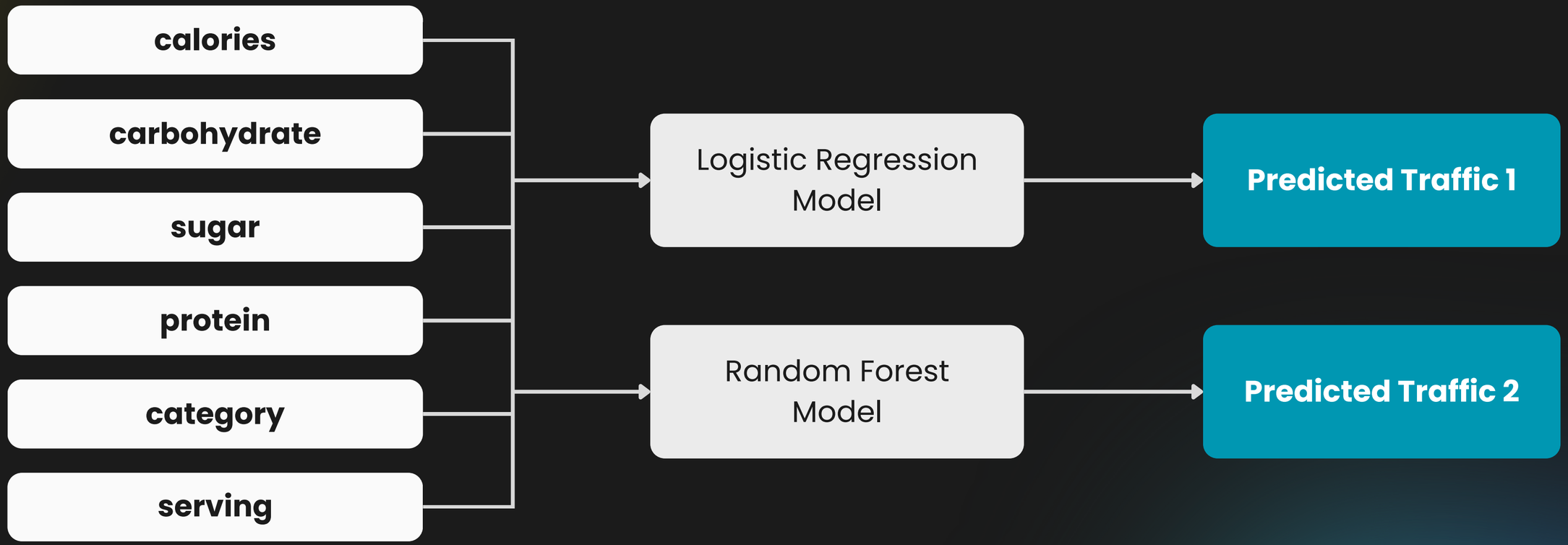
- Protein levels also exhibit comparable distributions, suggesting no strong correlation with recipe traffic.

KEY FINDINGS

- The dataset contains slightly more recipes labeled as High Traffic compared to Low Traffic, indicating a **minor class imbalance**.



Two models - Logistic Regression and Random Forest Model



OUTCOMES

Key Evaluation Metrics for Model Performance

Metrics	Details	Range
Precision	Ensures the correctness of High Traffic predictions	0 - 1
Recall	Measures the ability to identify all High Traffic recipes	0 - 1
F1-Score	Balances Precision and Recall for imbalanced datasets	0 - 1
AUC-ROC	Evaluates the model's ability to distinguish between High and Low traffic	0 - 1
Accuracy	Measures the overall correctness of the model predictions	0 - 1

