



City Research Online

City St George's, University of London

Citation: Alghamdi, S. (2022). Data-driven modelling of unhealthy snacking behaviour. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/28506/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Data-driven Modelling of Unhealthy Snacking Behaviour



Shaima Alghamdi

Department of Computer Science
City, University of London

A thesis submitted in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

City, University of London

July 2022

To my great parents, Dr. Mohammed Alghamdi and Mrs. Ibtesam Alhules, whose constant love, affection, and encouragement have led me to complete this project.

To my dearest husband, Dr. Wael Aljohani, for the love, support and sacrifice during the PhD journey,

To my lovely little heroes, Dimah and Faisal,

To my beloved brothers, sisters, and friends.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Shaima Alghamdi
July 2022

Acknowledgements

My deepest and sincere gratitude goes to my supervisors Dr. Tillman Weyde and Prof. Emmanuel Pothos, for offering me the opportunity to work on this project and for their continuous guidance and endless supply of ideas and comments throughout my PhD journey. My sincere thanks is also to Dr. Dympna O'Sullivan who was my second supervisor at the first stage of my PhD. I would like to express my many thanks to Dr. Katy Tapper who was always available to give assistance and advice on numerous occasions with kindness and great knowledge. I am grateful to Prof. Anne Roefs and colleagues at the faculty of Psychology and Neuroscience at Maastricht University for providing me the opportunity to work on their dataset collected on a similar topic.

This project would not have been possible without the scholarship awarded from King Abdulaziz University (KAU), providing all financial means to this project. Many thanks to the generous support and encouragement from the Department of Health Services and Hospitals Administration at KAU. My sincere thanks is also to my respected academic advisers in KAU, Prof. Omar Alsharqi and Prof. Abdulllah Al-Malaise for their full support and administrative guidance.

And finally, I would like to express my deepest thanks to my parents, husband, family members, and friends who endured this long process with me, always offering support and love.

Abstract

Unhealthy snacking is considered as a risk factor for many diseases including cancer, obesity, cardiovascular disease, and diabetes and many would like to reduce their snacking. However, changing behaviour is challenging and many attempts achieve limited success. Interventions providing suggestions or feedback, often called nudges, can help, especially if they are delivered at the right time, i.e. just before the potential snacking. Machine learning models can be used to predict behaviour and potentially drive the timing of the nudges. Accordingly, the main goals of this project are to a) develop an app to collect data on snacking, b) collect and analyse snacking data, and c) build predictive models to support the timing of interventions.

We developed an app, called ‘SnackTracker’ which enabled users to track their snacking behaviour. We used this app to run a longitudinal study, collecting unhealthy snacking data in the UK between Dec 2019 and April 2020. 184 participants completed the initial survey and 111 participants completed the app-based data collection over a period of approximately 4 weeks. The data about time and location of snacks forms our ‘UK dataset’. For comparison, we analyse another dataset collected with different methodology in the Netherlands (the ‘Dutch dataset’), covering general eating behaviour including unhealthy snacks.

The UK and Dutch datasets were analysed for trends and patterns. We observed demographic and other differences. In both datasets, we found that unhealthy snacking takes place most frequently in the afternoon and at home. The snacking frequencies were similar between weekends and weekdays. We found generally more unhealthy snacking in the UK dataset, and the daily and weekly number of unhealthy snacks tended to reduce during the experiment in both datasets.

Bearing in mind support for timed interventions, we built models to predict unhealthy snacks in different ways. We applied a number of different models, including linear and non-linear regressions, as well as fixed-context models, requiring fixed input size, and recurrent models. We evaluated the predictive models with standard classification and regression metrics. We also evaluated the trade-off between timing close to the event and the risk of intervening too late. We found that some of our models are usable for timing interventions under reasonable assumptions.

Based on the feedback from users and the observed reduction in snacking, even without interventions delivered by the app, the use of our tracking app can have positive effects. Given our modelling results, the implementation and evaluation of nudges timed by predictive models is now a practical next step for further development and research.

Table of contents

List of figures	xv
List of tables	xvii
1 Introduction	1
1.1 Research problem and motivation	1
1.2 Research questions and objectives	4
1.3 Research contributions	4
1.4 Thesis structure	5
2 Literature Review	7
2.1 Behaviour modelling and machine learning	7
2.1.1 Modelling human actions and interactions	9
2.1.2 Modelling human daily living activity	10
2.1.3 Modelling online purchase behaviour	11
2.1.4 Modelling eating behaviour	12
2.1.5 Machine learning and time series predictions	15
2.2 Behaviour change	16
2.3 Timed interventions	18
2.4 Ethical issues and considerations	20
2.5 Gap in the literature	20
3 Methodology	23
3.1 Survey and app design and data collection	23
3.2 Data pre-processing	24
3.3 Data analysis	26
3.4 Predictive models	26
3.5 Hypothetical interventions design and considerations	27

4	Snacking Behaviour Survey and Data Collection with the SnackTracker App	29
4.1	Ethics, privacy and security considerations	29
4.2	Survey design and execution	30
4.2.1	Questionnaire	32
4.3	App design and app-based data collection	35
4.3.1	The <u>Snack Tracker</u> app development and data flow	35
4.3.2	Technical description of the app design	35
4.3.3	The flow of the app-based part of the study	39
4.4	Discussion	43
5	Snacking Datasets Pre-processing and Descriptive Analysis	45
5.1	The Dutch dataset	45
5.2	Data cleaning	46
5.2.1	UK dataset	46
5.2.2	Dutch dataset	46
5.3	Feature engineering	47
5.4	Data transformations	48
5.5	Oversampling of imbalanced classes	49
5.6	Descriptive analysis	50
5.7	Discussion	58
6	Predictive Models of Unhealthy Snacking Behaviour	61
6.1	Predictive models	61
6.2	Metrics	62
6.3	Models validation	63
6.4	Classification	64
6.4.1	Prediction of unhealthy snacking occurrence in a time bin	67
6.4.2	Predict whether the next snack is healthy or unhealthy	72
6.4.3	Addressing class imbalance and loss asymmetry	74
6.4.4	Predicting the number of unhealthy snacks	77
6.5	Regression task	82
6.5.1	Predicting the time until the next unhealthy snack	82
6.6	Discussion	83
7	Application and Intervention Design Considerations	85
7.1	Distribution of residuals	86
7.1.1	Hypothetical Intervention for timing analysis	90

7.2	Discussion	94
8	Summary and Conclusion	97
8.1	Summary	97
8.2	Thesis contribution	100
8.3	Future work	101
8.3.1	Data collection	101
8.3.2	Data analysis and predictive models	102
8.3.3	Interventions	103
8.3.4	Other relevant human behaviours	104
	Bibliography	105
	Appendix A Ethics approval	121
	Appendix B Participants comments about the study	123
	Appendix C Snacks consumed in the Netherlands	125
	Appendix D UK Participants snacking trend throughout the study	129
	Appendix E Dutch participants snacking trend throughout the study	133

List of figures

4.1	Participant Flow in the two studies	31
4.2	data flow	36
4.3	App screenshots of splash screen1, login screen2, and home screen3	37
4.4	App screenshots of new snack screen2, time recording screen5, and location recording screen 6	38
4.5	App screenshots of recording save screen 7, and review recordings summary 8	39
4.6	App user flow	40
4.7	Three groups of participants in the app-based part of the study	42
4.8	App reminders	43
5.1	App-based Average Daily number of Unhealthy Snacks (ADUS) for different genders in the two datasets.	51
5.2	App-based Average Daily number of Unhealthy Snacks (ADUS) for different age groups in both datasets.	52
5.3	Daily unhealthy snacks for different Employment status in the two datasets.	53
5.4	App-based ADUS for different obesity levels in the two datasets.	54
5.5	App-based ADUS for education groups in the Dutch dataset.	55
6.1	Feature importance from XGBoost in the two main datasets, UK (self-collected) and Dutch while day time was sliced into 4 and 12 time bins. Ethnicity-4 is other ethnic groups, Location-1 is Home, Location-2 is work/study, Location-3 is other and Education-2 is post-secondary vocational education.	72
6.2	ROC curves for the prediction of unhealthy snack occurrence with 4 and 12 time bins in the two main datasets, UK and Dutch. Oversampling was applied on datasets with imbalanced classes in the target variable.	75

6.3	Precision/recall curves for whether a snack is healthy or unhealthy, with 4 and 12 time bins, in the UK and Dutch datasets, predicted with an LSTM model. Oversampling was applied in cases where there were imbalanced classes in the target variable.	76
6.4	ROC curves for the classification the type of snacking as healthy or unhealthy in time bins of 4 and 12, in the Dutch dataset. The predicted cases were from the LSTM model. No oversampling was required.	77
6.5	Precision recall curves for the classification of snacks as healthy or unhealthy in time bins of 4 and 12, in the Dutch dataset. The predicted cases were from the LSTM model. No oversampling was required.	77
7.1	Histogram of residuals for LR with 4 and 12 time bins, for the UK and Dutch datasets.	87
7.2	Histogram of residuals for RFreg when using 4 and 12 time bins in the UK and Dutch datasets.	88
7.3	Histogram of residuals for XGBreg with 4 and 12 time bins, for the UK and Dutch datasets.	88
7.4	Histogram of residuals for FFNN with 4 and 12 time bins, for the UK and Dutch datasets.	89
7.5	Histogram of residuals for LSTM with 4 and 12 time bins for the UK and Dutch datasets.	90
7.6	Case distribution of a 120-minute window for accepting the intervention in 4 and 12 time bins in the UK and Dutch datasets. The model used was RFreg.	92
7.7	Case distribution for a 60-minute window for accepting the intervention with 4 and 12 time bins in the UK and Dutch datasets. The model used was RFreg.	93
7.8	Case distribution with a 30 minutes window for accepting the intervention with 4 and 12 time bins, for the UK and Dutch datasets. The model used was RFreg.	94
A.1		122

List of tables

2.1	Description of the three persuasive systems designs strategies targeted in this thesis.	17
4.1	Demographics questions part of the survey	32
4.2	Questions regarding motivation for healthy eating in the survey	33
4.3	Three factor eating questionnaire (TFEQ) for restraint eating, emotional eating, external eating	34
4.4	Profile of participants in the sample. The total number of participants in the UK dataset was 100 (after excluding incomplete records).	41
5.1	Coding guide of Likert-scales related questions	46
5.2	Description of the new features created in the UK and Dutch datasets based on the collected features	47
5.3	Demographic information in the Dutch dataset. All frequencies in this table represent unhealthy snacking. Unhealthy snack events were extracted from generic eating information contained in the data. The total number of participants in the Dutch dataset was 196. Records with missing values were excluded from any processing.	50
5.4	Pearson correlation coefficients between the various variables, for the UK dataset, and p-values (2-tailed), N = 107. * Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.01 level. SADUS is Survey-based Average Daily number of Unhealthy Snacks. AADUS is app-based Average Daily number of Unhealthy Snacks.	52
5.5	Pearson correlation coefficients for the Dutch dataset and p-values (2-tailed), N = 196. ** correlation is significant at the 0.01 level. AADUS is app-based Average Daily Unhealthy Snacks.	53
5.6	CR, UE and EE in the UK dataset	55

5.7	Unhealthy snacking distribution over locations, days of week, and time bins in the two datasets. Healthy snacking instances were excluded from the Dutch dataset. The total number of unhealthy snacks in the UK and Dutch dataset are 4979 and 3297, respectively.	56
5.8	Time of snacking in the UK and Dutch dataset	56
5.9	Mean of snacks in the UK and Dutch dataset. ADUS is Average Daily number of Unhealthy Snacks	57
5.10	Trend of Snacking. 28 participants with only 1 week of recording were excluded from Dutch progress table.	58
6.1	Classification tasks description.	66
6.2	the size of the training and testing sets for the classification tasks in the UK and Dutch datasets with 4 and 12 time bins.	67
6.3	The test accuracy of imputed locations for time bins without snacking using, K-nearest neighbour imputation.	68
6.4	The accuracy of predicting the occurrence of unhealthy snacking in time bins of 4 and 12 in the two main datasets, UK (self-collected) and Dutch. The target variable was unhealthy snacking (yes, no).	68
6.5	Results of paired t-test with 4 degrees of freedom and 0.05 significance level on 5-fold CV, under the null hypothesis that the models have equal performance.	68
6.6	The Hyper parameters of the best performing models for predicting the of unhealthy snacking occurrence in time bins of 4 and 12 in the two main datasets, UK (self-collected) and Dutch. The models were optimised with random search. n-E is the number of estimators and LE is the learning Rate.	69
6.7	The metrics for predicting the occurrence of unhealthy snacking in time bins of 4 and 12 in the two main datasets, UK (self-collected) and Dutch. a) Binary-class metrics for the UK Dataset, b) Binary-class metrics for the Dutch dataset.	71
6.8	Prediction of healthiness of the next snack in the Dutch dataset. The frequencies of the target classes were 84% and 16% in 4 and 12 time bins.	73
6.9	The hyper parameters of the best performing model for the Prediction of healthiness of the next snack in the Dutch dataset. The models were optimised with random search.	73
6.10	The accuracies for predicting the number of unhealthy snacks consumed in time bins of 4 and 12 in the two main datasets, UK and Dutch. The classes of the target variable are 1,2 and 3+.	78

6.11	The metrics for predicting the number of unhealthy snacks consumed (1,2 and 3+) in time bins of 4 and 12 in the two main datasets, UK and Dutch. a) Multi-class metrics for the UK Dataset b) Multi-class metrics for the Dutch dataset	79
6.12	Estimating the number of unhealthy snacks consumed when using 4 and 12 time bins in the UK dataset after oversampling. a) Accuracies. b) Multi-class metrics for the UK dataset. The classes of the target variable are 1,2 and 3+. The target classes were balanced.	80
6.14	The Hyper parameters of the best performing model for predicting the number of unhealthy snacks consumed in time bins of 4 and 12 in the two main datasets, UK and Dutch. The models were optimised with random search after balancing the output classes.	81
6.13	Estimation of the number of unhealthy snacks consumed as 0,1,2 or 3+ in time bins of 4 and 12 in Dutch dataset. a) Accuracy b) Multi-class metrics for the Dutch dataset. The frequencies of the target classes in 4 time bins were 48%, 31% , 13% and 8%. The frequencies of the target classes in 12 time bins were 52%, 33%, 9% and 6%.	81
6.15	Regression task description.	82
6.17	The Hyper parameters of the best performing model for or predicting the time of next unhealthy snack in minutes, when using 4 or 12 time bins, in the two main datasets, UK and Dutch. The models were optimised with random search. n-E is the number of estimators and Min_SS is the minimum samples split.	83
6.16	MAE for predicting the time of next unhealthy snack in minutes, when using 4 or 12 time bins, in the two main datasets, UK and Dutch.	83

Chapter 1

Introduction

Studying and modelling human behaviour is useful in nearly every aspect of life. It has also aided progress in fields such as education, organizational behaviour management, and public health (Biglan et al., 2015). In healthcare, deeper understanding of health-related behaviour can play a significant role in improving people's health and well-being. Behavioural factors are important in preventing and treating many diseases, such as heart disease, cancer, and stroke (Kasl and Cobb, 1966). Diet, alcohol and tobacco are considered as major behavioural influences on health with variation across cultures. Such behaviours are found to be the leading causes of cardiovascular disease and early deaths (WHO, 2021).

Therefore, helping individuals to modify their behaviour can greatly reduce the personal, social, and economic loss associated with these behaviours (Weber and Fore, 2007).

1.1 Research problem and motivation

In today's world, where food and prepared snacks are constantly available, many people find it difficult to control their eating behaviour (Kelly and Barker, 2016). At the same time, obesity is a major health risk, and most people would like to lead a healthy life (Haththotuwa et al., 2020). This contradiction has been often recognised and is evident from the fact that there are countless magazines, television and Internet sites with recommendations and tips on what, how and when to eat (Kondracki et al., 2002). Yet, obesity has risen among adults, reaching epidemic proportions as the most serious health problem in both developed and developing countries. If current trends continue, a majority of the world's adult population will be overweight or obese by 2030 (Haththotuwa et al., 2020). Excessive fat accumulation is the major public health problem that increases the risk of cardio-metabolic multi-morbidity and mortality by up to twenty-three times. At the same time, the annual cost of treating obesity and its associated disorders is expected to exceed \$12 trillion by 2025. Despite the

fact that pharmacological (e.g., orlistat) and surgical procedures (e.g., bariatric surgery) have been effective and quick procedures for weight loss, people fail to sustain weight loss due to lifestyle (Chew et al., 2021).

Changing human behaviour is difficult and it has been found that most efforts made to change health behaviours achieve limited success (Kasl and Cobb, 1966). The complexity of changing health behaviour arises from psychological, social, political and economic factors that impact peoples' life. For example, advertisements prominently promote many unhealthy snacks and modern lifestyles often leave little time to prepare home-made meals. The apparent difficulty to control and change behaviour, despite the best of intentions, leaves the desire for support (Kelly and Barker, 2016).

One way to implement healthy eating interventions is in terms of 'nudges', i.e., reminders just before an unhealthy snack is consumed. Even without formal argument, it can hopefully be appreciated that a nudge to have a healthy snack right before a person plans to consume an unhealthy one could potentially be quite effective. Some preliminary work along these lines has already been conducted Spanakis et al. (2017). However, a key limitation is the capacity to predict these instances in the first place. Before we trial such interventions, we first need to understand just how well unhealthy snack eating behaviour can be predicted (and how quickly, given an assumed sparsity of the behaviour). Machine learning techniques can therefore be effective in supporting behaviour change by modelling and predicting human behaviour from data. More specifically, predicting human behaviour can enable the delivery of timed interventions in which could potentially lead to behaviour change.

Are there available data about personal eating behaviour? Despite the significance of this issue, we found that most — though not all — of the openly available datasets on nutrition or eating behaviour are aggregated, providing trends and patterns about groups. A very notable exception is the study by Spanakis et al. (2017), which involves an elaborate data collection procedure with questionnaires. Although such an approach is extremely valuable for perhaps theory development, it is impractical as a system for (a) predicting unhealthy eating behaviour and (b) offering timely corresponding nudges. Therefore, we decided to address this shortcoming in the literature by collecting our own dataset, restricted to just instances of unhealthy snacking, and by tracking a sample of individuals in the UK, where this project was conducted.

To accomplish data collection as naturally and unobtrusively as possible, we designed and developed an app for smartphones since, trivially, the ubiquity of smartphones would facilitate data collection. Then, with the collected data, we employed data-driven modelling with statistics and Machine Learning (ML) tools, following the state of the art in analogous

analytic situations (Solomatine et al., 2009). Note that the app development will enable the expansion of this project in future work to include interventions for behaviour change.

One question is how well the present approach and models may be generalisable to different cultures. To address this question, we will compare our data to the above mentioned dataset from the Netherlands by Spanakis et al. (2017).

We will consider analysing the collected and obtained datasets by applying descriptive analysis and by looking for differences in unhealthy snacking amongst different demographic groups. Applying descriptive analyses prior to ML models will aid in understanding the distributional characteristics of our dataset, as well as assessing if the various dataset partitions we will employ are balanced or not (Bonaccorso, 2018). The dataset modelling with ML tools will primarily aim to predict the occurrence of unhealthy snacks. Specifically, we will consider: (1) the occurrence of unhealthy snacking in different day/time windows, (2) the classification of snacking instance as healthy vs. unhealthy, (3) the prediction of the number of unhealthy snacks in day/time windows, and (4) the prediction of time until the next unhealthy snack.

