

# RECIPE DISCOVERY

Predicting High Traffic Recipes

By: Aisha Evering

# 02

## Problem

The wonderful company, Recipe Discovery, currently manually chooses recipes from their selection to display on their home page. They have noticed that traffic to the website increases by as much as **40%** if a popular recipe is chosen. More traffic means more subscriptions, making this process crucial for the company. They would like a more informed and reliable way of selecting popular recipes.



A photograph showing two white bowls filled with omelettes. The omelettes are topped with various ingredients including diced tomatoes, sliced green onions, mushrooms, and guacamole. The bowls are placed on a light-colored wooden surface.

03

## BUSINESS REQUIREMENTS

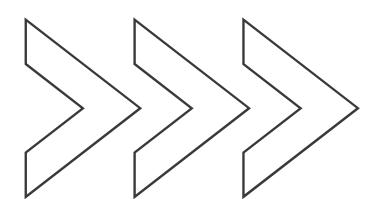
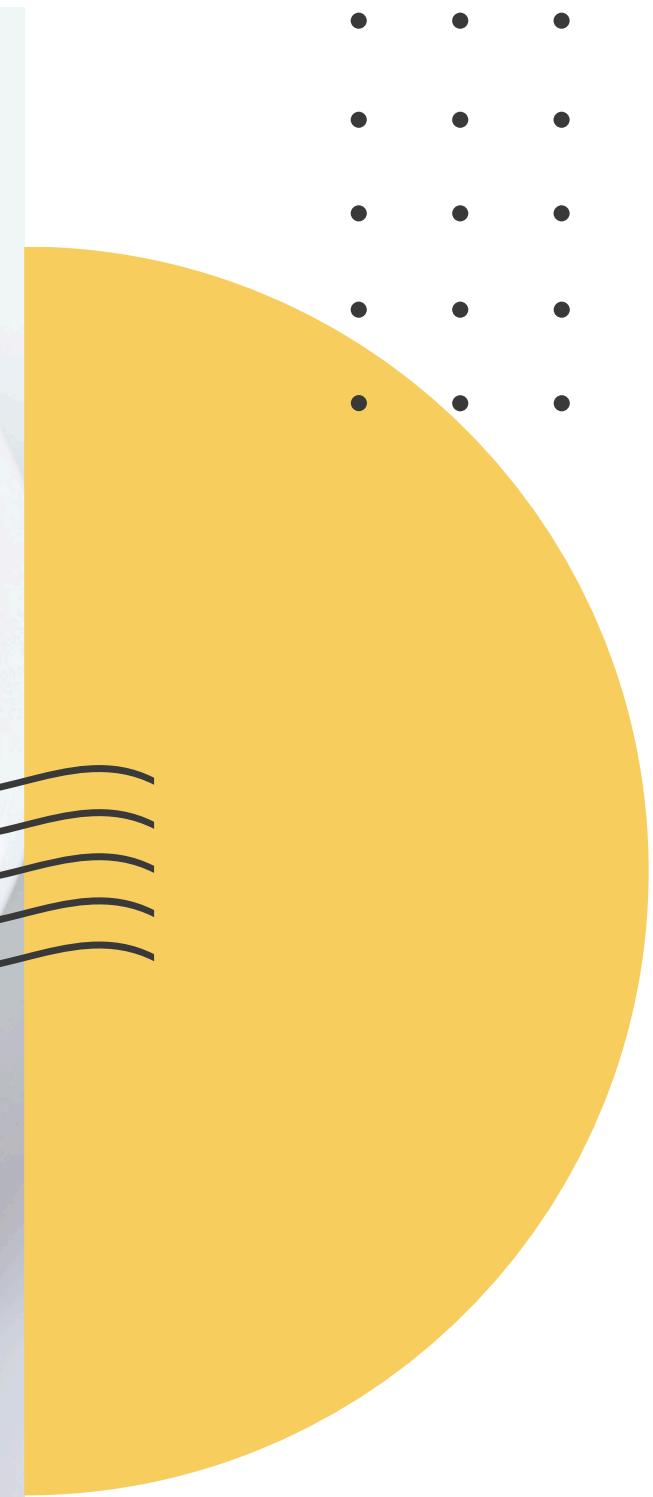
The wonderful company, Recipe Discovery, currently manually chooses recipes from their selection to display on their home page. They have noticed that traffic to the website increases by as much as 40% if a popular recipe is chosen. More traffic means more subscriptions, making this process crucial for the company. They would like a more informed and reliable way of selecting popular recipes.