OUTCOMES

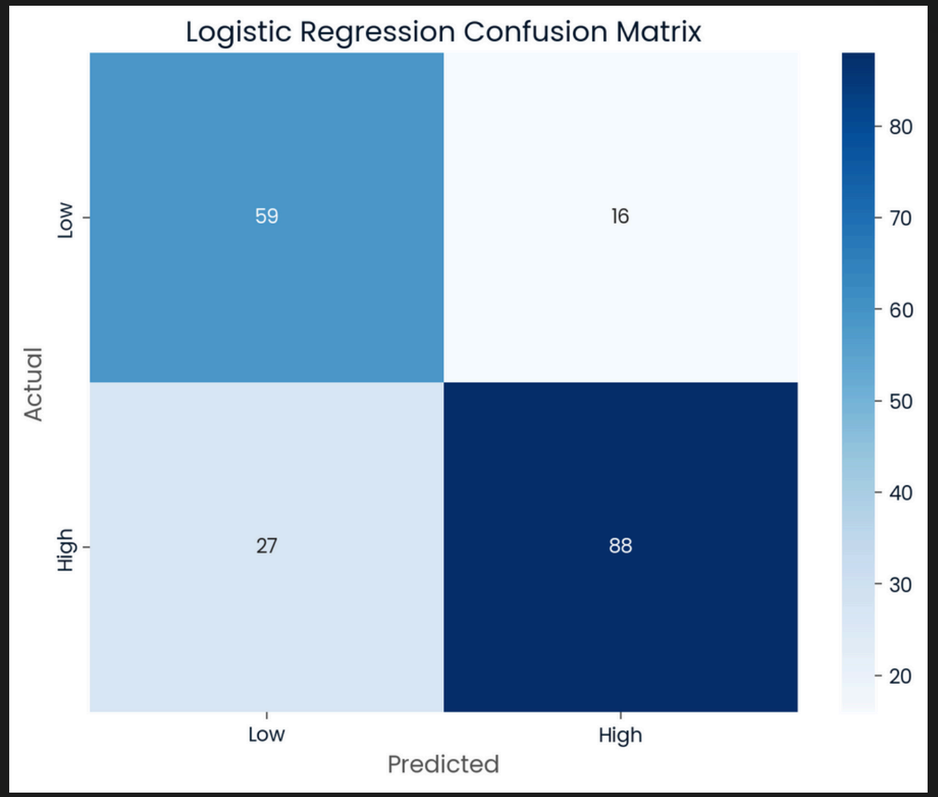
Model Evaluation Results

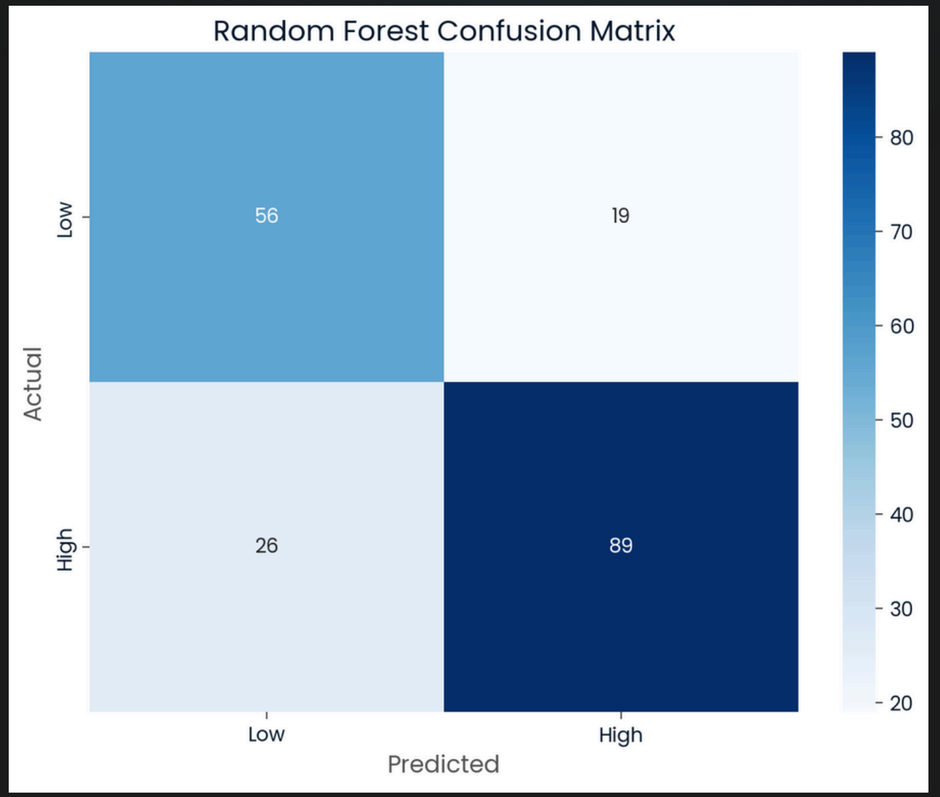
- Logistic Regression demonstrates slightly better performance overall.

Metrics	**Logistic Regression**	Random Forest
Precision	**0.85**	0.82
Recall	0.77	0.77
F1-Score	0.80	0.80
AUC-ROC	**0.86**	0.84
Accuracy	**0.77**	0.76