We will primarily assess the effectiveness of the various ML models in terms of their capacity to predict the time of unhealthy snacks (Cai, 2020). As noted above, a good time for an intervention to avoid doing something, such as not eating an unhealthy snack, is right before making a decision (Purohit and Holzer, 2019a).

Overall, through this project, we aim to contribute to behaviour change support technology, by designing an app for data collection on unhealthy snack eating behaviour, collecting corresponding data, analysing the data, and building and evaluating predictive models for unhealthy snacking. This research comprises the following steps:

1. The design and development of a smartphone app for collecting data on unhealthy snacking behaviour
2. The corresponding data collection process and considerations
3. The analysis of the collected data
4. The computational modelling of the collected data
5. The evaluation of several ML models regarding the prediction of unhealthy snacking behaviour

1.2 Research questions and objectives

In this thesis we aim to answer three research questions that we have identified based on the current literature from psychology, human behaviour, persuasive technology, and machine learning applications. Chapter 2 offers the background and related work for each of the research questions.

- **Research Question 1** How can we design and develop an app to collect data about unhealthy snacking?
- **Research Question 2** How can we understand data on unhealthy snacking behaviour?
- **Research Question 3** To what extent can we predict unhealthy snacking with the aim of supporting targeted interventions?

Our specific objectives that help answering these questions are:

- **O1** Designing and developing an app for collecting snacking behaviour.
- **O2** Collecting data on unhealthy snacking behaviour with the app.
- **O3** Performing analyses on the datasets for the purpose of understanding distributional patterns amongst different demographic groups.
- **O4** Performing analyses on the datasets for the purpose of understanding other patterns of variation.
- **O5** Applying machine learning models for predicting unhealthy snacking behaviour.
- **O6** Evaluating the machine learning models for their potential to support timed interventions.

1.3 Research contributions

This work makes the following four contributions to knowledge in the field of data collection, analysis, and applied machine learning for health-related behaviour change.

- Development of an app for tracking and collecting unhealthy snacking behaviour from individuals. Such an app is a unique research tool, which can be easily reused, extended, and modified both for related data collection exercises and for providing interventions at a future point.

- A rich dataset on individual unhealthy snacking behaviour. This dataset offers the first opportunity to analyse individual snacking behaviour of UK-based subjects.
- Analysis and comparison of two non-aggregated datasets on snacking behaviour (the one we collected and the one collected by Spanakis et al. (2017)).
- Models for predicting unhealthy snacking behaviour and analysing their potential for supporting timed interventions.

The analysis and models produced in this work are aimed at two beneficiaries:

- Health psychologists and software developers interested in modelling and developing interventions for unhealthy eating behaviour.
- Ultimately, individuals looking to improve their eating behaviour by using a tracking app and, in the future, targeted interventions.

1.4 Thesis structure

The remaining chapters of the thesis are organized as follows:

- **Chapter 2: Literature review** - this chapter provides an overview of previous studies related to human behaviour utilising machine learning models. It will highlight some related theories for interventions in the case of human behaviour. We will also consider smartphone apps implementing behaviour interventions.
- **Chapter 3: Methodology** - this chapter provides the methods for the whole thesis. Additionally, choices of methods, tools and important input features are better justified now.
- **Chapter 4: Collection of Snacking Data** – in this chapter we cover unhealthy snacking data collection considerations, process and tools. We develop the app design for collecting the data in this project.
- **Chapter 5: Snacking Datasets Pre-processing and Descriptive Analysis** – in this chapter we introduce the snacking behaviour datasets with descriptive analysis and looking at behaviour patterns.
- **Chapter 6: Models of Unhealthy Snacking Behaviour** – in this chapter, we develop predictive models, describing modelling methods, results and analyses.

- **Chapter 7: Application and Intervention Design Considerations** – in this chapter we evaluate the potential of predictive models for timing interventions.
- **Chapter 8: Conclusion** - this chapter concludes the thesis by summarizing the contributions and limitations, as well as suggesting future research options.

Chapter 2

Literature Review

2.1 Behaviour modelling and machine learning

Modelling and predicting human behaviour is a difficult task and Machine Learning (ML) is a promising approach to address it. The practical benefit we are working towards is that timing interventions based on ML models can optimize solutions that support behaviour change. The logic for this proposal is that many interventions operate at the level of nudges just prior to the target behaviour and, so, require timing information for the target behaviour (Chew et al., 2021). Although this work does not produce actual interventions, we aim to complete the essential, initial steps of designing an app for recording behaviour and testing the effectiveness of different ML models in predicting unhealthy snacking behaviour. This section will discuss generic ML models that have been applied to human behaviour modelling, followed by a summary and review of some previous studies where ML models have been applied to this task.

According to Bishop (2007), ML can be categorized into three main types, based on the nature of the learning target:

- **Supervised learning:** The inputs and their targeted output are offered to the algorithm and the learning process aims to produce a mapping that assigns inputs to target output. For example, predicting the next step of human action according to a given inputs.
- **Unsupervised learning:** Data are offered to the algorithm without labels. The algorithm should learn the structure of the input data. The goal of this type of learning is discovering hidden patterns in datasets that can be of benefit in tasks like anomaly detection or as a preparation for supervised learning (semi-supervised learning). Unsupervised learning could be used when the goal of the prediction is to recognise behaviour patterns in a group of individuals and partition the data according to behaviour patterns.

- The third category is Reinforcement learning. However, supervised and unsupervised learning are the categories applied to human behaviour modelling more commonly (Monekosso and Remagnino, 2009).

Supervised ML tasks can be categorized into classification ones or regression ones, depending on the targeted output variable and the nature of the problem (Bishop, 2007).

- **Classification:** In this category of tasks, inputs are divided into discrete classes of activities or behaviours and a model is constructed to assign inputs to the classes (Russell and Norvig, 2020). The most common approaches for classification problems are logistic regression, Decision Tree (DT), Random Forest (RF) and Artificial Neural Networks (ANN). Logistic regression is an algorithm for predicting a categorical dependent variable by modelling its relationship to the independent variables (Bhandari and Joensuu, 2011). Decision Trees (DT) are algorithms that break down classification into a sequence of decisions based on utility, as measured by, for example, entropy or the Gini coefficient. Despite their simplicity, DTs can be very effective classifiers (Giudici, 2005).
- **Regression:** This approach requires a continuous outcome variable (of course, one might counterargue that logistic regression is a type of regression too; we do not intend to imply that these two categories of techniques are sharply distinguished). For example, the prediction problem might be to quantify how much an individual is going to eat tomorrow. Common regression approaches are NN, linear regression, generalised linear models, and regression trees. Linear regression algorithms model the outcome in terms of a linear combination of independent variables. Regression trees are similar to DTs and employ “recursive partitioning”.

Another important and common task is Clustering, which is a learning task that divides the datasets into subsets, the clusters. Clustering techniques can be supervised or unsupervised. Clustering can be employed to reduce the number of dimensions in large datasets (Bishop, 2007).

For many decades, there have been numerous studies which have addressed the modelling of diverse aspects of human behaviours. For example, Hidden Markov Models (HMMs) and neural Networks are popular tools applied to human behaviour problems (e.g. Dix et al. (1992); Pentland and Liu (1999); Yang et al. (1997); Bregler (1997); Ashbrook and Starner (2002); Vasquez et al. (2008); Zheng et al. (2008); Mahmoud et al. (2013); Karki et al. (2016); Wang et al. (2015)). Hidden Markov models (HMMs) and NNs, particularly recurrent ones, are popular due to their ability to model sequential data. HMM and Recurrent

Neural Networks (RNN) offer a way to represent the sequential trace of behaviour in terms of state transitions.

Common aspects of modelling behaviour will be summarised and reviewed with relevant studies in the next sections.

2.1.1 Modelling human actions and interactions

One of the early studies tracing human actions, interactions and skills with ML was performed by Dix, Finlay and Beale in (Dix et al., 1992). They used a recurrent version of A Distributed Associative Memory (ADAM) model for modelling human interactions. This method considers the state transitions as a solution for time series data that represents human interactions. The results offered some success in tracking human actions in relation to some relevant context. Pentland and Liu (1999) used HMM in modelling human behaviour from sensory data. The model was tested with a task of recognising automobile drivers' behaviours and achieved 95% accuracy. The results illustrated that the methodology could be generalised to other dynamic human interactions. Yang et al. (1997) proposed a framework based on HMM for modelling human action-intent and skills from observations with no sensory feedback. The results were promising to the extent that they could recommend the transfer of human skills to robotic systems. Jebara (1998) constructed a model called Conditional Expectation Maximization to model human interactive behaviour aiming to analyse and synthesize human behaviour. The model was based on the probabilistic relationships between an action and its consequent reaction. The model was built and trained from data to predict the most likely behaviour resulting from an individual action. The results proved that the methodology could be successfully implemented in an Action-Reaction Learning framework. Although these are obviously cherry-picked examples from the extensive literature, they do illustrate that, for some time now, ML models have been employed with consistent success in the description of sequential tasks, not unlike the one we propose to study.

As another example, human action recognition has been extensively applied in smart manufacturing and human-robot collaboration. Most of this research is focused on computer vision ((Cheng et al., 2015); (Kong and Fu, 2018)), which is usually conducted in two steps: feature extraction and action classification. Feature extraction converts the data in video frames into feature vectors with action-related patterns (Bilinski and Bremond, 2011). Once the features are extracted, classifiers (e. g. RF and support vector machines can be used to classify the features into appropriate action classes (Hu et al., 2009).

Deep Learning (DL) has emerged as a promising model that allows for the automatic extraction of multi-level features through hierarchical characterisation of images, employing multiple network layers, followed by a conversion to conditional probabilities that demon-

strate how inevitably the input image relates to certain action types ((LeCun et al., 2015); (Wang et al., 2018)). DL models, such as the convolutional neural networks, have shown promising performance in recognising human actions, voice commands, and hand gestures ((Wang et al., 2018); (Liu et al., 2018); (Tao et al., 2020); (Mazhar et al., 2019); (Nuzzi et al., 2021)). Furthermore, the neural architecture has been examined as a specific way to extract information about the human pose (Papandreou et al., 2018). The output of this system served as input to a classifier for action detection, with some success, (Liu and Wang, 2021).

2.1.2 Modelling human daily living activity

The task of monitoring (and modelling) daily living activities has been addressed in several studies utilising sensors installed in living spaces. NNs have often been applied for the task of inferring activities from sensor data in order to identify the patterns of particular activities at a given time. Zheng et al. (2008) applied a self-organising neural network called Growing Self-Organizing Maps (GSOM) to cluster human activities. The approach was able to detect usual and unusual behaviours in daily activity data. Choujaa and Dulay (2010) used a Bayesian probabilistic model to quantify the relationships between mobile phone users' activities. The method allowed some prediction of future activities, as well as activities of other user (from relevant prior data). Mahmoud et al. (2013) applied dynamic RNN to study the daily activities of elderly occupants suffering from dementia. Occupancy sensors were utilized to monitor the motion patterns of the occupants. Behaviours were then predicted after converting the recorded data into temporal sequences of activities. The results demonstrated that the approach correctly extracts the long-term behaviour patterns of the occupants. Moreover, RNN have offered extraordinary capacity in discovering temporal relationships between input patterns. Corresponding results have been validated in several ways, for example, by utilizing hypothetical and real environment data (Mahmoud et al., 2013). Karki et al. (2016) proposed a framework based on DL models to infer daily activities from YouTube videos and public sources. DL models are suitable for learning tasks such that representation would benefit from network architectures with more than one hidden layer. The framework represented in the Karki et al. (2016) study could allow identifying diverse types of activities in surveillance video as well as allowing for a good degree of recognition of individuals.

In recent years, DL models have gained popularity owing to their promising performance and ability to learn high-level features ((Saba et al., 2019); (Biswas et al., 2019)). Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, as variants of RNN, have been applied in some previous studies. For example, Murad and Pyun (2017) applied LSTM to collect discriminative characteristics from raw sensor input sequences obtained

from body-worn sensors. Mekruksavanich and Jitpattanakul (2021) also implemented LSTM, utilising smartphone sensor data obtained from a public dataset. They compared five LSTM networks to investigate how different architectures affected the results. Liciotti et al. (2020) applied LSTM for modelling spatio-temporal sequences collected by smart home sensors. They evaluated their model on the Center for Advanced Studies in Adaptive Systems dataset. The findings showed that LSTM outperformed other ML and DL approaches. Vu et al. (2017) employed self-gated RNN, utilising sensor data collected from wearable devices. Their method outperformed the standard RNN model. However, it offered results similar to those of the LSTM and GRU techniques.

2.1.3 Modelling online purchase behaviour

Some of the existing studies have focused on online purchasing behaviour, because of its obvious importance on enhancing advertising tools and pricing procedures (Gupta and Pathak, 2014). Purchasing behaviour and eating behaviour share some commonalities (e.g., in both cases there is an element of sensitivity to reward), so models that work for purchasing behaviour may also work for eating behaviour. Naive Bayes, Logistic Regression and Support Vector Machines were used in the (Zhang and Pennacchiotti, 2013) study to predict online purchasing behaviour. The dataset was collected from Facebook, social network, and eBay, e-commerce website. The relations and interactions between user profiles in social network websites and user online purchase behaviour were studied using the algorithms. The results indicated that social media information and online purchases are significantly correlated and the purchase intention can be efficiently predicted utilising social media information alone. A Logistic Regression model was also applied in the Gupta and Pathak (2014) study to address the problem of predicting online purchases. The study focused on customers in segments for predicting the purchasing of specific items, rather than on individual customers' intentions. The data was captured from different sources such as: visitor attributes and purchase history. The data were used to identify appropriate pricing for items and the prediction for their likely purchase in particular pricing brackets. The results were encouraging to the extent that the prediction error rate was reduced when offering a better price range.

User interactions used to predict buying intention includes time spent per item, search frequency, returning visits, and clickstream data. The results demonstrated that DL models allowed a considerable improvement by detecting features from high dimensionality data. Wang et al. (2015) proposed a methodology that identifies correlation accounts of users between social network sites and e-commerce sites, enabling an inclusion of account profiles and past user behaviours. Utilising correlation patterns in the above data, a predictive method based on a DL model was employed, in order to predict purchasing behaviours. The experi-

mental results illustrated that the method was able to identify 70% of correlated accounts between Facebook and eBay. Yang et al. (2022) investigated energy price predictions using an Artificial Neural Network (ANN). They created a novel predictive model that includes adaptive data pre-processing, improved optimisation, a kernel-based model, and an optimum model selection strategy for more accurate energy price predictions.

2.1.4 Modelling eating behaviour

ML applications in eating behaviour have been addressed in the literature mainly in three categories: machine perception, predictive analytics, and real-time analytics with personalised micro-interventions. Machine perception refers to the application of ML to detect and extract features, classify and interpret (recognize) data received from wearable/non-wearable devices – e.g. a vision camera, proprioception (gestures), and audition (sound). Predictive analysis is the process of using historical data and statistical approaches (such as data mining and modelling) to forecast future events. Previous studies on predictive analytics focused on creating prediction models based on behaviour data (eating and exercise), nutrition, target achievement rates, anthropometric data, views (e.g., blog posts), and ecological factors for predicting weight loss and behaviour lapses. Real-time analytics refers to the immediate analysis of past and current data, in order to train, test, and optimize predictive models, as well as deliver relevant nudges for behavioural lapse risks and recommendations as micro-interventions.

Alshurafa et al. (2015) have developed a necklace as a novel form of wearable system that gathers data from an embedded piezoelectric sensor which has the ability to detect skin motion in the lower trachea during the consumption of food/drink. An output voltage with different frequencies over time was produced from the skin motion. In order to precisely discriminate between food types, for instance, liquid and solid, hot and cold drinks, and hard and soft foods, an algorithm based on time-frequency decomposition, spectrogram analysis of the piezoelectric sensor signals was developed. A smartphone was used to receive the data from the necklace, followed by processing the signals and classifying the food type, so that visual feedback could be provided to assist the user in observing his/her eating habits over time. In addition, the authors made a comparison between their approach (spectrogram analysis) and other time-frequency features (matching pursuit and wavelets), to conclude that their approach outperformed other techniques. This was in terms of various performance indices, such as high accuracy for discriminating between food types as well as the extraction of key characteristics (features) representative of the distinctive swallow patterns of numerous foods.

Aswani et al. (2019) developed an approach to increase human adherence to diet via personalised treatment to each patient. They built a framework used for predictive modelling through the use of utility functions depending on time-varying system and motivational states which evolve based on some modelled process. In turn, this would contribute to a model of behaviour change based on various qualitative social science models. In order to implement the predictive model, these researchers needed to solve a 'bilevel program' through building a mixed-integer linear program. This approach has allowed Bayesian inference that utilises empirical histograms as prior distributions. In addition, they investigated the ability of their predictive models with a dataset on weight loss intervention; this exercise included a validation of results.

Goldstein et al. (2018) attempted to combine group and personal data in order to predict diet lapses and prevent them through the utilisation of ML. They built an algorithm for predicting dietary lapses amongst adults who were overweight/obese (but note, their sample was small, $n = 12$). The participants were required to record their data using ecological momentary assessment connected with a smartphone (EMA) about lapse and corresponding triggers. They then used a software tool called WEKA for performing analyses using decision trees, to predict lapses with an accuracy of 72% for the group, but with poor results for individual participants. Additionally, they suggested that a period of a month for individual data collection would be the optimal data collection period to reach reasonable model performance.

Spanakis et al. (2017) provided a framework for utilizing ML techniques to fully exploit data acquired from a mobile application called "Think Slim," developed to measure eating habits using a so-called experience sampling approach. The goal of their study was to measure the impact of individual states of a person's status (emotions, location, activity, etc.) on unhealthy eating. they utilised a classification algorithm (decision tree tailored to longitudinal data), in order to warn people before a possible unhealthy eating event. They found that participants could be grouped into 6 groups according to their eating behaviour, and they were able to identify rules which could enable the classification of healthy and unhealthy eating instances, based on the other information collected at each time point. They utilised these rules to provide customised feedback to their participants so as to help them to avoid unhealthy eating. They concluded that the adoptive customised feedback was effective. This study is close in its objective to what we are presently interested in. A key difference is that in this work we aim for a minimal data collection procedure, whereas Spanakis et al. (2017) employed a rather detailed questionnaire (even if designed to be brief and easy to complete), at several time points on each day.