- Predict which recipes will lead to high traffic?
- Correctly predict high traffic recipes **80%** of the time?
- Minimize the chance of showing unpopular recipes.

# 04

## DATA

- A list of recipes with features: calories, carbohydrates, amount of sugar, amount of protein, the category they fall under, and number of servings.
- The data also includes which recipes lead to high traffic. This information will be used as the target variable.
- The dataset is imbalanced with respect to the target variable, comprising **60.6%** High Traffic recipes and **39.4%** Low Traffic recipes.



# 05

## Analysis

- Recipes with 4 servings are more likely to lead to high traffic
- Most recipes that are likely to drive **high traffic** are in categories
  - Vegetable
  - Potato
  - Pork
  - One Dish Meal
  - Meat
- Most recipes that are likely to drive **low traffic** are in categories
  - Beverages
  - Breakfast





# 06

## BINARY CLASSIFICATION

Considering that the target variable determines whether a recipe leads to **High** or **Low** traffic, the models I have selected for experimentation are well-suited for predicting binary classification data.

- • • • •
- • • • •
- • • • •
- • • • •
- • • • •

➤➤➤ Baseline model: Logistic Regression

➤➤➤ Comparison model: Random Forest Classifier

# 07

## KEY METRICS

**Recall:** percentage of actual high traffic recipes that the model correctly identified

**Requirement:** correctly predict high traffic recipes **80%** of the time

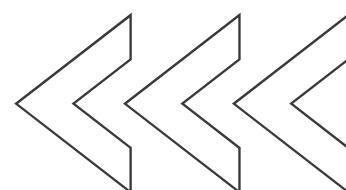
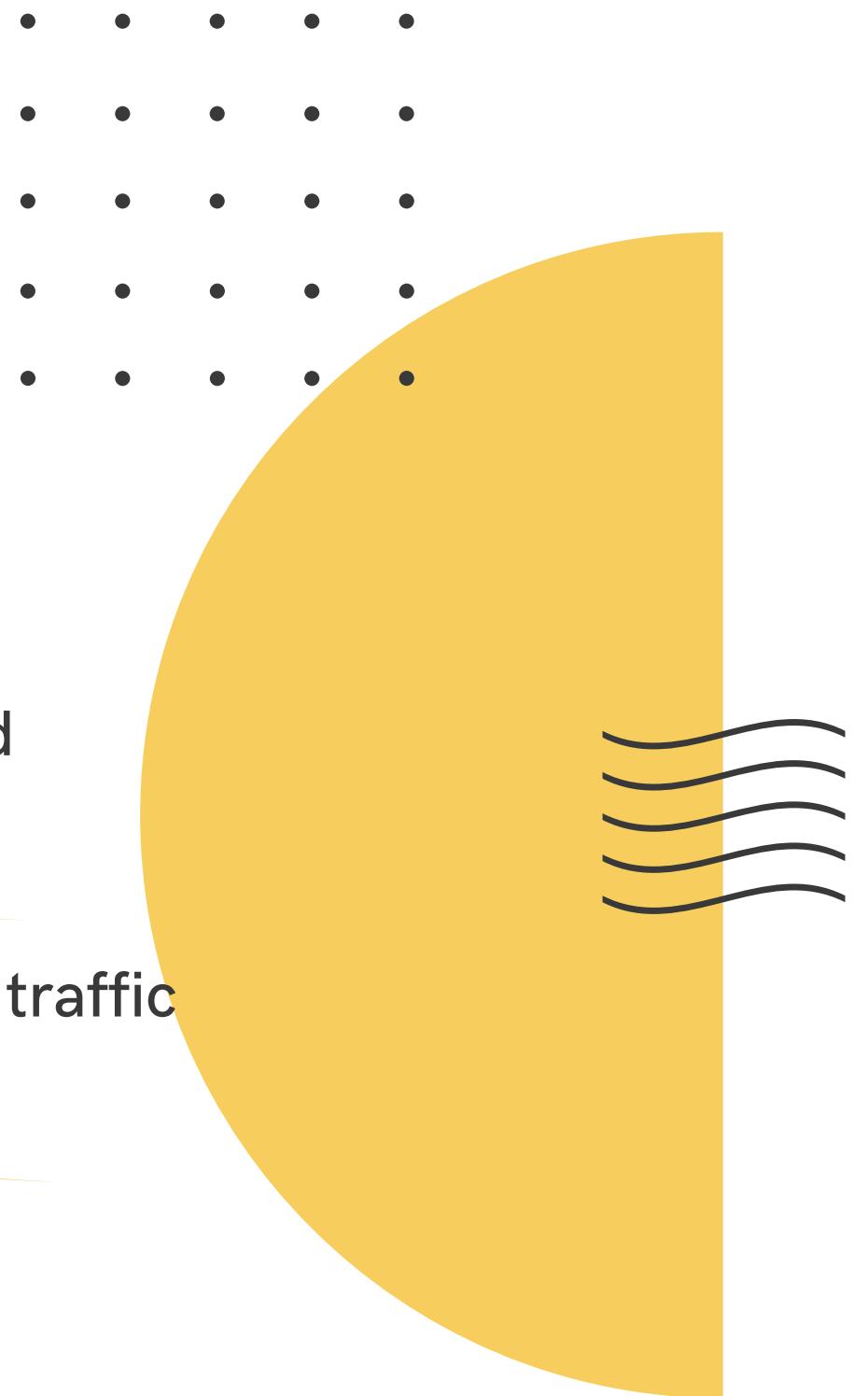
---

**Precision:** percentage of recipes predicted to be high traffic that are actually high traffic

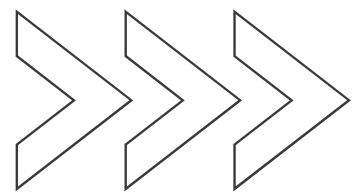
**Requirement:** minimizing the possibility of choosing recipes that lead to low traffic