OUTCOMES

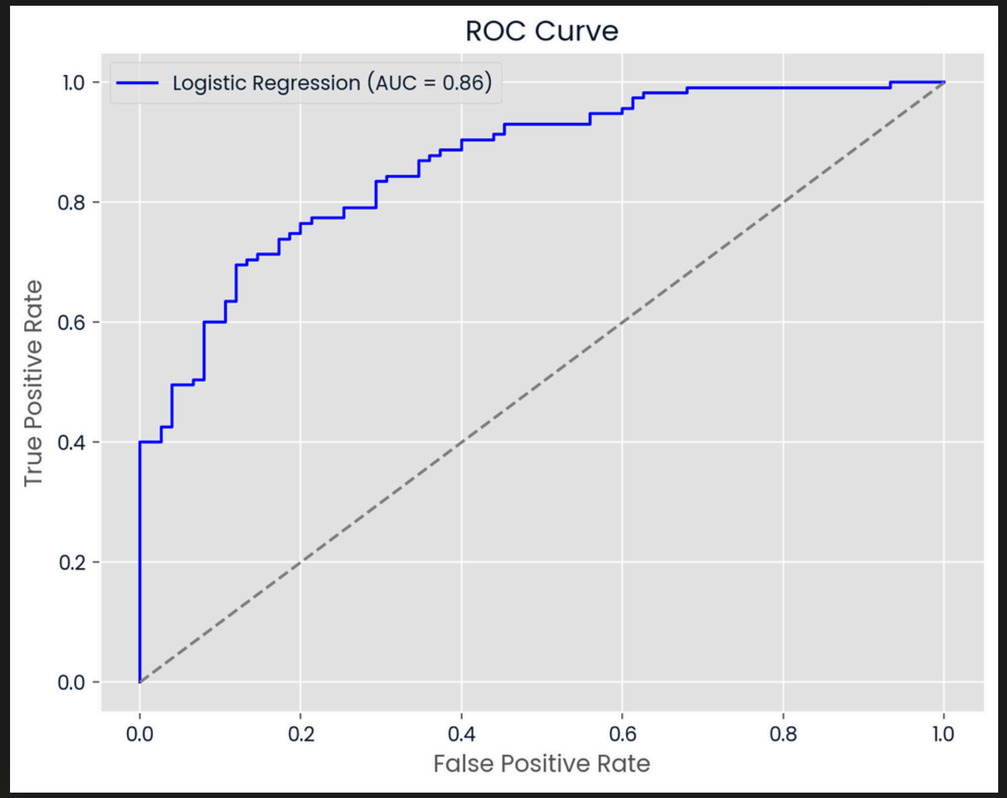
Logistic Regression shows fewer misclassifications for High and Low traffic compared to Random Forest, indicating better precision and recall.

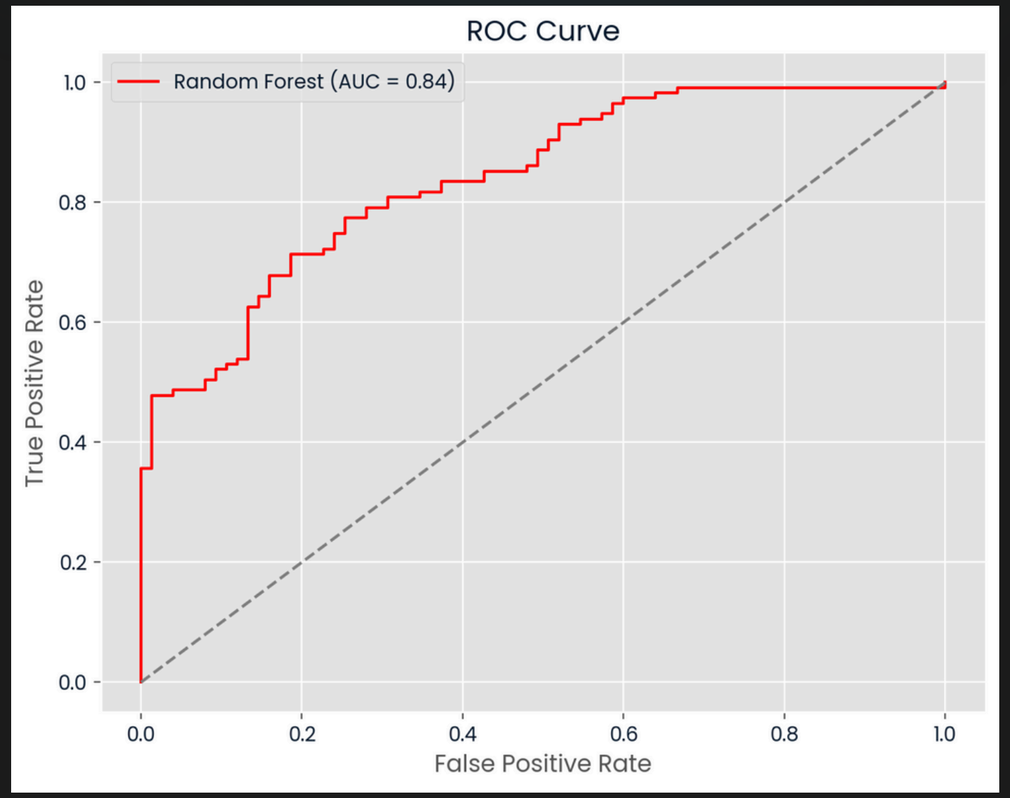




OUTCOMES

Logistic Regression has slightly higher AUC-ROC (0.86) compared to Random Forest (0.84), indicating a better balance between sensitivity and specificity.