Forman et al. (2019b) developed an app called “OnTrack” in order to exploit ML potential to predict dietary lapses (deviation from adherence to dietary prescription). Their aim was to help design and deliver a targeted intervention to prevent dietary lapses. Logit boost, bagging, random subspace, random forest, and Bayes nets were all used in OnTrack’s cost-sensitive ensemble model. The study was focused on a group of people ($N = 43$; 86% female; body mass index = 35.6 kg/m²), who were on a diet for a period of 8 weeks. They found that the participants were likely, with the help of app reminders, to submit their data on time engage with the intervention, and show high levels of contentment. In addition, during the course of the study, 3.13% was the average weight loss, and there was a decrease observed in unplanned lapses. Overall, the Just-in-time Adaptive Intervention (JITAI) that was developed was shown to be effective and to have helped with weight loss monitoring. The same research group, in another study, further employed the same OnTrack tool to collect data on lapses and triggers of lapses. They utilised a continuously improving ML model so as to predict any events of dietary lapses and deliver a JITAI when needed, similarly to what we described above. However, with this extension of their work, they offered a comparison between two approaches: the OnTrack app and the Weight Watchers (WW) digital program. They used a randomized group of people ($N = 181$) who were overweight/obese (MBMI = 34.32; 85.1% female; 73.5% White). Their participants received one approach, the OnTrack or WW one. They found that a general linear model revealed a treatment condition $\times$ diet plan interaction, such that OnTrack exhibited superior effectiveness only amongst those receiving Beyond The Scale (BTS). Furthermore, in comparison with Freestyle, participants who adopted BTS WW with OnTrack showed greater contentment with the intervention and engagement; in addition, algorithm accuracy was better. These researchers concluded that there were concrete results supporting the OnTrack approach and generally the ML-powered JITAI, for the purpose of facilitating weight loss through predicting and preventing any events of dietary lapses (Forman et al., 2019a). The same group just recently utilised the same app (OnTrack) based on JITAI, to provide personalised interventions, in order to prevent nonadherence to dietary prescription (Goldstein et al., 2021). This time, they used an EMA for self-reporting lapse triggers, utilising ML models for the prediction of lapses. As before, the objective was to prevent lapses through interventions. The study was conducted on a group of people ($N = 121$) who were overweight/obese (MBMI = 34.51; 84.3% female; 69.4% White). These participants used the app for 10 weeks. The researchers found that there was an association between the reduction of lapses and weight loss during the use of the app (OnTrack) using Linear mixed models.

In the next section will provide an overview of the ML models which will be applied in this project.

2.1.5 Machine learning and time series predictions

Due to the fact that human behaviour is usually expressed in time, it is often studied as time series data (Bhandari and Joensuu, 2011). Dynamic probabilistic graphical models (e.g. HMM and Bayesian Networks (BNs)) and RNN are often considered to be appropriate for modelling time series data (Dix et al., 1992), although they are not the only approaches. Static models such as Decision Trees (DT) and Random Forest (RF) could also be applied on time series data by working with time windows in a way that each time window is considered as one data point (Jović et al., 2012).

The ML community has used probabilistic graphical models to tackle uncertainty in data. These models offer a framework to represent “independence relationships between variables, learning probabilistic models of data, and inferring the values of latent variables”. BNs are graphical models representing a number of random variables and their conditional dependencies through a directed acyclic graph (DAG). An example of such network could be the representation of the relationship between disease and symptoms in probabilistic form. A BN represents the random variables as nodes and the conditional dependencies as edges in the DAG. Nodes that are not connected are conditionally independent. All nodes correspond to a probability function representing a probability distribution of the dependencies and relationships. In a BN, probabilistic inference analysis is applied and the maximum-likelihood estimator is used to learn the parameters.

HMM is a Markov model with unobserved (hidden) states, controlled through transition probabilities, and a dependent, output state. HMMs are dynamic BNs. The emission is controlled by a matrix of the conditional probabilities for the observed variable distribution, given the hidden variable (Jensen, 1998). Tree-based models produce predictions from one or more decision trees using a sequence of if-then rules. All tree-based models can be used for classification or regression. All tree-based models are built on the foundation of decision tree models. RF models are an "ensemble" method for creating several decision trees at the same time which are commonly trained using the "bagging" method. The bagging method's basic premise is that combining several learning models improves the overall output (Ali et al., 2012). Extreme Gradient Boosting (XGBoost) is another ensemble method, however, it uses boosting trees instead to achieve its outputs. To determine the optimum tree model, XGboost uses a gradient descent approach that leverages more precise approximations (Huang and Yen, 2019). A Random Forest regressor and XGBoost regressor are type of RF and XGBoost, however for regression problems (Jensen, 1998).

ANN are ML models that loosely resemble the biological neural networks in the brains of animals and humans. They are applied mostly to problems involving classification or regression, but they are also now used for unsupervised learning, dimensionality reduction,

and generation of images and other artefacts. ANNs are typically employed in prediction rather than knowledge extraction, due to the difficulty in interpreting their parameters. The feed forward neural network (FFNN) was the first and most basic artificial neural network to be developed. The information in this network flows forward from the input to the prediction output in the application (often called inference) and backwards from the outputs through the model during learning, in a process called back-propagation. In the FFNN, there are no cycles or loops (Singh and Chauhan, 2009). RNN is a recurrent versions of ANNs have feedback connections in one or more layers of the network, allowing them to process observations as sequence (Alpaydin, 2013). RNN work in cycles, using network activations from a previous time step as a series of inputs to make predictions for the current time step (Medsker and Jain, 1999). LSTM is a type of RNN developed as a solution of the exploding and vanishing of the gradients that comes from the backward propagation step (Yu et al., 2019).

LSTM is a type of RNN developed as a solution of the exploding and vanishing of the gradients that comes from the backward propagation step (Yu et al., 2019).

2.2 Behaviour change

Many diverse theories, frameworks, models, approaches, and strategies are used in behaviour change research (Pinder et al., 2018). Changing human behaviour requires some theoretical model of behaviour. Behaviour theories are based on a combination of ideas from psychology, social science, and health sciences. Human behaviour theory explains or predicts behaviour events based on several systematically interrelated definitions and propositions (Glanz and Bishop, 2010). Social Cognitive Theory, the Transtheoretical Model/Stages of Change, the Health Belief Model, and the Theory of Planned Behaviour are the most-commonly applied theories in human behaviour interventions (Glanz and Bishop, 2010). The theory of planned behaviour (TPB) is an intrapersonal theory that is based on a person's belief about their ability to perform a particular behaviour. In TPB, peoples' attitudes, subjective social norms, and perceived behaviour control are the components that predict a behaviour (Lee et al., 2017).

There are different approaches influencing people to make them select particular choices or change their behaviour through design processes and strategies. Nudges and persuasive technologies are examples of design strategies that aim to assist peoples' decision-making through design (Castmo and Persson, 2018).

Nudge theory can be considered as part of the TPB (Armitage, 2015). It applies ideas and strategies from behavioural science, politics, and economics to offer indirect suggestions as an attempt to motivate individuals in achieving non-forced compliance (Arno and Thomas,

2016). A little change made by ‘Nudges’ forms un-thought-of choices in individuals and focuses on the long-term effect of an action rather than a short-term gain (John et al., 2011). Nudge theory also relates to aspects of cybernetics. Cybernetics is about control and dynamic systems. It includes concepts such as bifurcation points and attraction basins. The idea of a nudge in this context is that with a small (and non-mandating) intervention we can change the direction of a system or lift it over a threshold and thus can reach a substantially different outcome later more likely than otherwise.

Persuasive technology refers to a group of technologies that use psychological foundations to affect users in their decision-making. Persuasion is defined by Fogg (2002) as “an attempt to change attitudes or behaviours or both (without using coercion or deception).” Persuasion is achieved by using persuasive principles or strategies. Researchers have proposed various persuasive strategies for behaviour change over the years. Among these are: Fogg’s seven technical tools for persuasion (Fogg, 2002), Fogg’s behaviour grid (Fogg, 2009a), Fogg’s behaviour model (Fogg, 2002), Fogg’s eight-step process (Fogg, 2009b), Cialdini et al.’s six persuasive principles (Cialdini, 2009), and Oinas-Kukkonen et al.’s 28 persuasive strategies (Oinas-Kukkonen and Harjumaa, 2009). For this thesis, we target three strategies (from Cialdini (2009)) that are widely used in persuasive systems design (Oyebode et al., 2021). These strategies – self-monitoring, reminder, and suggestion - are described in Table 2.1.

Table 2.1 Description of the three persuasive systems designs strategies targeted in this thesis.

Persuasive Strategy	Definition
Self-monitoring	Allows users to keep track of their own progress or performance.
Suggestion	Provides suitable suggestions for achieving the targeted behaviour during tracking behaviour.
Reminder	Reminds or notifies users of their desired behaviour when during tracking behaviour.

Many commercial smartphones applications have been designed for changing user’s behaviour towards a healthier lifestyle, based on the strategy of self-monitoring and suggestion. Samsung Health is an example of such an application for monitoring and tracking daily activities, exercise, sleeping patterns and eating habits (Samsung, nd). Another example, the “Perfect Diet Tracker” application, tracks users’ diet, so as to achieve consistent weight loss: maintaining weight loss after it has been accomplished requires additional aids, for example, the app might provide suggestions for healthy eating (perfect Diet Tracker, nd). Another example runs on Apple watches and is called “Hello Heart” for blood pressure monitoring,

based on recording and uploading vital signs from the users' wrists (Cohen, 2021). However, note that there is very little information available that is employed in internal, algorithmic methods in these apps.

Researchers in the field of Human-Computer Interaction (HCI) have increasingly applied these strategies in their systems or designs to motivate users and achieve specific behaviours in various domains, including health. Persuasive systems based on the strategy of self-monitoring have been used in several areas, such as promoting an active lifestyle [e.g. (Altmeyer et al., 2018); (Muneer et al., 2020); (Ganesan et al., 2016); (Consolvo et al., 2006); (Miller and Mynatt, 2014); (Toscos et al., 2008)], in smoking cessation [e.g. (Paay et al., 2015)] and in weight management [e. g. (Rothert et al., 2006); (Teeriniemi et al., 2018); (Anderson-Bill et al., 2011)]. A suggestion strategy was implemented by (Fuller-Tyszkiewicz et al., 2018) as a list of recommendations that can be accessed in the app used in their study to improve the well-being of people experiencing depressive symptoms. Other studies provided users with a list of app-based dietary suggestions [e.g. (Salim et al., 2017); (de la Torre Díez et al., 2017); (Oyebode et al., 2020)]; or weight management suggestions (e. g. (Anderson-Bill et al., 2011); (Hunter et al., 2008)). A reminder strategy was employed in previous studies through pop-up messages or push notifications. Some studies considered meal reminders [e. g. (Haque et al., 2018); (van der Lubbe and Klein, 2019); (Wang and Reiterer, 2019), physical activity reminders (e. g. (Paldán et al., 2019); (Hurling et al., 2007)], smile recording reminders (Moore et al., 2017) or daily step count reminders (Edney et al., 2017). Reminders have also been used in smoking cessation (Karim et al., 2017); (Fenicio and Calvary, 2015) (Obermayer et al., 2004), diabetes in young people (Franklin et al., 2006), bulimia nervosa (Robinson et al., 2006) and weight management [e. g. (Teeriniemi et al., 2018); (Anderson-Bill et al., 2011)].

Eating behaviour change has been addressed within app platforms, where interventions implemented for mobile devices aim to support and influence health-related behaviour changes. Kerr et al. (2016) implemented mHealth for eating behaviour. Another study modelled eating behaviour, attempting to change overweight adults' behaviour also using mHealth (Mahmoud et al., 2013). Various researchers (Forman et al., 2019a,b; Goldstein et al., 2021; Spanakis et al., 2017), discussed in Section 2.1.4, also implemented interventions for weight loss and behaviour lapses.

2.3 Timed interventions

In Persuasive technologies for influencing behaviour, timing is a critical factor responsible for success of digital nudges (Purohit and Holzer, 2019b). Timed interventions relate to two

different types of thought processes that may influence behaviour: reflective and impulsive processes. Reflective processes are those that require more control, such as plans and tactics, whereas impulsive processes are those that are more automatic (and perhaps more amenable to "nudges"). The first category of thought processes is the target for the majority of health/behavioural change programs (e.g., education campaigns). People engaged in reflective processes, however, must be in a position that allows them to plan, practice self-control, and so on, to accomplish a change (clearly, this may be challenging for some people, Tapper (2021). Regarding nudges, their effectiveness has been shown to differ depending on when they are delivered. Early intervention may cause forgetting, while late intervention may limit the amount of time available to change behaviour (Gillitzer and Sinning, 2020). Overall, the smaller the gap between the intervention and the behaviour, the more likely the intervention will influence behaviour. In other words, items we have recently been exposed to are more cognitively accessible. Hofmann et al. (2008) present a dual-processing account of health behaviour to explain how impulsive (automatic) processes or reflective (controlled) influences can impact behaviour. Unhealthy snacking can be an automatic activity, therefore reflexive techniques might be required to prevent or change it (i.e. exercise self-control). Situational moderators, according to Hofmann's concept, can shift the balance between reflective and impulsive processes (Hofmann et al., 2008). Our behaviour is only influenced by plans/intentions (e.g., to minimize unhealthy snacks) via impulsive / inhibitory forces, according to PRIME theory, introduced by West and Brown (2013). According to West and Brown (2013), an intention must create a specific emotional state in order to influence an inhibiting force (for example a desire to eat more healthily or a feeling of anxiety about becoming overweight).

McBride et al. (2003) developed the concept of teachable moments in health education research to urge people to modify their behaviour. Naturally occurring health events are expected to drive individuals to adopt risk-reducing health habits spontaneously, according to the definition of a teaching moment. During the perinatal period, for example, a smoking cessation strategy for women could be beneficial (Ockene et al., 2002). Some studies presented the importance of timed nudges to achieve a personal health target. For instance, Intille et al. (2003) suggested that the relevant data should be loaded onto a PDA personal digital assistant when the user is at the point of purchase (relevant nudge moment). Another study looked into how to change eating habits in obese teens to help them lose weight. The researchers provided participants real-time feedback (nudge moment) during meals to encourage them to eat more slowly (Ford et al., 2010).

2.4 Ethical issues and considerations

The ubiquity of mobile devices and the advancement of information technologies have resulted in a tremendous rise in the amount and diversity of data, called big data, collected about individuals. Big data collection and analysis have created serious privacy and ethical concerns (Gambs, 2018). Although privacy risks have existed for a long time because of access to personal data, big data magnifies these risks and makes them more difficult to realise and predict because of the capability to integrate data from multiple sources. Furthermore, the inferential ability of big data models increases privacy and ethical challenges (Gambs, 2018).

Persuasive technologies can be intrusive due to their reliance on data about users' lives. Consequently, they can even be misused as a surveillance tool (Purpura et al., 2011). It is therefore important to ensure that user data is private, especially in health applications because health data are particularly sensitive.

Furthermore, it is crucial that the line between persuasion on the one hand and coercion and deception on the other hand is not crossed, even if it is sometimes perceived to be blurred (Aldenaini et al., 2020). According to Aldenaini et al. (2020), the ethical considerations of persuasion that need to be considered in persuasive technologies are as follows:

- The app needs to be tailored solely to the needs of users.
- Collecting data and the intrusion between the privacy and autonomy of the user must be balanced.
- Persuasive technologies need to be developed, so that users know how and why their behaviour is tracked, especially for persuasive technologies that track health-related behaviour.
- The persuasive technology design needs to be built in such a manner that the autonomy of the users is not compromised.

In sections 4.1 and 4.3, we demonstrate how we consider the ethical concerns described above in all empirical work conducted in this project.

2.5 Gap in the literature

As presented in this chapter, previous studies have highlighted the importance of ML in human behaviour modelling for influencing or changing human behaviour.

Some studies have also addressed eating behaviour modelling and interventions for weight loss and behaviour lapses. In contrast, our focus is on predicting unhealthy snacking behaviour to support behaviour change. Predicting unhealthy snacking behaviour could support a range of interventions to prevent obesity problems, or even for underweight and normal weight people. Despite the significance of this problem, we found very little work has been done on modelling snacking behaviour, except for Spanakis et al. (2017) study, which modelled generic eating behaviour, including snacks. Spanakis et al. (2017) conducted an extensive data collection involving brief questionnaires in the Netherlands. Although such an approach is very beneficial for theory development, it is impracticable for systems that (a) predicts unhealthy eating behaviour and (b) provides timely nudges in response. Consequently, we decided to address this gap in the literature by collecting our own dataset with a snacking-focused smartphone app in the UK, to understand how well unhealthy snacking can be predicted in this context to support behaviour change interventions.

Chapter 3

Methodology

In this chapter, we describe the methodology we employ in this research. The methodology will be presented separately for each of five main stages: (1) survey and app design for data collection, (2) data pre-processing, (3) data analysis, (4) predictive models, and (5) hypothetical interventions designs and considerations. We will outline each of the stages in a separate section.

3.1 Survey and app design and data collection

As noted in Section 2.5, data collection was required in the thesis because no dataset directly suitable for our purpose was available. We first obtained ethical approval from the Research Ethics Committee of City, University of London's Psychology Department before proceeding with the data collection. Additionally, we addressed privacy and security considerations to ensure compliance with the General Data Protection Regulation (GDPR, <https://gdpr-info.eu>), as detailed in Section 4.1.

The data collection was conducted in two parts: first, we created a survey to explore various assumptions about the target behaviour of interest, and, second, we implemented an app to obtain realistic data for app users in the UK. The survey was self-administrated, including demographics and questions about eating behaviour. Eating behaviour questions were based on the Three Factor Eating Questionnaire (TFEQ), which is related to the development of obesity (Cappelleri et al., 2009), as shown in Table 4.3. TFEQ is widely employed in psychology studies to assess the eating behaviour aspects of Cognitive Restraint (CR), Uncontrolled Eating (UE), and the Emotional Eating (EE) (Cappelleri et al., 2009), again as detailed in Section 4.2.1.

The app-based part of the study was conducted with our “Snack Tracker” app, which was developed specifically for this project to record unhealthy snacking. I designed the app

myself, and the system was implemented by Dev Technosys (for more details about the company, see <https://devtechnosys.com>). Mobile app development is usually divided into two main components, frontend and backend (see figure 4.2). In this project, the frontend is the app on Android and iOS devices with which users navigate and interact. The backend includes the database, a web-based admin panel, and a hosting server. Based on the results from the survey, participants who reported having two or more daily unhealthy snacks were invited to the app-based part of the study, which involved monitoring their snacking behaviour for 28 days. We asked the participants to keep track of their location whenever they ate an unhealthy snack. Each recording was automatically saved with the current date and day of the week. Figure 4.1 shows the flow of the two studies. The cost of this data collection was funded by the Saudi Arabian Cultural Bureau (SACB) in London. The data collection and mobile app development are described in Chapter 4.

3.2 Data pre-processing

Real-world data are often incomplete, inconsistent, and sometimes inaccurate. Data pre-processing is the process of data cleaning, integration, or transformation in order to increase the collected data quality and achieve reliable analyses and predictions (Hellerstein, 2008). Each dataset is unique and different, which makes data preparation challenging. The number of variables, the types of variables (integer, binary, nominal, ordinal), the scale of the variables, and other factors all affect the size of the dataset and lead to different approaches to data pre-processing (Brownlee, 2020). Data pre-processing includes data filtering, normalisation, feature selection, and other procedures, depending on the dataset description and research methodology (Kuhn and Johnson, 2013). The data pre-processing techniques employed on our UK and the Dutch dataset, that we obtained from another research group, include the following:

- Data cleaning:

Data cleaning involves detecting and treating inaccurate, incomplete, and duplicate records in a raw dataset (Kuhn and Johnson, 2013). The UK dataset was a raw dataset in CSV format. It was exported from our app server. Although the Dutch dataset was already cleaned for other analyses, it required some additional cleaning and pre-processing steps to put it in a similar format to what we had for the dataset we compiled directly (in the UK)) so that the same analyses and modelling could be applied in both datasets.

