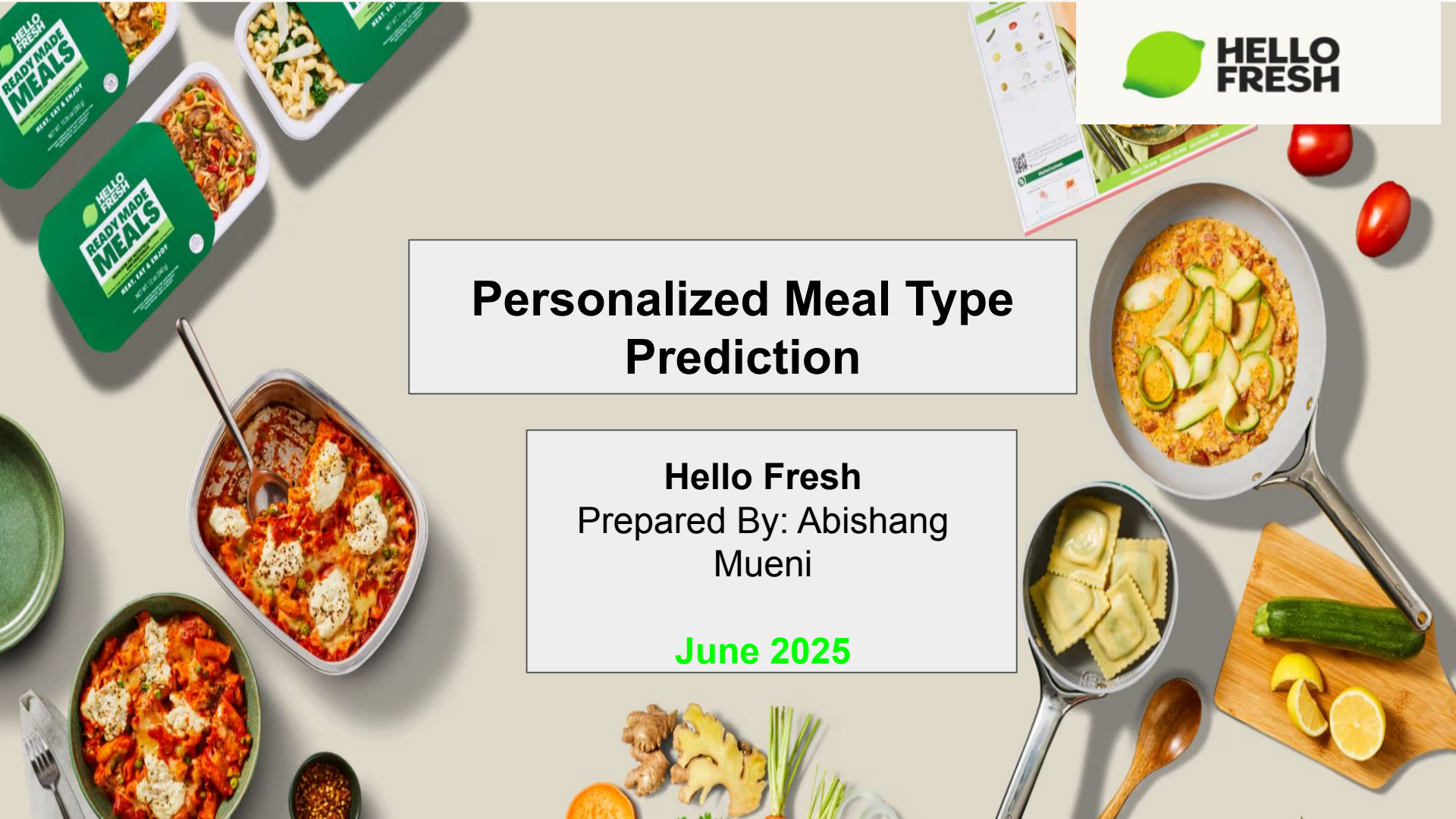

Personalized Meal Type Prediction

Hello Fresh

Prepared By: Abishang Mueni

June 2025

Project Overview

HelloFresh is a global meal kit company that helps people cook fresh, healthy meals at home.

Their goal is to make home cooking easier and more enjoyable with ready-to-use ingredients and step-by-step recipes.

This project aims to develop a classification model that predicts a user’s preferred meal type (Breakfast, Dessert, Dinner, Lunch, or Snack) and increase personalization accuracy using user reviews and recipes data.

Data & Methodology

Data Source: Data from Food.com (via Kaggle)

Contains 2 dataset;

1. Recipes dataset - Over 500,000 records,28 columns
2. Reviews dataset - Over 250,00 records, 8 columns

Key data: Nutritional info, cook time, review text, ratings.

