The purpose of this deck is to fulfill the requirement of DSI 39, Project 4.

Role: Health Promotion Board (HPB) Data Science Team

Audience: HPB's Senior Management

Problem Statement: Singaporeans spend a significant % of their diet on snacks. However, there is no guidelines that apply specifically to snacks.

Done by: The Colony (Jackie Seah, Li Cheng, Mei Qi, Yvonne Lim)

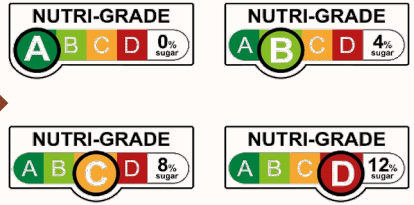
Snack-O-Meter: a tool to inform consumers on consumption of biscuits

HPB Senior Management Meeting (13 Oct 2023)

The National Population Health Survey highlighted need to focus on healthy diets and lifestyles

Singaporeans' diet is showing an **increasing consumption of fats, sodium, and sugar**. While there are ongoing efforts to promote healthy consumption of beverages and food dishes, there is a **gap in covering the nutritional values in snacks**.

Beverages



- Focuses on **Sugar** and **Saturated Fat**
- Nutri-Grade shifted consumption patterns and sugar intake has reduced

Food Dishes

Singapore is studying possible regulatory measures to reduce sodium content in food

- Focuses on **Sodium**

Snack

?

- Focuses on **Sugar**, **Sodium** and **Fat**

Biscuit is used as the initial proof-of-concept due to its popularity.

Leading choice of snacks

57%

Survey respondents purchased biscuits over other snacks

Market Size

5%

Expected CAGR in the next 5 years

https://www.statista.com/forecasts/1231123/snack-food-market-volume-by-segment-worldwide
https://www.statista.com/statistics/1341575/singapore-most-bought-sweet-snacks-in-the-past-week/

Objective

To build a user-friendly tool that can classify if a biscuit is healthy or not, helping consumers make healthier choices

Workflow

Web-scraping

- Scraping biscuits and their nutritional data from Fairprice website.

Data Cleaning and Labelling

- Using regular expressions to extract information from the unstructured data.
- Labelling “healthy” vs “unhealthy”

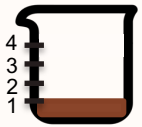
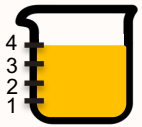
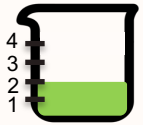
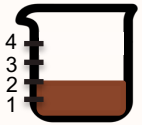
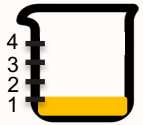
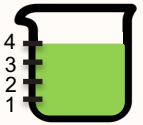
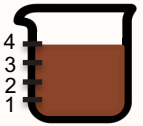
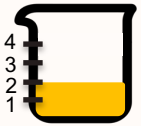
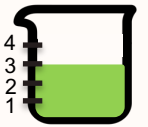
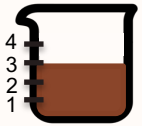
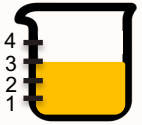
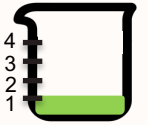
Modelling

- Fitting and evaluating multiple classification models to classify “healthy” vs “unhealthy” based on its nutritional values.

Web Application Design

- Developing a user-friendly web application for consumers and businesses to use

Variations in nutrients within each type and among the types call for need to consider a combination of the nutrient amount

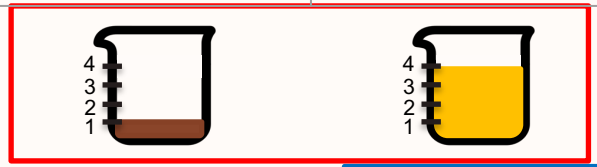
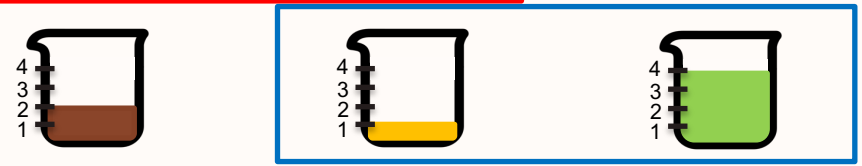
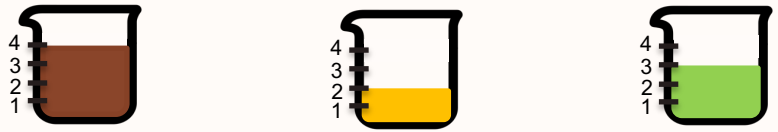
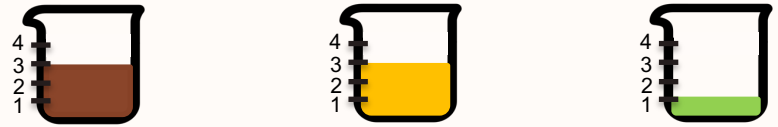
	Cream-filled biscuit has the highest amount of fats	Chocolate cookies has the highest amount of sugar	Cracker has the highest amount of sodium
	Fats	Sugar	Sodium
Chocolate cookies			
Cracker			
Cream filled biscuit			
Wafer			