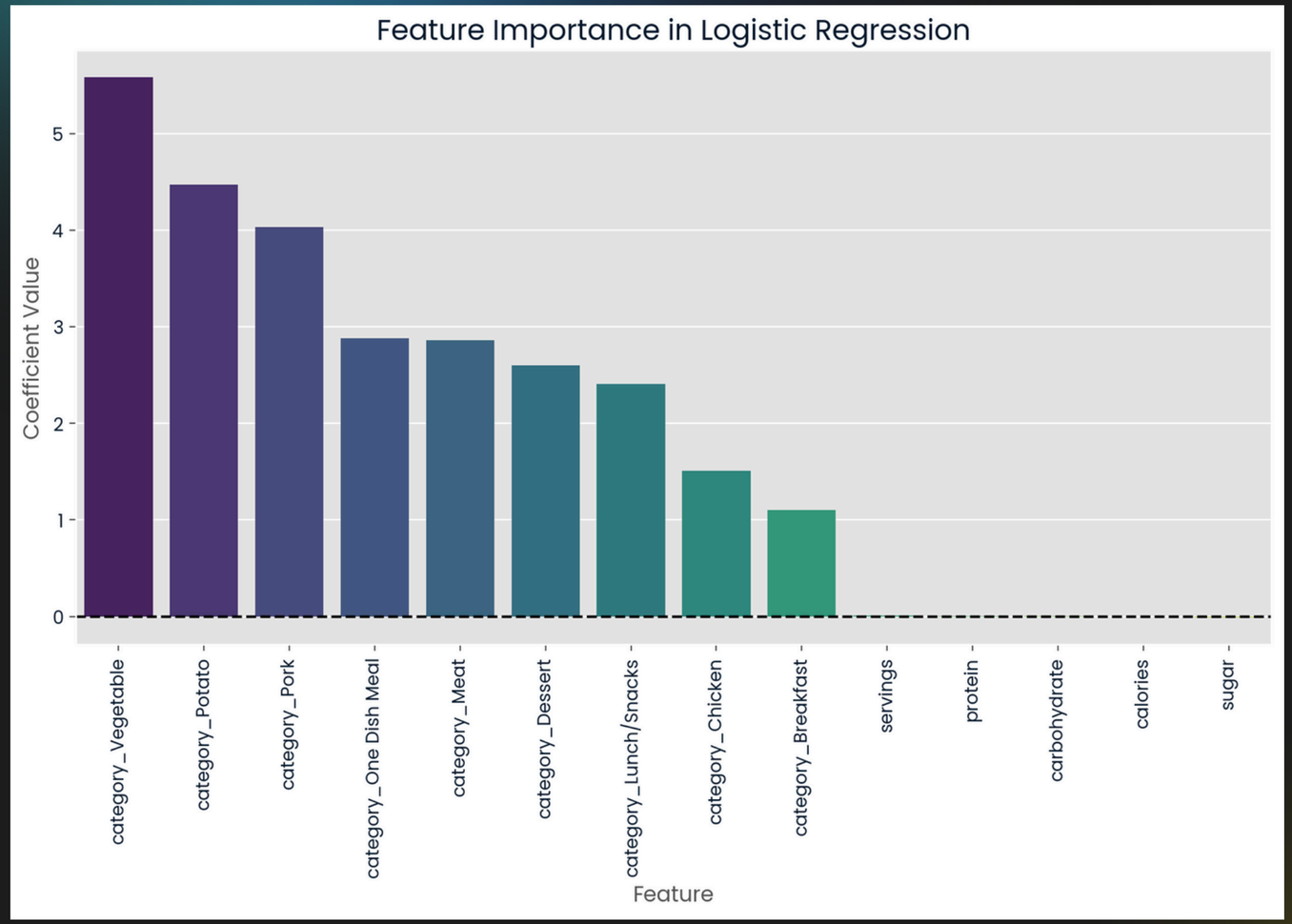


OUTCOMES

- The **top** contributors to predicting high traffic are the **Vegetable**, **Potato**, and **Pork** categories.

- Nutritional features like **Calories**, **Carbohydrates**, and **Protein** play a **lesser** role in the logistic regression model's predictions.

OUTCOMES



RECOMMENDATION

- **Test the Logistic Regression Model** on new recipes to validate performance.
- **Identify and fix errors** to improve accuracy after testing.
- **Deploy the model** using API integration or platform implementation.
- **Focus on high-impact features** like Vegetable and Potato for recommendations.
- **Collect additional data** on underrepresented recipe categories.
- **Monitor model performance** and retrain with updated data regularly.

THANK YOU SO MUCH!

Presented by Rolando John R. Aca-ac