- Feature engineering:

A feature is a variable that is useful or meaningful to a phenomenon being observed. Any variable could be a feature, as long as it is valuable to the model (Nargesian et al., 2017). Feature engineering involves extracting features from raw data using domain knowledge. Feature engineering is sometimes crucial to aid the analysis of data and the performance of Machine Learning (ML) algorithms (Nargesian et al., 2017). We constructed several features in the UK and Dutch datasets required for the analysis and/ or ML models, as well as to achieve better overall results.

- Data transformation:

Data transformation converts raw data into a format or structure suitable for modelling (Heer et al., 2015). It is important to ensure a sufficient number of instances representing each level of a categorical variable to obtain robust models and analyses. This means that categories with few instances need to be merged. Moreover, most ML models require input data to be integers, and most algorithm implementations adhere to this requirement (Brownlee, 2020). Feature scaling is also required for ML models to ensure all features have a standardised range or distribution. Standardisation aligns the variables so that the resulting distribution of the mean and standard deviation becomes 0 and 1, respectively. Feature scaling is useful and often improves the performance of ML algorithms that use gradient descent as an optimisation technique, such as logistic regression and neural networks (Albon, 2018).

We considered feature transformation in the pre-processing stage to ensure a sufficient number of instances for each level of a categorical variable. Additionally, we ensured that all features were in a format acceptable to the ML models we considered and that all features were scaled.

- Oversampling to re-balance classes

Oversampling is a method for adjusting a target variable's class distribution by creating more examples if the class is less frequent. The simplest method is to duplicate samples, but this can lead to an over-representation of a small number of samples for infrequent classes. This can be addressed by Synthetic Minority Oversampling TEchnique (SMOTE). SMOTE is an oversampling technique that involves choosing examples in the feature space that are close together, drawing a line in the feature space between the examples, and creating new samples at points on that line (Chawla et al., 2002).

3.3 Data analysis

Descriptive analysis is crucial before applying ML to develop some understanding of the dataset. In addition to enabling a holistic view of the data, descriptive statistics can give you better insight into each attribute and its distribution (Bonaccorso, 2018). We report descriptive statistics to better understand the participants' attributes, such as gender, age group, and employment status in both datasets. We also applied correlation analysis to the relationships between rates of unhealthy snacking among different sets of participants, based on demographics and personal eating characteristics in both the UK and Dutch datasets. We also use box plots to visually present and understand the data distribution based on a five-number summary (minimum value, lower quartile, median, upper quartile, and maximum value). These plots provide information about the outliers and their values and show whether the data are regularly distributed or skewed (Williamson et al., 1989). Ultimately, we explored snacking distribution over locations, days of the week, and time bins in the two datasets to compare and analyse snacking in both datasets.

3.4 Predictive models

At this stage, we focused on the predictive models that might support timed behaviour change interventions. We utilised classification models for predicting whether an unhealthy snack instance would occur in a time bin, predicting the property of the next snack as healthy or unhealthy (for the Dutch dataset), and estimating the (binned) number of unhealthy snacks instances in the next time bin. Furthermore, we applied ML models for predicting the time until the next unhealthy snack as a regression task. Spanakis et al. (2017) suggested that participants' demographics and individual status (such as: emotions, location, and activity), were effective predictors for general unhealthy eating in their study. Therefore, it is particularly interesting to focus predictive models only on location and day of the week, as we aimed for a minimal data collection procedure. Some of our models were additionally provided with the features of CR, UE, and EE, collected in the survey-based part of the study, to investigate whether they were effective predictors for unhealthy snacking behaviour or not.

As discussed in Chapter 2, previous studies have successfully applied a selection of ML models for modelling human behaviour and/or eating behaviour. We found that the promising models applied in previous studies can be divided into two types: fixed context and recurrent models. Recurrent models, such as Long Short-term Memory (LSTM), can process an input sequence of arbitrary length. In contrast, fixed context models require fixed input size, which can be achieved by windowing a longer sequence. Recurrent neural networks function in

cycles that feed the network activation from a previous time step, as a sequence of inputs to the network, to produce predictions at the current time step (Medsker and Jain, 1999).

We applied both fixed context models and recurrent models including, Feed Forward Neural Networks (FFNN), Extreme Gradient Boosting (XGBoost), Random Forest (RF) and Long short-term memory (LSTM). We also applied a Bayesian network (BN) model, as such models have been considered a powerful tool in psychology for understanding human cognition and behaviour (Tauber et al., 2017). In the modelling part we also divided the data into training and testing parts to estimate model performance to predict unseen situations. We tuned the hyper-parameters in all models to maximise generalisation.

3.5 Hypothetical interventions design and considerations

At this stage, we analysed the best time for delivering interventions based on psychological processes for behavioural decisions. We considered a more tailored analysis for such applications based on the explicit timing criteria desirable for a potential behaviour change intervention. We did not use intervention in the app in this project; we provided a time analysis, which is a basic requirement for a range of such interventions. We evaluated the usefulness of the developed predictive models for the timing of the interventions to reduce unhealthy snacking behaviour by testing whether predicted snacking times fell into time windows of 120, 60 and 30 minutes before the actual unhealthy snack time. In this experiment we used different time windows to analyse the feasibility of different intervention timings with our models.

Chapter 4

Snacking Behaviour Survey and Data Collection with the SnackTracker App

In this chapter we describe the methodology for the collection of our own dataset, the ‘UK dataset’, including survey design, app design, and ethics.

4.1 Ethics, privacy and security considerations

In order to run our study with human subjects, we had to address ethics, privacy, and security considerations. Initially, we applied for the ethical approval from the Research Ethics Committee of City, University of London’s Psychology Department (see appendix A). In order to fulfil ethics compliance, all participants signed consent forms to participate in the study and for their anonymous data to be made accessible for research purposes. Furthermore, on the first page of the online survey, participants were clearly informed about the background and objectives of the study. Additionally, they were notified that they may be selected to take part in the following part of the study (app-based part) for one month. Participants were informed that they were permitted to withdraw at any time, without giving a reason, and that all information and data provided would be anonymous and confidential.

Extensive efforts were made with the Data Protection office at City, University London (which is registered with the Information Commissioner’s Office, registration number Z8947127) to ensure that Personal Identifier Information (PII) were securely collected and participants’ privacy were not by any means violated to the best of our knowledge. We additionally had conversations with a company named Dev Technosys for implementing the app. Finally, we used Linode (a cloud hosting company) for providing virtual private servers, as well as Twillo, a company for providing programmable messaging services. These companies

are compliant with the General Data Protection Regulation (GDPR) (<https://gdpr-info.eu>) and use encrypted communication.

4.2 Survey design and execution

Our survey-based dataset was collected using a self-administered survey designed using Qualtrics (<https://www.qualtrics.com>) as a web-based survey tool. which is a web-based survey tool. The survey was run on Prolific Academic (<https://www.prolific.co>), a platform for finding candidates willing to participate in scientific studies, who can be selected according to various demographic criteria. The study was published with a target sample of 200 participants on 23 Dec 2019 and closed on 13 Feb 2020, when the full number had been reached.

Only UK citizens between the age of 18 and 60 were allowed to join the study. Participants were excluded from the study if they did not have a smartphone or were unwilling to participate in the follow-up study, i.e. the app-based part study. Furthermore, the app-based part involved sending a few reminders during the course of the study, detailed in Section 4.3. T-Mobile, O2, Vodafone, Three and EE were the only SIM card providers supported by the messaging service provider. Therefore, participants with other SIM card providers were excluded from the survey-based part of the study. Eventually, 16 participants were excluded and 184 participants completed the survey.

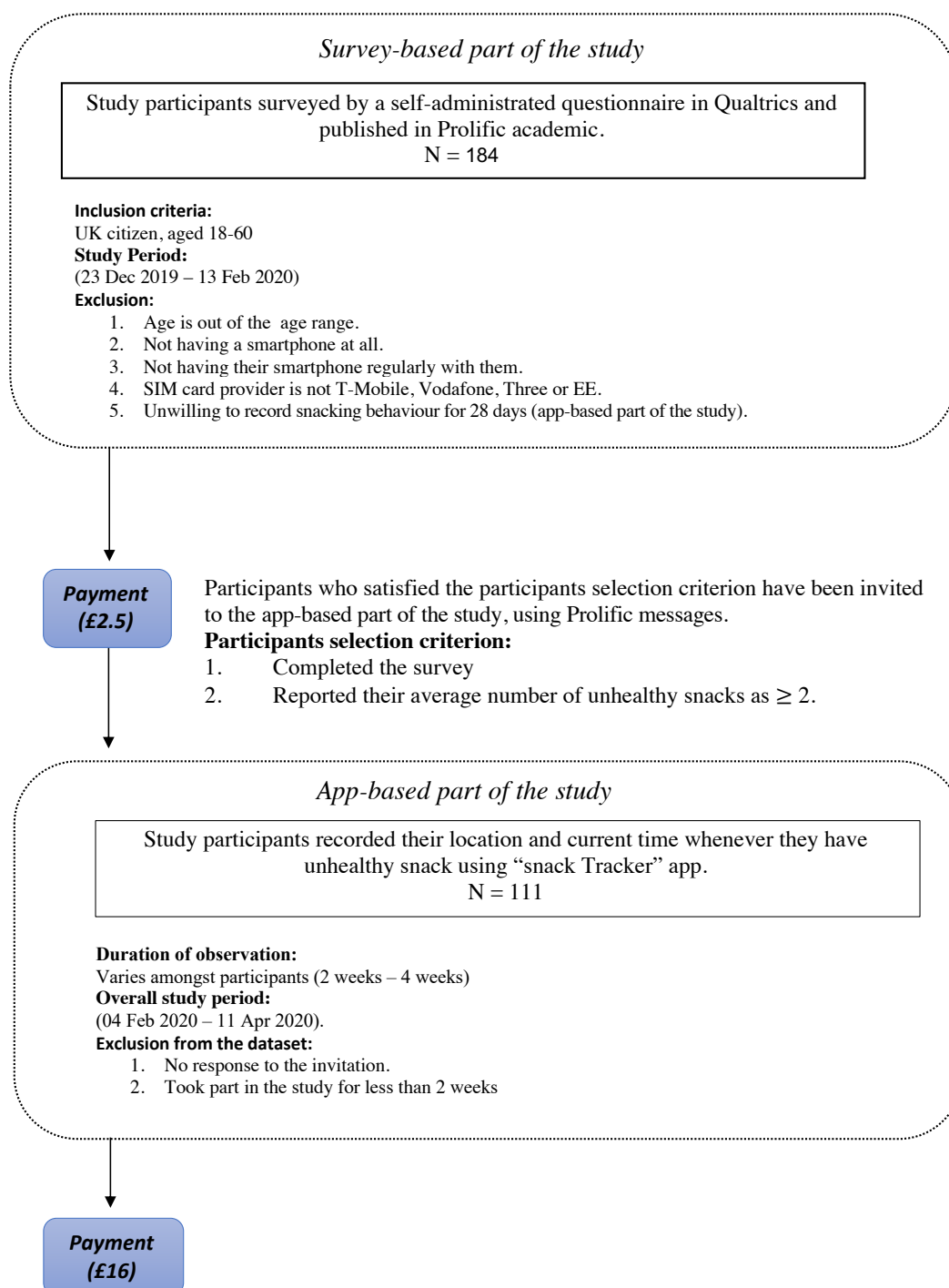


Fig. 4.1 Participant Flow in the two studies

4.2.1 Questionnaire

The survey questions were organised into three sections: demographics (Table 4.1), healthy eating motivation (Table 4.2) and Three-Factor Eating Questionnaire (TFEQ) (Table 4.3). Each of these sections were described to participants briefly as aiming to assess their healthy eating motivation or TFEQ.

Table 4.1 Demographics questions part of the survey

Question	Cardinality	Choices\value
What is your age?	-	18-60
What is your gender?	3	Male, Female, Other
Please indicate your ethnicity:	7	White
		Mixed/Multiple ethnic groups
		Asian/Asian British
		Black/African/Caribbean/Black British
		Other ethnic group
Please indicate your employment status:	5	Unknown
		Prefer not to say
		In paid employment
		Studying full time
		Unemployed Homemaker
What is your weight?	-	Retired
		Other (please specify)
What is your height?	-	-

All questions for the two sections shown in Tables 4.2 (motivation) and 4.3 (TFEQ) were in Likert scale format, except for the first two of the motivation section. The Likert scale is considered as one of the fundamental psychometric tools in social sciences research to measure attitude, perceptions, and opinions. It offers an ordered set of response categories (e.g. from strongly disagree to strongly agree) for participants to illustrate their level of agreement with a given statement (Joshi et al., 2015).

Table 4.2 Questions regarding motivation for healthy eating in the survey

Question	Cardinality	Choices\value
1. On a typical day, approximately how many sugary, salty or fatty snacks do you usually consume in between main meals?	-	-
2. Are you currently dieting to lose weight? (dieting)	3	Yes , No, I'd rather not say
3. It is important to me to watch my weight (IWW)	7	1-7 scale anchored by 'strongly disagree' and 'strongly agree'
4. It is important to me to eat a healthy diet (IEHD)	7	1-7 scale anchored by 'strongly disagree' and 'strongly agree'

The third part of the questionnaire, shown in Table 4.3, is based on the Three Factor Eating Questionnaire (TFEQ) (Cappelleri et al., 2009). As mentioned in Chapter 3, TFEQ is to assess the following three aspects of eating behaviour: Cognitive Restraint (CR), Uncontrolled Eating (UE), and Emotional Eating (EE). CR is defined as consciously restricting food consumption to control body weight management for the purpose of weight loss. UE refers to losing control in food consumption when feeling hungry or when external cues occur (e.g., delicious food). EE refers to the tendency to overeat in response to negative moods, such as anxiety or irritability (Aoun et al., 2019). In Table 4.3, questions 2,11,12,15,16 and 18 are related to CR while questions 3,6,10 are related to EE. The UE eating factor is associated with questions 1, 4, 5, 7, 8, 9, 13, 14, 17 .

Once a participant had completed the three sections of the questionnaire, he/she was approved in Prolific to receive a payment of £2.5.

Table 4.3 Three factor eating questionnaire (TFEQ) for restraint eating, emotional eating, external eating

Question	Cardinality	Choices\value
1. When I smell something delicious, I find it very difficult to keep from eating, even if I have just finished a meal.	4	Definitely true (4), mostly true (3), mostly false (2), definitely false (1)
2. I deliberately take small helpings as a means of controlling my weight.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
3. When I feel anxious, I find myself eating.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
4. Sometimes when I start eating, I just can't seem to stop.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
5. Being with someone who is eating often makes me hungry.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
6. When I feel blue, I often overeat.	4	Definitely true(4), mostly true(3), mostly false(2) definitely false(1)
7. When I see a real delicacy, I often get so hungry that I have to eat right away.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
8. I get so hungry that my stomach often seems like a bottomless pit.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
9. I am always hungry so it is hard for me to stop eating before I finish the food on my plate.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1),
10. When I feel lonely, I console myself by eating.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1),
11. I consciously hold back at meals in order not to weight gain.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
12. I do not eat some foods because they make me fat.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1),
13. I am always hungry enough to eat at any time.	4	Definitely true(4), mostly true(3), mostly false(2), definitely false(1)
14. How often do you feel hungry?	4	Only at meal times(1), sometimes between meals(2), often between meals(3), almost always(4)
15. How frequently do you avoid "stocking up" on tempting foods?	4	Almost never(1), seldom(2), usually(3), almost always(4)
16. How likely are you to consciously eat less than you want?	4	Unlikely(1), slightly likely(2), moderately likely(3), very likely(4)
17. Do you go on eating binges though you are not hungry?	4	Never(1), rarely(2), sometimes(3), at least once a week(4)
18. On a scale of 1 to 8, where 1 means no restraint in eating (eating whatever you want, whenever you want it) and 8 means total restraint (constantly limiting food intake and never "giving in"), what number would you give yourself?	4	1-8

4.3 App design and app-based data collection

4.3.1 The Snack Tracker app development and data flow

The frontend in this project is the app on the Android/iOS devices that users navigate and interact with. The app was designed with the goal of being easy to use with a simple sequence of screens. The screens are: splash screen 1, login screen 2, home screen 3, new snack screen 4, time recording screen 5 (time picker), location recording screen 6, recording save screen 7, and review recordings summary 8, as shown in Figures 4.3, 4.4 and 4.5. The current date and day of the week were automatically captured with each recording made, in order to minimize user effort in the data collection. The app works online allowing, users to log in and record the snacks they had. In case of connection loss, the app allows users to save their records and passes the data to the server once the mobile device is connected to the internet again.

In the backend, the database is for storing the data and user credentials, and the web based admin panel is for the admin to manage the project, and the web-based hosting server is for controlling project requests and operations from the app and the web-based admin panel. When participants accept the invitation to the app-based part of the study, I (as a project admin) would register participants manually using the admin panel. This was to prevent random users from using the app and ensure that the stored data are only from participants of our study. Other admin functions in the web admin panel included viewing, filtering, searching and downloading records from the database. The app-related operations and requests from users initially included logging in and viewing previous records. Additionally, after a snack record is created, users could select one of the options of save and record another snack, go back and edit, or just to save. The screen designs are shown in Figures 4.3, 4.4 and 4.5. The app sends the selected request to the server to execute it. The app development included designing and delivering messages as SMS and app push notification, in order to keep the users engaged and remind them to record snacks 4.8, in case they forgot to do so. The two methods of notification, MS and app push notification, were utilized to ensure that participants were notified even if their SIM cards or WiFi connections were lost.

4.3.2 Technical description of the app design

In app frontend development, React Native was used as an open source JavaScript framework for writing IOS and Android applications. All operations performed by app users and project admins were handled by Rest APIs created in Node JS (Server operations). Rest APIs were used to communicate with the database (MongoDB) for store and retrieve operations. The

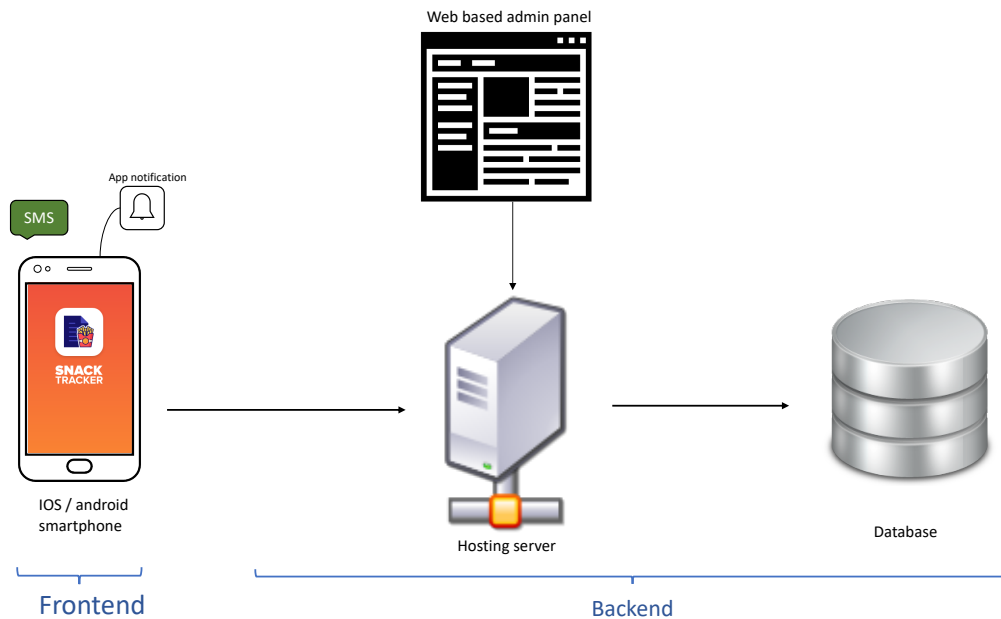


Fig. 4.2 data flow

server used in this project is a cloud-based server (Linode) that controls all operations and allows management of the application environment.