1 refers to lowest amount of the component, and 4 refers to the highest amount of the component, in the biscuit type, where amount is based on the median value of each biscuit type

Variations in nutrients within each type and among the types call for need to consider a combination of the nutrient amount

Lowest amount of total fats, but highest amount of sugar

Lowest amount of sugar, but highest amount of sodium

	Cream-filled biscuit has the highest amount of fats	Chocolate cookies has the highest amount of sugar	Cracker has the highest amount of sodium
	Fats	Sugar	Sodium
Chocolate cookies			
Cracker			
Cream filled biscuit			
Wafer			

1 refers to lowest amount of the component, and 4 refers to the highest amount of the component, in the biscuit type, where amount is based on the median value of each biscuit type

Labelling biscuits based on recommended values of 3 nutrients of concern

Sodium

0.28g/100g

Sugar

21g/100g

Fat

25g/100g

Sugar and Fat based on Healthier Choice Symbol Nutrient Guidelines (as of 2020)

A classification machine learning model is trained, and used to classify if a biscuit product is healthy or unhealthy

Biscuit	Fat	Sodium	Sugar	Predictor
A	X	X	X	Healthy
B	X	X	X	Unhealthy
C	X	X	X	Healthy

Data preparation
- Split data
- Upsample imbalanced train data

Training models

Model Evaluation
- Accuracy
- Under/Over-Fitting
- Run Time

Tuning

Best model

Deploy best model

Healthy: Eat in Moderation
OR
Unhealthy: Eat Less Of

Models Used

Logistic Regression
A statistical model that models the probability of an event taking place by fitting a sigmoid function to the data.

Decision Tree
A machine learning algorithm that constructs a tree-like structure which represents a series of decisions and their possible consequences.

Decision Tree (with bagging)
Multiple decision trees (dataset is randomized) are used to make a single prediction.

Random Forest
Multiple decision trees (both dataset and features are randomized) are used to make a single prediction.

Model Evaluation

Accuracy
The total number of accurate predictions out of all predictions. The higher the better.

Cross Validation Accuracy
The accuracy of the model based on different subsets of the dataset. It should be close (within 10%) to the train and test accuracy. This suggests that the model is generalizable and will show similar performance on unseen data.

F1 Score
Harmonic mean of the precision and recall scores of a model. The higher the better.

Run Time
Time taken to train the model. The lower the better.

Decision Tree is the best performing model which takes the shortest time to run

Models	Train accuracy	Test accuracy	Cross Validation Train accuracy	Train F1 score	Time Taken to run (in secs)
Logistic Regression	0.77	0.86	0.66	0.77	0.06
Random Forest	1.00	1.00	0.92	1.00	1.53
Decision Tree	1.00	1.00	0.95	1.00	0.03
Decision Tree with Bagging	1.00	1.00	0.92	1.00	0.22

DEMO

https://snack-o-meter.streamlit.app/

Recommendations

Increase Awareness to Public

Through marketing campaigns (offline and online campaigns)

Expand the model to include other snack types

Other snack types like nuts, chips.

Integrate tool into HPB's existing Health 365 app

Intergration is beneficial as it makes Health 365 as "one stop app"

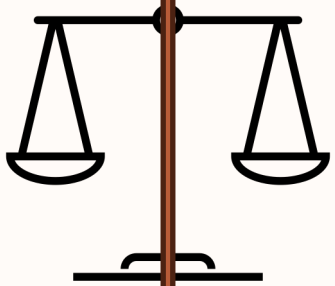
Benefit of tool overweighs cost by 25x

Cost
(estimated per year)

$500,000

Marketing Campaign:
S$400,000

App Development & Maintenance:
S$100,000



Benefit
(estimated per year)

$12,840,000

Healthcare cost of metabolic risk:
S$642,000,000

Reduction of daily sodium, sugar, fats consumption:
2.0%

- In FY2022, HPB spent $400k on Programme, supplies & marketing.

- From FY2017 to FY2019, the median sugar level of beverages decreased from 8.5 to 6.3 grams per 100 ml (25%) due to the Nutri-Grade campaign.
- Assuming one consumed a serving of Hello Panda Chocolate a day, that would constitute 8% of the daily average fats, sodium and sugar intake overall.
- All else constant, if this label has the same success as the Nutri-Grade campaign, it can reduce daily fats, sodium and sugar intake by 2%.

Summary

Problem statement:

- Singaporeans' diet is showing a worrying trend of increasing consumption of fats, sodium, and sugar. While there are programs in place to mitigate those, there is a gap in covering snacks, which is a significant part of Singaporeans' diet.

Deliverables:

- Build a classifier model with accuracy of > 0.9
- Deploy an easy-to-use self-help tool using the classifier model to provide quick information for individuals to improve their snacking habits

Thank You.