---

**F1 Score:** balance between Recall and Precision



# Logistic Regression

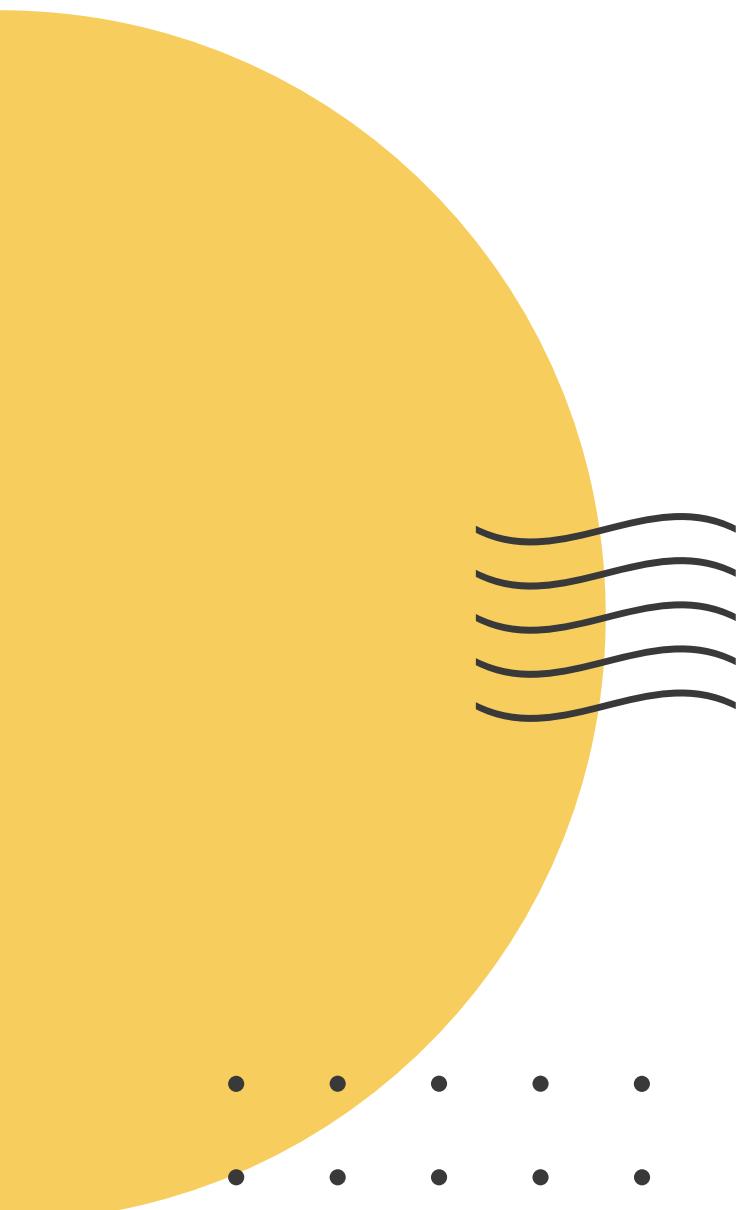
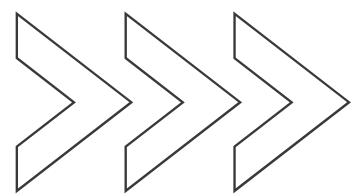


77%

- Recall for High Traffic: 77%
- Precision for High Traffic: 85%
- f1-score for High Traffic: 81%

- ✖ Does not meet Business Requirement

# Random Forest Classifier



09

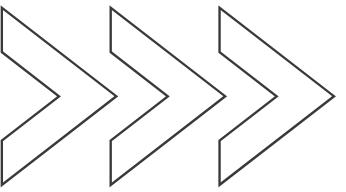


# 81%

- Recall for High Traffic: 81%
- Precision for High Traffic: 82%
- f1-score for High Traffic: 81%



Meets Business Requirement



# 10 SUMMARY

- The best performing model is the Random Forest Classifier.
- Recipes likely to lead to high traffic are 4-serving recipes in categories
  - Vegetable
  - Potato
  - Pork
  - One Dish Meal
  - Meat

# BUSINESS SUGGESTIONS

- Add more 4-serving Vegetable, Potato, Pork, One Dish Meal, and Meat recipes.
- The current dataset is small and imbalanced. Gathering more data will likely improve the model's performance and provide more accurate predictions.
- Deploy the model in the cloud or a platform that supports an API to initiate the inference pipeline and automatically deliver popular recipes to the Food Discovery web application.
- Periodically re-evaluate the model using fresh data to ensure it continues to perform well. This will help maintain the model's accuracy and relevance.
- Implement a system for gathering feedback from stakeholders and users to continuously improve the model and its predictions. This feedback will be invaluable for making iterative improvements to the model.

12

# Thank You!

Aisha Evering

---

Portfolio: <https://aishaeportfolio.com/>

LinkedIn: <https://www.linkedin.com/in/aisha-evering-91580a3/>

Email: [shovon3000g@gmail.com](mailto:shovon3000g@gmail.com)