For automated SMS messages, described in the previous section, Twilio was used as a cloud communication tool allowing the software to programmatically send SMSs, utilising its web service APIs. For more information on this see <https://www.twilio.com>.

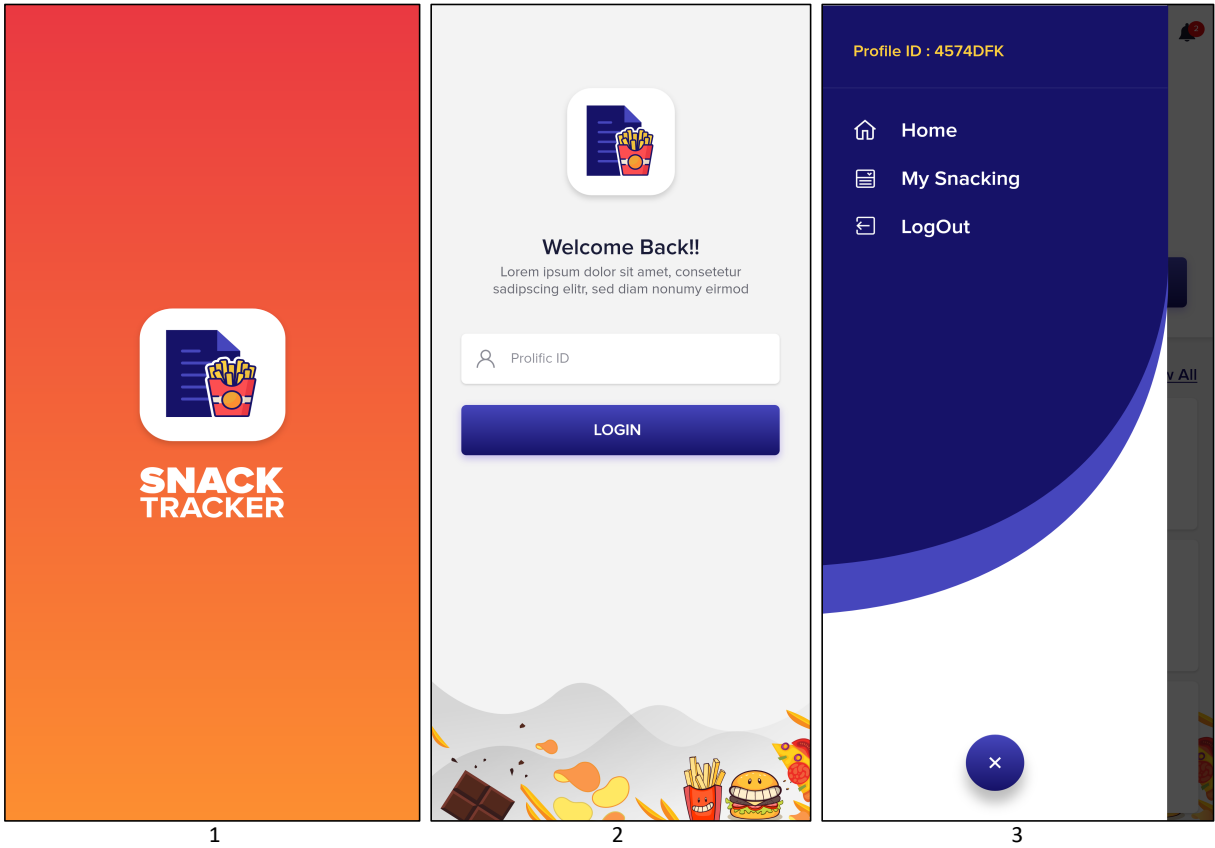


Fig. 4.3 App screenshots of splash screen1, login screen2, and home screen3

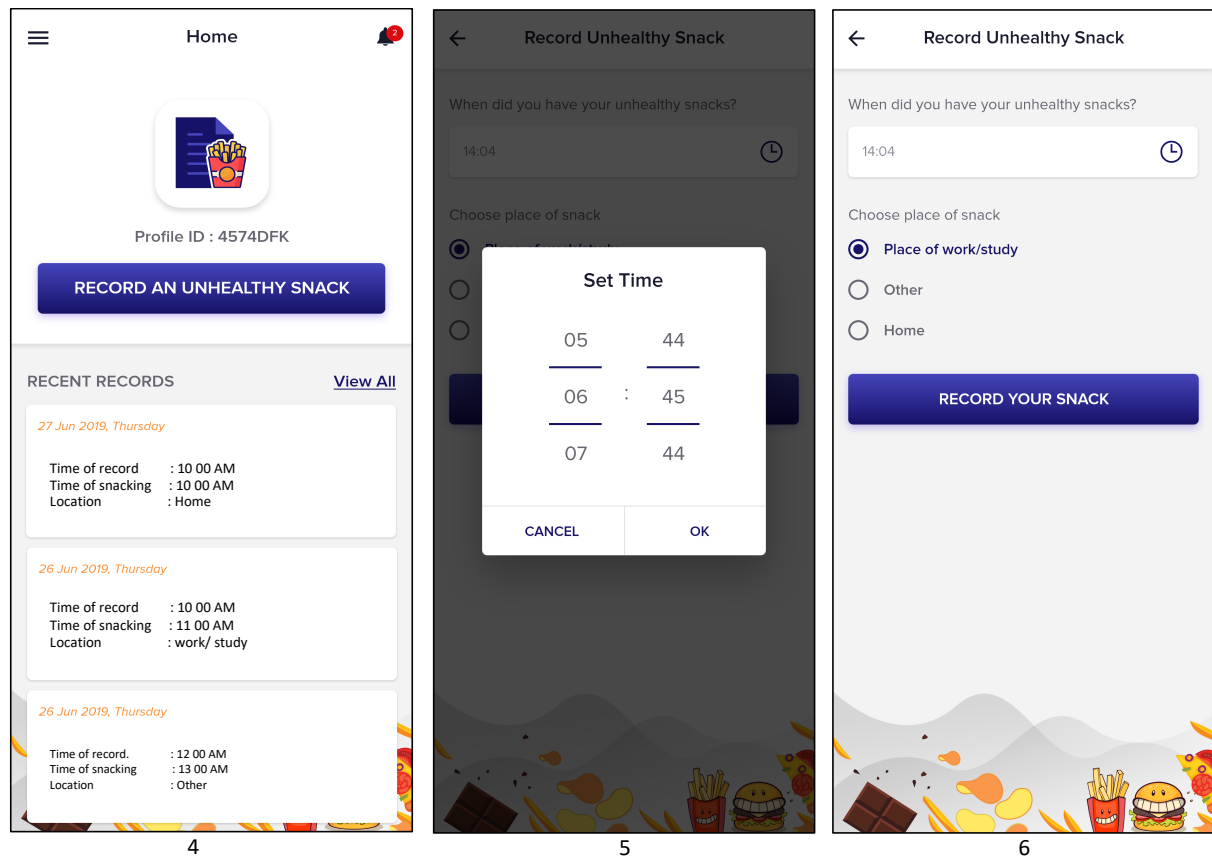


Fig. 4.4 App screenshots of new snack screen2, time recording screen5, and location recording screen 6

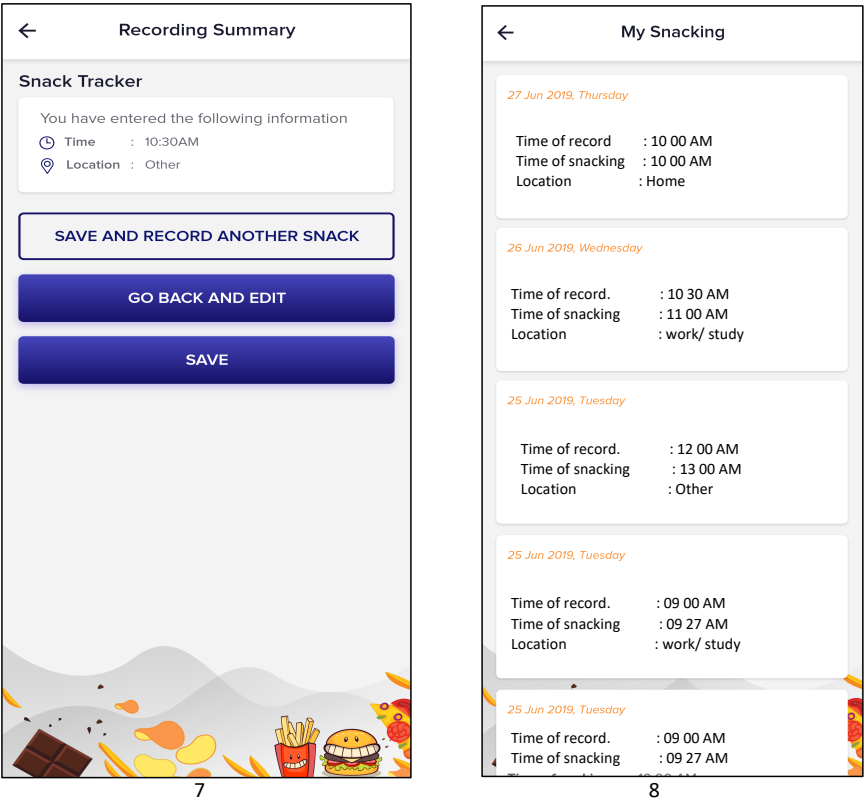


Fig. 4.5 App screenshots of recording save screen 7, and review recordings summary 8

4.3.3 The flow of the app-based part of the study

The app user’s flow for recording their time and location when they have unhealthy snacks is illustrated in Figure 4.6. As stated in Section 4.2, 184 participants completed the survey. The main criterion for selecting the participants for the app-based part of the study was an estimate that 2 or more unhealthy snacks are typically consumed per day. 170 of the participants satisfied this criterion.

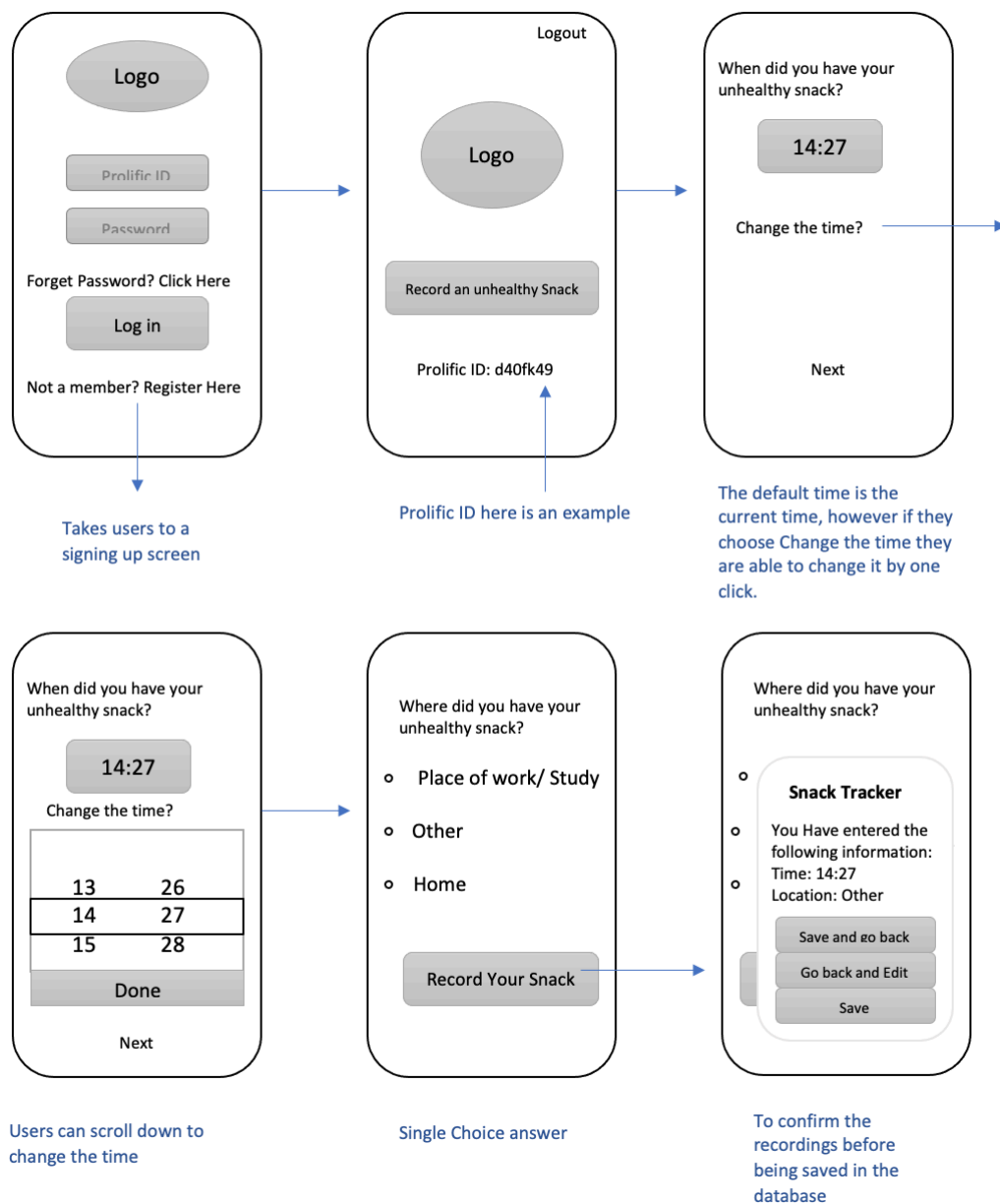


Fig. 4.6 App user flow

We initially decided to select 100 participants for the app-based part of the study. The Saudi Arabian Cultural Bureau (SACB) in London generously agreed to cover the payments of the participants. The invitation was sent in three groups and different periods of times in Feb and March 2020, as shown in Figure 4.7; different periods were utilized due to the limitation of available participants at initial recruitment periods. Participants received the

invitation via Prolific messages and when they accepted the invitation, the app installation link was then sent along with the app usage instructions.

Initially, 100 participants were invited to the study. However, only 68 accepted the invitation and completed the study, see Figure 4.7.

We then sent the invitation to another 45 participants to cover intended sample size of 100, of which 27 replied and completed the study. Out of the final group of 25 invitees, 16 completed the study. The total number of participants who accepted and completed the study was 111. The profile of participants in the sample is shown in Table 4.4. As shown in the table, the sample is somewhat unbalanced concerning different categories of age, Gender, employment status and Obesity level. Clearly, as the recruitment approach was random sampling amongst adult (18 plus) participants in Prolific Academic, based in the UK (for the first part of data collection), it would have been extremely unlikely that, by chance, participants would end up being balanced across such individual characteristics. As is common in behavioural sciences, we took into consideration such differences in statistical ones. For example, we merged the minority classes of Unemployed, Homemaker / Retired, Other into a one single class of Other in order to ensure that there is a sufficient number of instances in each category, see Section 5.4.

Table 4.4 Profile of participants in the sample. The total number of participants in the UK dataset was 100 (after excluding incomplete records).

Variable	Categories	UK	
		N	%
Gender	Female	34	31.78
	Male	73	68.22
Age group	18-24	28	26.17
	25-34	39	36.45
	35-44	28	26.17
	45-54	10	9.34
	55-60	2	1.87
Employment status	Employed	70	65.42
	Student	18	16.82
	Unemployed	10	9.35
	Homemaker/ retired	5	4.67
	Other	4	3.74
Obesity	Underweight	4	3.74
	Normal	43	40.19
	Overweight	20	18.69
	Obese	40	37.38

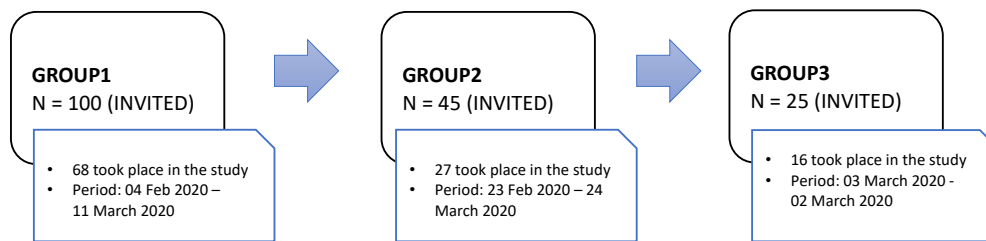


Fig. 4.7 Three groups of participants in the app-based part of the study

Participants were instructed to take part in the study for 28 days. However, exact start and end dates of the study differed amongst participants as would be expected. Also, some participants did not stay on the study for the full duration.

Participants who took part in the study for less than 1 week were excluded from the dataset. The majority of participants recorded their unhealthy snacks for the full 28 days. A few participants continued recording their snacks for an additional 6-11 days after reaching their 28 day and even after a completion message had been sent.

A follow-up message was sent to participants in Prolific under some conditions:

- If they had at least one recording a day or they had no recordings in any day.
- If they had multiple recordings (2 or more) on every day for a period, but then the number of recordings abruptly decreased.

This was to ensure that there were no technical issues/ problems with any of the tools employed in the study, as outlined in Section 4.3.2, which might prevent records to be stored in the database. Additionally, a few random messages were sent via Prolific messages in order to check if participants were receiving the study reminders detailed in Figure 4.8. Responses from participants did not indicate that there was an issue in any of the tools used in this part of the study. Additionally, we received some encouraging and positive comments about the study in general. Some mentioned that recording their unhealthy snacks made them aware of the amount of unhealthy snacks consumed daily and helped them to reduce them, see appendix B for participants comments.

Participants were asked to record their location whenever they were having an unhealthy snack. The current date and day of the week (i.e. Monday, Tuesday .. etc) were automatically stored with each recording made. The total number of records made by participants was 5391.

Participants received 3 kinds of notifications throughout the study as an app notifications and SMS messages (see 4.8). The two forms of notification were used to ensure that participants were notified in case of losing SIM card or WiFi connections. Daily reminders were received at 7 pm reminding them to record any forgotten unhealthy snacks. Additionally, a weekly notification was sent at the end of each week in order to keep participants engaged especially towards the end of the study. If a weekly notification was sent, no daily reminder was also sent on that day, to ensure that participants did not receive too many messages.