Data Preprocessing

- Preserved Raw Data
- Handled Missing Values
- Standardized Time Format
- Formatted Dates
- Identified & Removed Outliers
- Merged Datasets
- Saved Final Dataset
- Created “Meal Type” labels based on recipe categories
  Final meal types: Breakfast, Lunch, Dinner, Snack, Dessert

Key Focus Areas

Meal Type Trends

Insights on breakfast, lunch, dinner, snacks, desserts

Label Cleaning

Consolidated 40+ meal types into 5 main categories

breakfast, lunch, dinner, snacks, desserts

Model Evaluation

Tested Logistic Regression and Decision Tree classifiers

Modelling Approach

First Model – Logistic Regression

Initial Model Results

- Accuracy: 46%
- Good at predicting Dinner
- Struggled with smaller categories like Snack and Breakfast
- Conclusion: Not enough complexity or balance

```
Logistic Regression Accuracy Score: 0.4641902328897338

Confusion Matrix:
 [[ 1863  1970  5645     3   260]
 [  650  3348  2933     0   487]
 [ 1301  1322 23012   885   269]
 [ 1095   899 16343   773   206]
 [   67   300  1440     0  2257]]

Classification Report:
              precision    recall  f1-score   support

           0       0.37      0.19      0.25      9741
           1       0.43      0.45      0.44      7418
           2       0.47      0.86      0.60     26789
           3       0.47      0.04      0.07     19316
           4       0.65      0.56      0.60      4064

    accuracy                           0.46     67328
   macro avg       0.48      0.42      0.39     67328
weighted avg       0.46      0.46      0.38     67328
```

Fixing the Imbalance

Handling Uneven Meal Type Distribution

- Used techniques to balance classes:
  - Class weights
  - SMOTE (synthetic sample creation)
- Results: Fairer model but accuracy dropped to 38%

```
Balanced Logistic Regression Accuracy: 0.38414923954372626
Confusion Matrix:
 [[3793 3022  712  624 1590]
 [ 894 4404  197   37 1886]
 [3805 4458 5524 9806 3196]
 [2860 2504 2851 8987 2114]
 [ 352  338  156   62 3156]]
Classification Report:
              precision    recall  f1-score   support

           0       0.32      0.39      0.35      9741
           1       0.30      0.59      0.40      7418
           2       0.59      0.21      0.30     26789
           3       0.46      0.47      0.46     19316
           4       0.26      0.78      0.39      4064

    accuracy                           0.38     67328
   macro avg       0.39      0.49      0.38     67328
weighted avg       0.46      0.38      0.37     67328
```

```
Logistic Regression with SMOTE Accuracy: 0.38496613593155893
Confusion Matrix:
 [[3812 3004  716  626 1583]
 [ 886 4425  197   37 1873]
 [3803 4452 5542 9811 3181]
 [2882 2508 2860 8984 2082]
 [ 350  336  160   62 3156]]
Classification Report:
              precision    recall  f1-score   support

           0       0.32      0.39      0.36      9741
           1       0.30      0.60      0.40      7418
           2       0.58      0.21      0.31     26789
           3       0.46      0.47      0.46     19316
           4       0.27      0.78      0.40      4064

    accuracy                           0.38     67328
   macro avg       0.39      0.49      0.38     67328
weighted avg       0.46      0.38      0.37     67328
```

Switching to Decision Trees

A Better Model: Decision Tree

- Accuracy improved to 61%
- Better at identifying underrepresented meal types
- Still room for improvement

```
Decision Tree Accuracy Score: 0.6091819153992395

Confusion Matrix:
 [[ 4843   898  3237   667    96]
 [  636  3988  2190   343   261]
 [  912  1042 21867  2843   125]
 [ 1368   754  9472  7590   132]
 [  168   317   616   236  2727]]

Classification Report:
              precision    recall  f1-score   support

           0       0.61      0.50      0.55      9741
           1       0.57      0.54      0.55      7418
           2       0.58      0.82      0.68     26789
           3       0.65      0.39      0.49     19316
           4       0.82      0.67      0.74      4064

    accuracy                           0.61     67328
   macro avg       0.65      0.58      0.60     67328
weighted avg       0.62      0.61      0.60     67328
```

Tuning the Tree

Optimized Model: Tuned Decision Tree

- Accuracy jumped to **94%**
- Great precision and recall across all mealtypes
- Very few misclassifications
- Reliable predictions for all users

```
Accuracy: 0.9396833412547528
Confusion Matrix:
 [[ 9160    96   254   181    50]
 [  135  6822   230   166    65]
 [  242   231 25346   864   106]
 [  197   132   858 18044    85]
 [   20    40    62    47  3895]]
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.94      0.94      9741
           1       0.93      0.92      0.93      7418
           2       0.95      0.95      0.95     26789
           3       0.93      0.93      0.93     19316
           4       0.93      0.96      0.94      4064

    accuracy                           0.94     67328
   macro avg       0.94      0.94      0.94     67328
weighted avg       0.94      0.94      0.94     67328
```

Key Takeaways & Recommendations

Key Takeaways

- Cook time and nutritional values like(Calories,Fat Content,Protein Content) helped predict meal type
- Advanced models (Decision Tree) outperformed the simple one(Logistic Regression)
- Decision Tree Model works well even with uneven data

Recommendations

- Include more Snacks and Desserts in weekly menus to meet diverse customer preferences.
- Continue focusing on Lunch and Dinner, as they remain top choices among users.
- Shift from older models like Logistic Regression to tree-based models for better handling of complex, imbalanced data.
- Use predictive models to recommend meal types customers are most likely to enjoy.
- Leverage user history for tailored recommendations to boost satisfaction and retention.

Conclusion & Next Steps

Next Steps

- Explore user-level preferences for even deeper personalization.
- Test model in a pilot feature on the HelloFresh platform.
- Gather feedback and continuously improve recommendations.

Thank You

- **Questions?**
- **Contact: Abishang Mueni**

[Email + Linkedin]