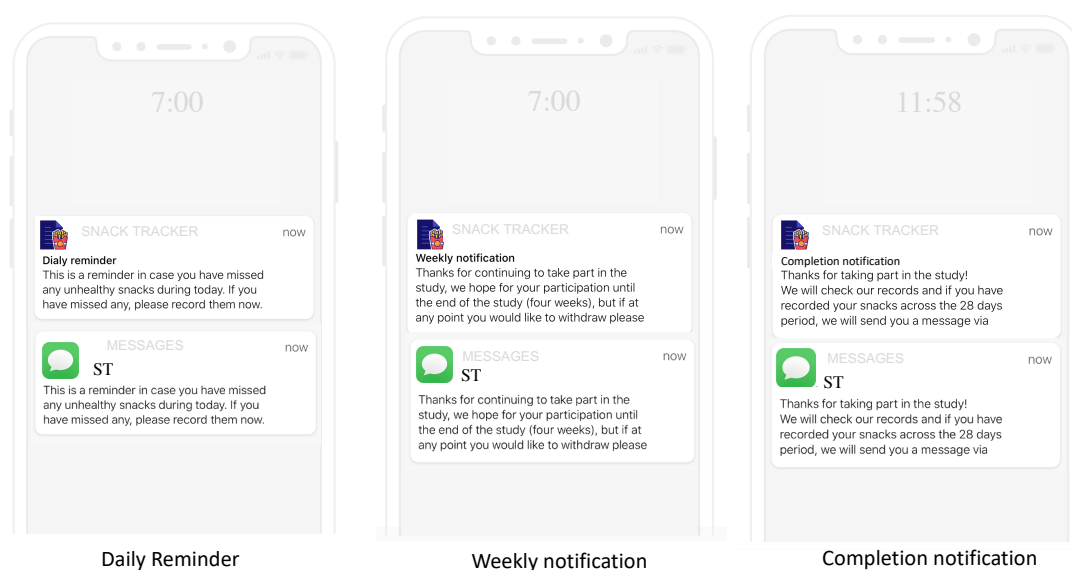


Fig. 4.8 App reminders

Once participants reached the end of the 28 day period from their first app record, they were sent a completion notification to thank them for their participation in the study, as shown in Figure 4.8. In addition, the completion message was to notify them that the study completion code was to be sent to them using Prolific messages. This would happen after we had checked our records and made sure that they had made the required recordings via the app. Then we would administrate payment via Prolific (£16).

4.4 Discussion

In this chapter we described the survey, app design and implementation, and app-based data collection, which was the basis for modelling snacking behaviour in the following chapters. Our “Snack Tracker” app was designed with the goal of gathering snacking data along with time and location. All tools used to collect our UK dataset met the General Data Protection

Regulation (GDPR) and City, University London's ethical, privacy, and security requirements. The dataset was collected in the United Kingdom (hence we call it the 'UK dataset').

The data collection consisted of two parts: a survey and an app-based study. Initially, 200 participants between the ages of 18 and 60 took part in a self-administered survey that covered demographics, healthy eating motivation, and the Three-Factor Eating Questionnaire (TFEQ). 184 participants completed the survey, and 170 of them were selected and invited to the study's app-based part of the study.

We designed our "Snack Tracker" app to track snacking behaviour. We ran the experiment, using the app as a frontend and custom backend system to collect data and send messages.

Participants used the app for four weeks to track their unhealthy snacking; their data forms our 'UK dataset'. The study was fully completed by 111 participants in three groups at different times. Most of the messages and feedback received from participants in the survey-based or app-based parts were positive about their participation in the study. This is encouraging about the method, which we may use for future data collection and perhaps also extend with timed message delivery based on our models.

Chapter 5

Snacking Datasets Pre-processing and Descriptive Analysis

This chapter covers the pre-processing techniques employed on our UK dataset and the Dutch dataset that we obtained from another research group. We then report the descriptive analyses, correlations, and some notable patterns for both datasets. All data pre-processing steps and data analyses discussed in this chapter were conducted using the Python programming language (van Rossum and Drake Jr, 1995) and several relevant packages, particularly SciPy (<https://www.scipy.org>) and scikit-learn (<https://scikit-learn.org>).

5.1 The Dutch dataset

Prof Anne Roefs and colleagues at Maastricht University's Faculty of Psychology and Neuroscience provided us with another relevant dataset (the 'Dutch dataset'). The dataset was collected with another app called "Think Slim" in the Netherlands. In their study, around 15 variables were used in a data collection exercise, which involved several daily measurements, in order to assess participants mood, emotion, and activity, along with the participant's current location (Spanakis et al., 2017). Their focus was classifying food (including meals and snacks) and cluster participants, in order to provide adaptive feedback aiming to enhance eating behaviour (Spanakis et al., 2017). Since the topic of this project is unhealthy snacking behaviour, we extracted measurements corresponding to unhealthy snacks from the more general food data in the original Dutch dataset. The location categories were reduced from 950 categories (entered as free text) to the 3 categories of home, study/work, and other. Further pre-processing steps are detailed below.

5.2 Data cleaning

5.2.1 UK dataset

Data cleaning for the UK dataset included the following steps:

- Removing rows with missing entries in the survey-based dataset and 413 duplicate entries in the app-based dataset.
- Homogenising measurement units, so that for weight we employed lb and height inches.
- Encoding Likert-scale responses (i. e. questions 3-4 in Table 4.2 and 1-18 in Table 4.3) as shown in Table 5.1. For the Three Eating Factors (TEF) questions, scores of the questions corresponding to each factor were aggregated for each participant, in order to find the participant's overall score for the various factors measured by the TEF: these were Cognitive Restraint (CR), Uncontrolled Eating (UE), and Emotional Eating (EE). The coding guide and score aggregation was based on the method employed in Karlsson et al (Karlsson et al., 2000). Questions 3-4 in Table 4.2 formed Importance of Watching Weight (IWW), and Importance of a Healthy Diet (IEHD) ordinal variables.

Table 5.1 Coding guide of Likert-scales related questions

Likert point	Score
1,2	1
3,4	2
5,6	3
7,8	4

5.2.2 Dutch dataset

Data cleaning for the Dutch included the following steps:

- Translation of the data from Dutch to the English language.
- Manual classification of 950 different locations, entered as free texts, into the 3 locations of 'home', 'place of study or work' and 'other'.
- Extracting snacks from general food recorded in the dataset.

5.3 Feature engineering

(Nargesian et al., 2017). (Nargesian et al., 2017). Table 5.2, shows the new features created in the datasets: Average Daily number of Unhealthy Snacks (ADUS), Age groups, Body Mass Index (BMI), Obesity level, Snacking healthiness, Number of unhealthy snacks with a description, range of values, and rationale of the features. Note that the equation below has been used in creating the feature BMI.

$$BMI = Weight(kg)/height^2(m) \quad (5.1)$$

Table 5.2 Description of the new features created in the UK and Dutch datasets based on the collected features

Feature	Dataset/s	Range of values	Description/ steps of creation	Rationale
Average Daily number of Unhealthy Snacks (ADUS)	Both datasets	1-8 in the UK and 0-3 in the Dutch dataset.	1. Filtering participant records. 2. Aggregating the number of unhealthy snacks recorded across all days. 3. Averaging per participant.	For facilitating the analysis of ADUS and different demographics in Section 5.6.
Age groups	Both datasets	18-24, 25-34, 35-44, 45-54 and 55-60.	The age groups were employed for consistency with the national statistical report on obesity, Physical Activity and Diet in England, NHS (2020).	For analysing the relationship between age and ADUS in Section 5.6 and snacking predictions in Sections 6.4.1 - 6.4.4.
Body Mass Index (BMI)	UK dataset, was already available in the Dutch dataset.	20-47	Calculated using the Equation 5.1.	For analysing the relationship between obesity and ADUS in Section 5.6.

Continued on next page

Feature	Dataset/s	Range of values	Description/ steps of creation	Rationale
Obesity level	Both datasets	Underweight (<18), normal (18.5-24), overweight (24.5-29.9) and obese (> 30) as suggested in the literature.	Created from the BMI feature.	For analysing the obesity levels and the relationship between obesity and ADUS in Section 5.6 and Figure 5.4.
Snacking healthiness	Dutch dataset, no information on healthy snacks was collected in the UK dataset.	Binary (yes/ no)	Created Manually with the classification of snacks as healthy or unhealthy. This was done in consultation with Dr. Katy Tapper from the Psychology Department at City, University London.	For predicting whether the next snack is healthy or unhealthy in Section 6.4.2.
Number of unhealthy snacks	Both datasets	0,1,2 and 3+	The number of snacks consumed by a participant in a specific time bin on a specific day of the study.	For the predictive model in Section 6.4.4.

5.4 Data transformations

Categories with few instances have been merged to ensure that there is a sufficient number of instances in each category. For instance, in Table 5.3 Unemployed, Homemaker / Retired, Other have few instances in both datasets, therefore, these categories were combined into a single category Other. Furthermore, the Number of Unhealthy Snacks per time bin was rarely more than 3. Therefore, regarding snacking frequency, 3 and more than 3 observations were combined into one class in the feature Number of Unhealthy Snacks. Data preprocessing also involves transforming or encoding data into a representation that machine learning (ML) models can more easily process (Iliou et al., 2015). Many ML models require numeric input

variables. Ordinal variables have a natural ordering of their levels, whereas nominal variables do not have an ordering relationship between their levels (Cerdeira et al., 2018). For the latter, one-hot encoding is commonly used in ML, in order to address the non-ordered nature in nominal variables. This type of encoding requires converting each category of a variable to a separate variable taking only the value of 0 to or 1 to indicate the presence of the category (Berry et al., 1998). Nominal variables in both datasets (i.e. gender, ethnicity, employment status, dieting, day of the week, location and time bin) were converted into one-hot encoding, while ordinal (i.e. Importance of Watching Weight (IWW) and Importance of a Healthy Diet (IEHD)) and continuous variables (e.g. Age, weight .. etc) were kept in their numeric format, because of their ordered nature. The time of snacks was treated as a continuous variable in most related analyses. However, in Bayesian Networks (BNs), continuous variables cannot be directly employed, because a standard BNs is based on probability theory involving conditional probabilities amongst the discrete states of categorical variables (Fang et al., 2017). Hence, Time was discretised in BNs in Section 6.4.1.

All input variables collected and created, as described in Tables 4.4, 5.2 and 5.3 were scaled in order to ensure a unified distribution for all input variables. In addition, feature scaling is required in Neural Networks (NN), as this is an essential step in the calculation of gradient descent (Wan, 2019). We applied standardisation before FFNN models using Equation 5.2, while we applied normalisation before LSTM models using Equation 5.3, as we noticed that normalisation slightly enhanced the LSTM model's accuracy.

$$Z = (X_i - \text{Mean}) / SD \quad (5.2)$$

Where Z is the standard score for Standardization, X_i is the observed value in a variable X, and SD is the standard deviation.

$$Y = (X_i - \min(X)) / (\max(X) - \min(X)) \quad (5.3)$$

Where X_i is the observed value in a variable X, Y is a scaled X_i .

5.5 Oversampling of imbalanced classes

SMOTE was employed as a solution to class imbalances in some of the tasks in Chapters 6 and 7, specifically in Sections 6.4.4 and 6.4.3.

5.6 Descriptive analysis

This section reports descriptive statistics and correlation analysis for the Dutch and UK datasets. (Williamson et al., 1989). In Figures 5.1 – 5.5, the box shows the inter-quartile range (IQR) as usual and then the lower and upper whiskers are coded as follows:

$$Lower = Q1 - 1.5 * IQR \quad (5.4)$$

$$Upper = Q3 + 1.5 * IQR \quad (5.5)$$

Then the outliers, i.e. data points below the lower or higher than the Upper bound, are shown as small circles.

Table 5.3 Demographic information in the Dutch dataset. All frequencies in this table represent unhealthy snacking. Unhealthy snack events were extracted from generic eating information contained in the data. The total number of participants in the Dutch dataset was 196. Records with missing values were excluded from any processing.

Variable	Categories	Dutch	
		N	%
Gender	Female	165	84.18
	Male	31	15.82
Age group	18-24	39	19.9
	25-34	46	23.47
	35-44	33	16.84
	45-54	61	31.12
	55-60	17	8.67
Employment status	Employed	125	63.78
	Student	37	18.88
	Unemployed	16	8.16
	Homemaker/ retired	18	9.18
	Other	0	0.00
Obesity	Underweight	0	0.00
	Normal	43	21.94
	Overweight	77	39.29
	Obese	76	38.78

In Table 5.3, we see that in the UK dataset male participants outnumber female participants roughly 2 to 1, whereas in the Dutch dataset, less than 1 in 6 participants were male. In Figure 5.1, we see that in the UK dataset, male participants most frequently consumed 3 unhealthy snacks per day, while female participants most frequently consumed 2. Moreover, there was more variation in female participants who reached up to 6 daily unhealthy snacks.

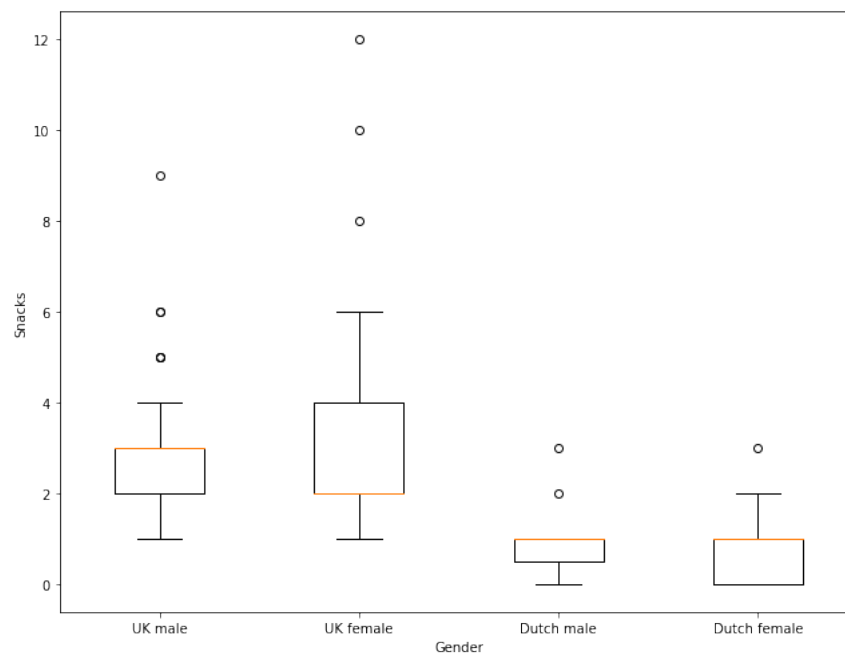


Fig. 5.1 App-based Average Daily number of Unhealthy Snacks (ADUS) for different genders in the two datasets.

The lower variation in male participants' unhealthy snacks consumption can be also observed in the Dutch dataset.

In the Dutch dataset, most participants were in the 45-54 years old age group, whereas in the UK dataset most participants were in the age group 25-34, see Table 5.3. The mean age of participants in the UK and Dutch datasets were 32.63 and 37.91 respectively.

Figure 5.2 illustrates that the Average Daily Unhealthy Snacks (ADUS) in the age group of 35-44 was most frequently 3, whereas it was typically lower in other age groups in the UK dataset. In the Dutch dataset, the ADUS was typically 1 in all age groups. There was more variation in the average daily number of unhealthy snacks amongst participants within the same age group in the UK dataset, compared to participants in the same age groups in the Dutch dataset. Overall, this picture is somewhat similar to the finding in the Piernas and Popkin study, that participants in the middle-aged group (40-59) accounted for the highest number of daily snacks compared to other age groups (19 years old and older; Piernas and Popkin (2010). Tables 5.4 and 5.5 show that there is no strong correlation between age and ADUS, hence the variability in the UK age groups may not have been associated with unhealthy snacking.

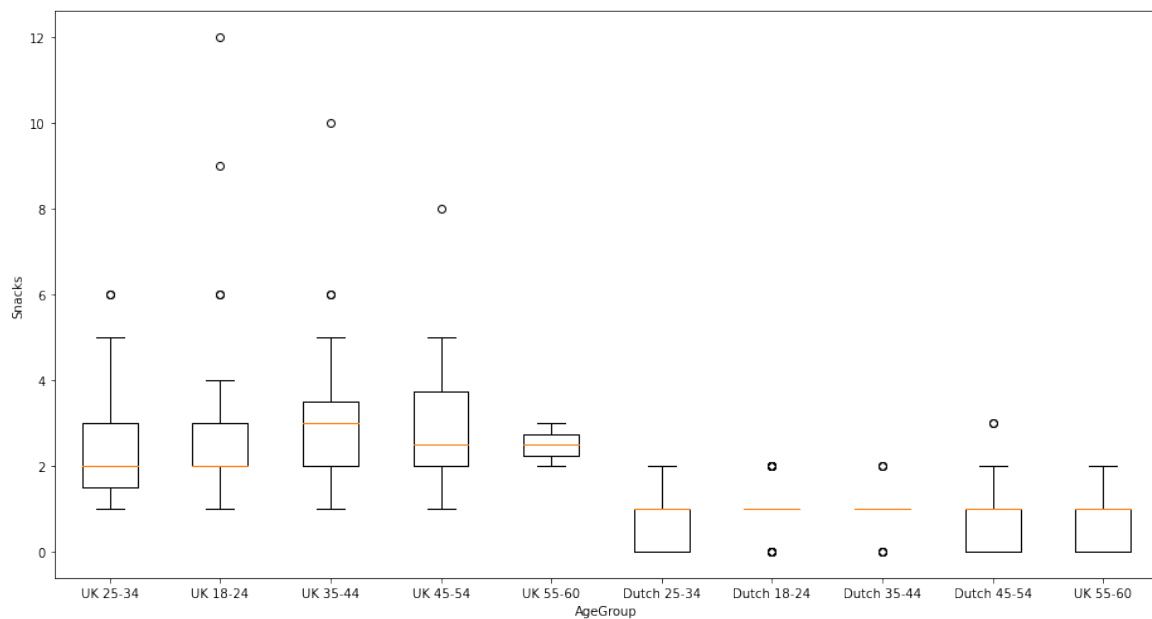


Fig. 5.2 App-based Average Daily number of Unhealthy Snacks (ADUS) for different age groups in both datasets.

Table 5.4 Pearson correlation coefficients between the various variables, for the UK dataset, and p-values (2-tailed), N = 107. * Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.01 level. SADUS is Survey-based Average Daily number of Unhealthy Snacks. AADUS is app-based Average Daily number of Unhealthy Snacks.

	Age	BMI	SADUS	CR	UE	EE	AADUS
BMI	0.19*	1					
p	0.04						
SADUS	0.05	0.12	1				
p	0.58	0.24					
CR	0.01	-.12	-.28**	1			
p	0.91	0.22	0.002				
UE	-.05	0.34**	0.30**	-.32**	1		
p	0.60	0.00	0.00	0.00			
EE	0.00	0.39**	-.01	-.15	0.47**	1	
p	0.96	0.00	0.93	0.13	0.00		
AADUS	0.06	-.10	0.27**	-.15	0.18	0.05	1
p	0.56	0.32	0.01	0.12	0.07	0.58	

Table 5.5 Pearson correlation coefficients for the Dutch dataset and p-values (2-tailed), N = 196. ** correlation is significant at the 0.01 level. AADUS is app-based Average Daily Unhealthy Snacks.

	Age	BMI_pre	BMI_post	AADUS
BMI_pre	0.284**	1		
p	0.00			
BMI_post	0.264**	0.991**	1	
p	0.00	0.00		
AADUS	-.123	-.133	-.122	1
p	0.085	0.063	0.089	

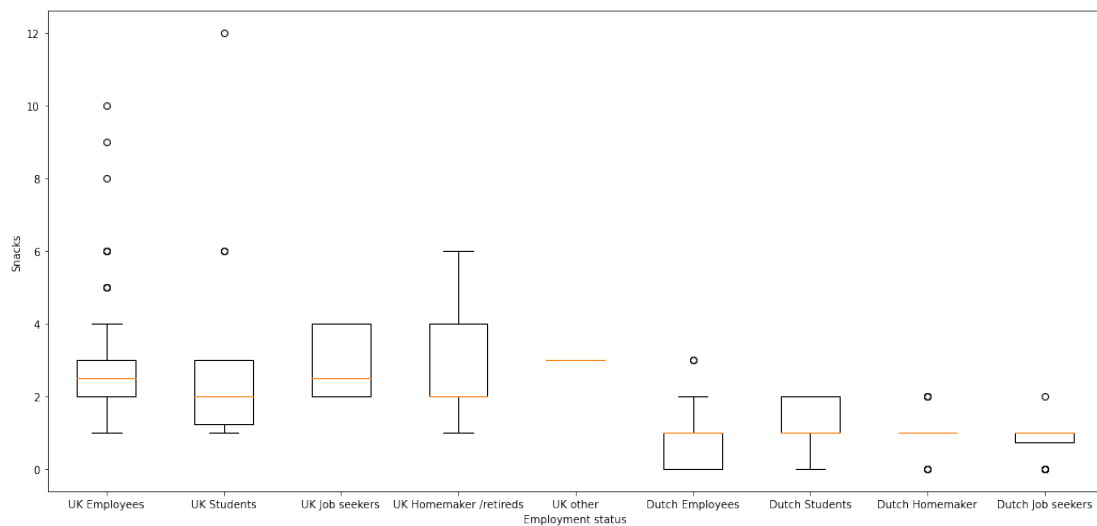


Fig. 5.3 Daily unhealthy snacks for different Employment status in the two datasets.

Concerning employment, as shown in Table 5.3, most participants were employed, in both datasets. Furthermore, the app-based ADUS variable was commonly the highest amongst employees and job seekers in the UK dataset, as shown in Figure 5.3. By contrast, results from the (Almoraie et al., 2021) study suggest that unhealthy snacking is associated more with students and employees, as the stress and lack of time were considered obstacles to a healthy dietary pattern in these groups. There was significant variation amongst UK home keepers/ retired individuals in the app-based ADUS variable, compared to participants in other categories regarding the employment status variable. On the other hand, no significant variation could be observed amongst different levels of the employment status variable in the Dutch dataset.

Regarding BMI, UK participants were more so in the ‘normal’ BMI category or obese, while Dutch participants were more overweight and obese, as shown in Table 5.3. In Figure 5.4, we can see that obese and overweight participants more often consumed higher unhealthy snacks in the UK dataset. In the Dutch dataset, participants across all levels of obesity more often consumed 1 daily unhealthy snack. Furthermore, there was no correlations between BMI and app-based ADUS in both countries. Rates of consuming unhealthy snacks in the UK dataset are consistent with demonstrations of a relationship between BMI and unhealthy food consumed, along with lower levels of physical activity Shah et al. (2014).

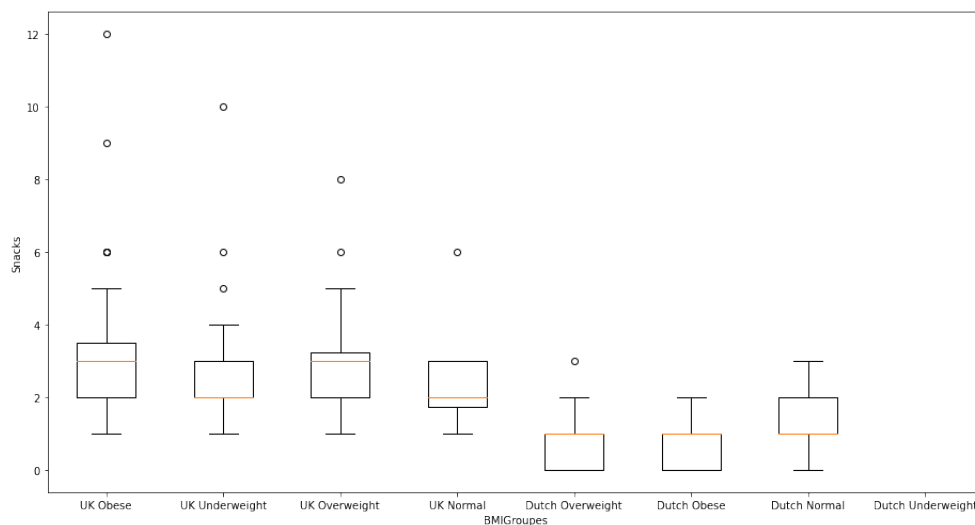


Fig. 5.4 App-based ADUS for different obesity levels in the two datasets.

Regarding ethnicity, amongst UK participants, 77.6 of them were white, while the rest of the participants represented other ethnicities such as Asian, Black and mixed. Around 77% of the UK participants were not dieting; 70% of them stated that they believe it is important to watch their weight; and around 75.7% of them stated that they believe it is important to eat a healthy diet.

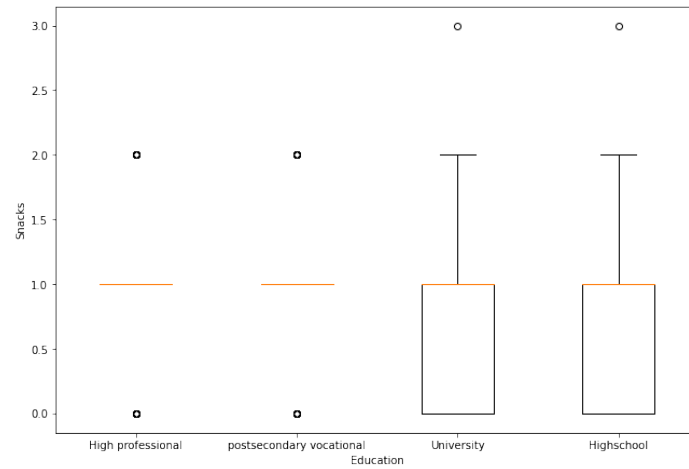


Fig. 5.5 App-based ADUS for education groups in the Dutch dataset.

In the Dutch dataset, the number of participants in different education groups ranged from 33.67% having High professional education to 17.86% having only High school education. Figure 5.5 demonstrates that there were no considerable variations of app-based ADUS between education groups in the Dutch dataset. Note, Finger et al. (2013) found a positive association between education level and unhealthy food consumption.

The mean of UE was the highest, compared to the CR and EE. Moreover, no association was observed between TEF variables (i.e. CR, EE and UE) and app-based ADUS, while CR and UE were moderately correlated with the survey-based ADUS. Note, we would expect Banna et al. (2018), we would expect a positive correlation between CR and BMI, as well as a positive correlation between UE and both EE and a variable concerning food left over from dinner.

Table 5.6 CR, UE and EE in the UK dataset

Variable	Mean	SD	Max	Min
CR	8.39	1.99	13	6
UE	14.98	2.67	18	9
EE	4.94	1.20	6	3

Table 5.7 Unhealthy snacking distribution over locations, days of week, and time bins in the two datasets. Healthy snacking instances were excluded from the Dutch dataset. The total number of unhealthy snacks in the UK and Dutch dataset are 4979 and 3297, respectively.

Variable	Categories	UK		Dutch	
		<i>N</i>	%	<i>N</i>	%
Location	Home	3393	68.15	2125	64.45
	Work/ study	992	19.92	517	15.68
	Other	594	11.93	655	19.87
Day	Monday	719	14.44	485	14.71
	Tuesday	714	14.34	481	14.59
	Wednesday	813	16.33	478	14.50
	Thursday	741	14.88	477	14.47
	Friday	663	13.32	508	15.41
	Saturday	589	11.83	459	13.92
	Sunday	740	14.86	409	12.41
	Early morning	248	4.98	30	0.91
	Morning	934	18.76	859	26.05
Time bin	Afternoon	2090	41.98	1338	40.58
	Evening	1707	34.28	1070	32.45

Home was the most frequent location for unhealthy snacking in both datasets, as shown in Table 5.7. Dunneram et al. (2013) found an association between low physical activity level and the consumption of unhealthy foods. The high frequency of home location for unhealthy snacks may be due to the low physical activity at home. The number of consumed unhealthy snacks in early morning was higher in the UK dataset than in the Dutch dataset, see Table 5.7. This might be due to the higher rate of younger participants in the UK dataset. Other trends are that participants snacked more towards the afternoon, then unhealthy snacking decreased in evenings and during the night period, in both datasets. The mean frequencies of unhealthy snacking in the UK and Dutch datasets were relatively similar (15.34 and 15.68 respectively), as shown in Table 5.8.

Table 5.8 Time of snacking in the UK and Dutch dataset

	Min	Max	Mean
UK dataset	00:00	23:97	15:34
Dutch dataset	00:03	23:82	15:68

In both datasets, unhealthy snacks were consumed relatively equally in all week days, Saturday and Sunday were the days with the lowest number of snacks in the UK and Dutch dataset respectively. This is in line with the findings of Bochud et al. (2015) that there are similar patterns of snacking during weekdays and weekends.

In the Dutch dataset (see appendix C), fruits were the most frequent snack. This could be an artifact of data collection in the Dutch study, because there was no specification of fruit type, whereas for snacks there were 17 different types (e.g. candy, fries, etc).

The total number of unhealthy snacks in the UK was 3297, whereas the total number of snacks (including healthy and unhealthy snacks) recorded in the Dutch dataset was 3297. This indicates that the unhealthy snacks recorded in the UK dataset were considerably more than the number of unhealthy snacks recorded in the Dutch dataset. This may well be due to the participant selection criteria in the UK study (i.e. selecting participants who reported more than two unhealthy snacks daily, as reported in the UK survey-based part of the study). In contrast, participants were free to participate in the Dutch study without a prior assessment of their snacking behaviour.

Table 5.9 Mean of snacks in the UK and Dutch dataset. ADUS is Average Daily number of Unhealthy Snacks

Variables	N	Min	Max	Mean	SD
Survey-based ADUS	107	2	12	3.83	1.77
App-based ADUS	107	1	8	2.4	0.88
App-based ADUS	196	0	3	0.93	0.69

As shown in Table 5.9, the mean of the UK Survey-based ADUS was 3.83 (SD=1.77, range:0-12) while the mean of UK app-based ADUS was 2.4 (SD=0.88, range: 2-8). This shows a remarkable difference between the estimated and recorded numbers of unhealthy snacks. This may indicate that people overestimate their snacking or that participants sometimes fail to record their unhealthy snacks. Another possible reason for the difference between the UK Survey-based and app-based ADUS is that tracking the snacks resulted in reducing the number of unhealthy snacks consumed towards the end of the study. This was noted by some UK participants, who reported that recording their daily unhealthy snacks made them aware of the number of unhealthy snacks they typically consume, see appendix B. This reasoning agrees with the findings from other studies that tracking behaviour can assist in behaviour change (Jin et al., 2022). App-based ADUS in the Dutch dataset was 0.93 (SD=0.69, range: 0-3). The criteria set when collecting the UK dataset, discussed earlier in this chapter, may explain the higher snack means in the UK study.

The increasing tendency to use technologies, such as mobile apps and wearable systems, leads to more common use of self-tracking technologies, that may help to change behaviour, e.g. in order to fulfil fitness goals (Asimakopoulos et al., 2017). The way weekly unhealthy snacking changed with them was analysed in the Tables presented in appendices D and E and Table 5.10 in order to address the effect of self-tracking, in changing the behaviour. 41% of UK participants had their weekly number of unhealthy snacks reduced by the end of the study in the UK dataset, while 30% participants had their weekly number of unhealthy snacks reduced in the Dutch dataset. The percentages of participants who had their weekly average number of unhealthy snacks increased was 29% and 20% in the UK and Dutch datasets respectively. 30% of the UK participants and 50% of the Dutch participants had their weekly average number of unhealthy snacks unchanged over the course of the study.

Table 5.10 Trend of Snacking. 28 participants with only 1 week of recording were excluded from Dutch progress table.

		decreased	unchanged	increased	Total
UK	<i>N</i>	44	32	31	107
	<i>%</i>	41	30	29	100
Dutch	<i>N</i>	51	84	33	168
	<i>%</i>	30	50	20	100

Generally, these findings are far from conclusion since the number of participants is not big in both datasets.

5.7 Discussion

This chapter covered the pre-processing steps on the UK dataset that we collected and the Dutch dataset provided by Maastricht university. The pre-processing includes data cleaning, translation to English language (for the Dutch dataset), encoding survey questions into scores, one-hot encoding, features engineering and scaling, and over-sampling. Additionally, it includes the extraction of snacks from the general food categories recorded in the Dutch dataset. The datasets were then analysed for understanding the behaviour in both datasets and with a view to identify general patterns and trends. The descriptive analyses showed many different patterns and trends, with more variation in the UK dataset compared to the Dutch dataset. Female participants in both datasets showed more variation in their number daily unhealthy snacks than male participants. Some studies in the literature, for instance, Cleobury and Tapper (2014); Renner et al. (2012); Verhoeven et al. (2015), have found that

younger people consumes more unhealthy snacks than older people. In addition, they found that women reported higher rates of unhealthy snacking than men when dealing with negative emotions. We have not found such correlations in our data, and this could be due to the particular characteristics of our datasets. Gender and age can be significant predictors in various contexts (e.g., Kant and Graubard (2015)), (Murakami and Livingstone (2016)) (Liu et al. (2017)), (Dunford and Popkin (2017)), (Genena and Salama (2017)). Concerning the specific behaviour of interest, unhealthy snacking, our conclusions do not support a role of these variables in predictive models. Of course, our conclusions are restricted by the limited time our study was carried out and the particular characteristics of our sample. Employees were over-represented in both datasets compared to participants who were in any other employment status (e.g. students or home/keepers). Employed, job seekers, over-weight, and obese more commonly consumed more unhealthy snacks than other demographic groups in the UK dataset.

Home was the most frequent location when unhealthy snacks were consumed. This may have been due to the lower level of physical activity expected at home. Participants were found to snack unhealthily more so in the afternoon than at other day times with no variation in app-based ADUS between weekdays and weekends, in both datasets. Average unhealthy snacks per day per person reported in the UK dataset were much higher than in the Dutch dataset. This may well be due to the study inclusion criteria for participants in the UK study, since we sought participants who already consume a (relatively speaking) high number of unhealthy snacks. Finally, tracking behaviour itself can raise awareness of unhealthy snacking and subsequently lead to behaviour change, as has been found in the daily and weekly analysis of unhealthy snacking in both datasets.

These findings were important because they helped provide a comprehensive understanding of the data and offered some insight into the shape and distribution of each feature before applying predictive models in Chapter 6. For instance, in Table 5.3, unemployed, homemaker / retired, and other have few instances in both datasets. These categories were combined into a single category named “other” before applying the models in Sections 6.4.1 to 6.4.3. Additionally, the analysis has helped us answer the questions below, using our predictive models in Chapter 6:

1. Because there are more variations in the snacking behaviour in the UK dataset compared to the behaviour in the Dutch dataset, the question arises whether the performance of the predictive models may vary across the two datasets.
2. Given that participants snacked unhealthily more in the afternoon than other day times, the question arises whether the models could predict unhealthy snacking occurrence in other time slots, given sparse data for these other time slots.

3. Because there were more instances of unhealthy snacking in the UK dataset compared to the Dutch dataset, the question arises whether the predictive models would perform better on the UK dataset.

Chapter 6

Predictive Models of Unhealthy Snacking Behaviour

This chapter focuses on the predictive models that could aid in implementing timed behaviour change interventions. We address the unhealthy snacking behaviour in terms of classification and regression.

Initially, we describe the predictive models in Section 6.1, followed by the metrics used in the models in Section 6.2 and the models validation procedures in Section 6.3. Moreover, the modelling details for the classification and regression tasks are discussed in Sections 6.4 and 6.5, respectively.

6.1 Predictive models

We consider the following three types of ML models for classification and regression tasks:

1. Fixed context models include Linear Regression (LR), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Random Forest regressor (RFreg), eXtreme Gradient Boosting regressor (XGBreg), and Feed Forward Neural Network (FFNN). All participant data were included across all time points and fed into the models with a fixed input size of one data point.
2. A recurrent model is the Long Short-term Memory (LSTM), developed with observation sequences of four time steps. We fed the previous target into the observation sequence, known as a "teacher forcing". This is an appropriate approach for our application scenario of predicting one time bin or one event ahead.

3. Bayesian Network (BN) captures conditionally dependent and conditionally independent relationships between random variables. We used a maximum-likelihood estimation for computing the probability distribution to identify parameters that best describe the observed data.

We performed hyper-parameter tuning through a random search, with predefined parameter ranges via `scikit-learn`, detailed below, only performed in the one and two models. This is because the BN had no hyper-parameter to be tuned.

- RF and RFreg: no. estimators': [100, 150, 200, 250, 300], 'max depth': [2, 3, 4, 5, 6, 7], 'criterion' : [' gini' , 'entropy'].
- XGBoost and XGBreg: 'min samples split' :[2,3,4], 'no. estimators' : [100, 150, 200, 250, 300], 'max depth' : [2, 3, 4, 5, 6, 7], 'gamma' : [0.01, 0.05], 'learning rate' : [0.01, 0.05, 0.1, 0.5, 1].
- FFNN: three hidden layers with two dropouts regularisation of 0.4 and 0.6 after the 1st and 3rd hidden layers, respectively. Units: min value=128, max value=512, step=64. ReLU activation functions were utilised in the dense layers, whereas in the output layer, we employed softmax or ReLU (Kim, 2017) in the classification and regression tasks, respectively.
- LSTM: two hidden layers with two dropouts regularisation of 0.4 and 0.6 after the LSTM layer and 2nd hidden layers respectively. Units: min value=128, max value=512, step = 64. ReLU activation functions were adopted in the dense layers, whereas softmax was adopted in classifiers and ReLU in regressors in the output layer.

We then retrieved and reported the hyperparameters of best performing architecture of the models, as shown in Tables 6.6, 6.9, 6.14 and 6.17.

6.2 Metrics

For classification tasks in Section 6.4, we used several evaluation metrics. The most common one is accuracy, i.e. the proportion of correctly classified examples. Especially for multi-class tasks and for imbalanced classes, it is useful to look additionally at specificity, sensitivity, Youden's J, recall, precision, and combined F1 score.

These metrics are calculated for one class that is defined as positive against all other classes. This is done for each class in a multi-class case.

Sensitivity and specificity are commonly used metrics in clinical testing. They calculate the rate of true positives vs. all positives and true negatives vs. actual negatives, respectively.

$$\begin{aligned} \text{sensitivity} &= \text{recall} = TP / (TP + FN) \\ \text{specificity} &= TN / (TN + FP) \\ J &= \text{sensitivity} + \text{specificity} - 1 \end{aligned} \tag{6.1}$$

Youden's J combines sensitivity and specificity into a single metric for a classifier.

Precision, recall, and F1 are popular metrics in information retrieval. Precision is another name for sensitivity and recall uses all retrieved items as the denominator, rather than all actual negatives as in specificity.

$$\begin{aligned} \text{precision} &= TP / (TP + FP) \\ F1 \text{ score} &= 2(\text{recall} \times \text{precision}) / (\text{recall} + \text{precision}) \\ &= TP / (TP + 1/2(FP + FN)) \end{aligned} \tag{6.2}$$

The F1 score combines precision and recall as their geometric mean.

As to the regression task in Section 6.5, we used the Mean Absolute Error (MAE). Mean Absolute Error (MAE) and Root mean squared error (RMSE) are the most used metrics for measuring accuracy for continuous output. We wanted to know how far our predictions are from the actual time until the next unhealthy snack occurrence. Additionally, we wanted to quantify the extent of positive and negative errors to use in the analysis conducted in the next chapter (Chapter 7), creating hypothetical interventions and analysing them.

6.3 Models validation

To ensure the validity of our models, we adopted three approaches, including rerunning the models five times (five runs), applying Cross-Validation (CV), and utilising significance tests, but only for the fixed contest models because they possess the same type of data structure facilitating the comparison among them.

Regarding to the Cross-Validation (CV), we applied either CV or time series CV, which is a special case of CV involving, multiple folds/ splits in each run to allow the evaluation of our models on unseen data, as well as optimising our models using the randomised search described in Section 6.1. In each data split, the datasets were divided into 80% for training and 20% for testing. In addition, the class distributions in the training and testing

subsets were ensured to be similar to that in the whole dataset, using stratified data sampling (Reitermanova, 2010). A five-fold CV was applied over all the participants' data in all models except LSTM, which was dealt with through a time series CV. Time series CV is more suitable for time-series and temporal data modelling because the training set only includes the observations occurring before creating the testing set. It begins with a small subset of data for training, followed by predicting the next data points. The predicted data points are then included in the next training subset in order to generate new predictions (Hyndman and Athanasopoulos, 2014). Because we are dealing with sparse datasets (the UK and Dutch), we decided to use the time series CV with a minimum number of splits (two splits), allowing for the use of all available data.

As to the significance tests, we investigated whether there were significant differences in the performance of the fixed context models. Specifically, we applied the paired t-test in the task described in Section 6.4.1 in every split of a CV, adopted to compare RF and XGBoost and XGBoost and FFNN. We computed the t-statistic based on the Student's t-test with four degrees of freedom, under the null hypothesis that the models have equal performance as follows:

$$t = \frac{\bar{d}}{SE} \quad (6.3)$$

Where $\bar{d}$ is the mean of the 5 differences between classifiers' accuracies (in the 5-folds CV) and SE is the standard error of the differences calculated as follows:

$$SE = \frac{\sigma}{\sqrt{n}} \quad (6.4)$$

where σ is the standard deviation of the k differences. The t-statistic was then converted to a p-value, employing the well-known student t distribution and was compared to a significance level $\alpha = 0.05$. If the p-value is less than α , we can reject the null hypothesis and accept that the two models are significantly different.

6.4 Classification

In this section, we address the tasks of predicting unhealthy snacking occurrence in Section 6.4.1 and whether a snack is healthy or unhealthy in Section 6.4.2. We model the number of unhealthy snacks consumed in the 'next' time bin as a multi-class classification problem in Section 6.4.4. Table 6.1 summarises the details of the models used in each task as well as the output and input variables. The Table also illustrates the common and unique variables provided in the UK and Dutch datasets because not all the input variables are available in both

datasets. Note that the models' architecture does not change with different input variables in both datasets.

The sizes of the training and testing sets for the classification tasks in the UK and Dutch datasets with 4 and 12 time bins are shown in Table 6.2. The size of the training and testing sets only differs in the UK and Dutch datasets with the different number of time bins. This is because we imputed the location of the target class (no unhealthy snacking occurred), as it was not provided in both datasets; see section 6.4.1 for more details.

Table 6.1 Classification tasks description.

Task	Model	Output variable	Common input variables in both datasets	UK dataset specific input variables	Dutch dataset specific input variable/s
Predicting the occurrence of unhealthy snacking in a timebin in Section 6.4.1.	BN, FFNN, LSTM, XGBoost	Unhealthy (yes/ no)	Age, Gender, Employment Status, Body Mass Index (BMI), Current Location, Current Day of the Week, and Current Time Bin	Cognitive Restraint (CR), Uncontrolled Eating (UE) Emotional Eating (EE), Importance of Watching Weight (IWW), Importance of a Healthy Diet (IEHD)	Education
Predicting whether the next snack is healthy or unhealthy in Section 6.4.2	FFNN, LSTM	Snack (healthy/ unhealthy)	-	-	Age, gender, employment status, BMI, education, current location, current day of the week, and current time bin
Predicting the number of unhealthy snacks in Section 6.4.4	FFNN, LSTM	The number of unhealthy snacks (1,2, 3+)	Age, Gender, Employment Status, Body Mass Index (BMI), Current Location, Current Day of the Week, and Current Time Bin	Cognitive Restraint (CR), Uncontrolled Eating (UE) Emotional Eating (EE), Importance of Watching Weight (IWW), Importance of a Healthy Diet (IEHD)	Education

Table 6.2 the size of the training and testing sets for the classification tasks in the UK and Dutch datasets with 4 and 12 time bins.

Task	UK dataset				Dutch dataset			
	4 time bin		12 time bin		4 time bin		12 time bin	
	Train	Test	Train	Test	Train	Test	Train	Test
Predicting the occurrence of unhealthy snacking in a timebin in Section 6.4.1.	8243	2061	23360	5841	16984	4247	39564	9891
Predicting whether the next snack is healthy or unhealthy in Section 6.4.2.	-	-	-	-	5695	1424	5695	1424
Predicting the number of unhealthy snacks in Section 6.4.3	3982	996	3982	996	2778	927	2778	927

6.4.1 Prediction of unhealthy snacking occurrence in a time bin

The objective of this task is predicting the occurrence of unhealthy snacking in one of 4 or 12 time bins in the UK and Dutch dataset. Day time was sliced into 4 and 12 time bins, as described in chapter 5, in order to explore the models' performance in coarser and finer time slots.

This task was treated as a binary classification task. We used Age, Gender, Employment Status, Body Mass Index (BMI), Cognitive Restraint (CR), Uncontrolled Eating (UE), Emotional Eating (EE), Importance of Watching Weight (IWW), Importance of a Healthy Diet (IEHD), Education, Current Location, Current Day of the Week, and Current Time Bin, as input variables. Since the CR, UE, EE, IWW, and IEHD variables were only collected in the UK dataset, while education was only collected in the Dutch dataset, these variables were only utilised in the modelling for their respective datasets. The target variable was unhealthy snacking (yes or no). The class 'no' covers two situations: healthy snacking and not snacking at all. Time bins with not snacking had no location information associated with them in both datasets, while, as noted, healthy snacking information was only collected in the Dutch dataset. Hence, we imputed the location of the situation of not snacking at all for the Dutch dataset. Additionally, it was necessary to impute the location information for the time bins with no snacking (both situations) in the UK dataset. The imputation was performed

using a K-nearest neighbours method (KNN) (Peng and Lei, 2005). The KNN imputer was trained and tested using the location collected in both datasets, with separate training and test sets. Table 6.3 shows the results of the imputed locations.

Table 6.3 The test accuracy of imputed locations for time bins without snacking using, K-nearest neighbour imputation.

	UK dataset	Dutch dataset
4 time bins	73%	72%
12 time bins	73%	72%

From Table 6.3, it can be seen that the location was imputed with a reasonable accuracy of 73% in the UK dataset and 72% in the Dutch dataset, using $k=5$. Moreover, accuracies did not differ when imputing the location in 4 and 12 time bins, in both datasets. Then BN, FFNN, LSTM and XGBoost were applied as described in Sections 6.1, 6.2 and 6.3.

Table 6.4 The accuracy of predicting the occurrence of unhealthy snacking in time bins of 4 and 12 in the two main datasets, UK (self-collected) and Dutch. The target variable was unhealthy snacking (yes, no).

model	UK				Dutch			
	4 time bins		12 time bins		4 time bins		12 time bins	
	Train acc	Test acc	Train acc	Test acc	Train acc	Test acc	Train acc	Test acc
BN	89%	89%	89%	88%	86%	85%	90%	90%
RF	89%	89%	88%	88%	77%	77%	80%	80%
XGBoost	92%	91%	88%	88%	87%	87%	93%	91%
FFNN	90%	90%	88%	88%	87%	86%	91%	90%
LSTM	91%	90%	88%	88%	87%	86%	91%	90%

Table 6.5 Results of paired t-test with 4 degrees of freedom and 0.05 significance level on 5-fold CV, under the null hypothesis that the models have equal performance.

	UK		Dutch	
	4 time bins	12 time bins	4 time bins	12 time bins
RF vs. XGBoost	Equal t-Statistic: -1.5 P-value: 0.2	Equal t-Statistic: -1.5 P-value: 0.2	Different t-Statistic: -54.1 P-value: 0.00	Different t-Statistic: -56.7 P-value: 0.00
XGBoost vs. FFNN	Equal t-Statistic: 0.7 P-value: 0.5	Equal t-Statistic: 0.7 P-value: 0.5	Equal t-Statistic: -2.5 P-value: 0.1	Different t-Statistic: 4.18 P-value: 0.01

Table 6.4 shows the accuracies obtained from the classifiers with 4 or 12 time bins in both datasets. The significance test results in Table 6.5 show that the two fixed context models provide similar accuracies on the UK dataset with 4 or 12 time bins. In contrast, the accuracies on the Dutch dataset were significantly different except between XGBoost and FFNN with four time bins.

Table 6.6 The Hyper parameters of the best performing models for predicting the of unhealthy snacking occurrence in time bins of 4 and 12 in the two main datasets, UK (self-collected) and Dutch. The models were optimised with random search. n-E is the number of estimators and LE is the learning Rate.

	UK		Dutch	
	4 time bins	12 time bins	4 time bins	12 time bins
RF	'n_E': 250, 'max_depth': 7, 'criterion': 'gini'	'n_E ': 100, 'max_depth': 7, 'criterion': 'entropy'	'n_E ': 150, 'max_depth': 7, 'criterion': 'entropy'	n_E ': 300, 'max_depth': 7, 'criterion': 'gini'
XGBoost	'n_E ': 200, 'max_depth': 5, ' LR: 0.1, 'gamma': 0.05	'n_E ': 250, 'max_depth': 4, ' LR: 0.05, 'gamma': 0.05	'n_E ': 200, 'max_depth': 6, ' LR: 1, 'gamma': 0.05	n_E ': 200, 'max_depth': 7, 'LR: 0.5, 'gamma': 0.01
FFNN	dense: 384, dense_1: 384, dense_2: 384, dense_3: 384, dense_4: 2	dense: 265, dense_1: 265, dense_2: 265, dense_3: 265, dense_4: 2	dense: 265, dense_1: 265, dense_2: 265, dense_3: 265, dense_4: 2	dense: 448, dense_1: 448, dense_2: 448, dense_3: 448, dense_4: 2
LSTM	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 2	Lstm: 448, dense: 512, dense_1: 512, dense_2: 2	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 2	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 2

Table 6.6 shows that most of the hyper parameters of the best models varies between the models in the UK and Dutch in 4 and 12 time bins. The precision, recall, and F1 score metrics of the classifiers are shown in Table 6.7. In Table 6.4, the accuracies for all classifiers are high with both 4 and 12 time bins, in both datasets. XGBosst provided the best performance, i.e. highest accuracy, precision, recall, and the F1 score in most cases. With a closer look at the precision, recall, and F1 score for all the models in Table 6.7 b, it could be noticed that the classifiers' ability to predict both classes (i.e., yes vs. no, the presence of an unhealthy snack vs. not) is better when the classes are balanced (i.e. with the UK dataset with 4 time

bins). In all other cases, i.e. in the UK dataset with 12 time bins and the Dutch dataset with both 4 and 12 time bins, the classifiers did not predict the class of yes (the minority) as well as predicting the class of no (the majority class). In these imbalanced cases, in predicting the class of yes (unhealthy snack), recall is not high compared to precision. This indicates that the classifiers do not predict all unhealthy snacks, but when they offer a prediction, it is mostly correct. High recalls are required in applications with sensitive or risky predictions, such as predicting cancer (Karim et al., 2019). In this task, the ‘cost’ of not detecting all unhealthy snacks is not high, although it would be ideal to obtain unhealthy snack predictions with both high recall and precision. Therefore, unhealthy snacks predictions with moderate recall might be acceptable in predicting whether an unhealthy snack is going to take place or not.

When intending to improve any of the recall or precision, a precision recall trade-off is expected, due to the inverse relationship between the two measures (Gordon and Kochen, 1989). Chapter 7 covers a recall precision trade-off analysis for an intervention design (that is, a design which might be employed in a putative intervention), in addition to addressing the effect of oversampling on model precision and recall.

Feature importance analysis

In feature importance analysis, each feature is assigned a score based on its relative predictive power. Feature importance analysis can be used for eliminating features with little or no predictive power, reducing computational cost and helping to prevent over-fitting (Razmjoo et al., 2017). In modelling unhealthy snacking behaviour, feature importance can help with choosing the optimal features in future data collection for intervention designs.

We retrieved the relative importance scores for each input feature used in Section 6.4.1, for predicting unhealthy snacking behaviour, across a randomized search, 5-fold cross-validation approach, with the top Machine Learning (ML) classifier (ie, XGBoost), in Figure 6.1. The score of each input feature was based on the explained variance each feature contributes to the model. Although the ranks of features varied substantially, the analysis shows that location and time bin are the most important features in predicting whether an unhealthy snack is going to occur in a time bin. BMI-pre, Job Seeker, Age and Post-secondary Vocational Education were highly ranked only in the case of (c), the Dutch dataset with 4 time bins. Ethnic group was only a significant feature in case of (a), the UK dataset with 4 time bins; overall there is little evidence that this feature influences unhealthy snacking behaviour.

Table 6.7 The metrics for predicting the occurrence of unhealthy snacking in time bins of 4 and 12 in the two main datasets, UK (self-collected) and Dutch. a) Binary-class metrics for the UK Dataset, b) Binary-class metrics for the Dutch dataset.

a) Binary-class metrics for the UK Dataset. The frequencies of the target classes in 4 time bins were 54% and 46%. The frequencies of the target classes in 12 time bins were 83% and 17%.

model	Metric	4 time bins		12 time bins	
		yes	No	yes	No
BN	Precision	93%	87%	100%	88%
	Recall	83%	95%	32%	100%
	F1 score	88%	91%	49%	93%
FFNN	Precision	96%	86%	95%	88%
	Recall	82%	97%	34%	100%
	F1 score	88%	91%	50%	93%
LSTM	Precision	93%	87%	96%	87%
	Recall	83%	95%	33%	100%
	F1 score	88%	91%	49%	94%
XGBoost	Precision	92%	91%	90%	88%
	Recall	90%	93%	36%	99%
	F1 score	91%	92%	52%	93%

b) Binary-class metrics for the Dutch dataset. The frequencies of the target classes in 4 time bins were 84% and 16%. The frequencies of the target classes in 12 time bins were 93% and 7%.

Model	Metric	Dutch			
		4 time bins		12 time bins	
		yes	No	yes	No
BN	Precision	61%	86%	71%	95%
	Recall	16%	98%	17%	100%
	F1 score	26%	92%	27%	97%
FFNN	Precision	73%	87%	69%	95%
	Recall	19%	99%	18%	99%
	F1 score	31%	92%	29%	97%
LSTM	Precision	75%	86%	75%	95%
	Recall	19%	99%	19%	100%
	F1 score	31%	92%	30%	97%
XGBoost	Precision	72%	88%	76%	95%
	Recall	25%	98%	26%	99%
	F1 score	38%	93%	39%	97%

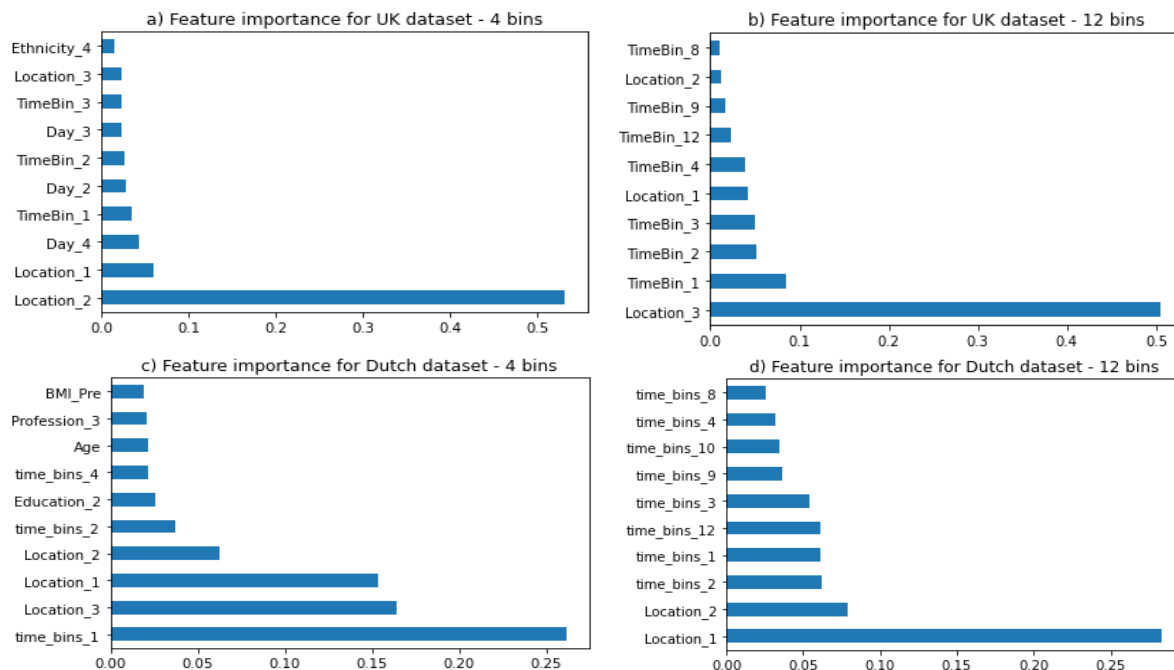


Fig. 6.1 Feature importance from XGBoost in the two main datasets, UK (self-collected) and Dutch while day time was sliced into 4 and 12 time bins. Ethnicity-4 is other ethnic groups, Location-1 is Home, Location-2 is work/ study, Location-3 is other and Education-2 is post-secondary vocational education.

6.4.2 Predict whether the next snack is healthy or unhealthy

The objective of this task was to predict whether the next snack is healthy or unhealthy. Since healthy snacks were only collected in the Dutch dataset, this task was only performed using the Dutch dataset. We used the age, gender, employment status, BMI, education, current location, current day of the week, and current time bin as input variables. The target variable was whether a snack was healthy or unhealthy and so this task was framed as a binary classification problem. The differences between the two binary prediction tasks corresponding to unhealthy snacking in Section 6.4.1 and healthy or unhealthy snacking in the present section is that in the previous section we aimed to predict the occurrence of unhealthy snacking in each time bin, whereas presently we aim to predict the type of next snack as healthy or unhealthy. For the present task, time bins were treated as an input variable, whereas in the Section 6.4.1 as intervals for the prediction of unhealthy snacking.

The FFNN and LSTM models were applied in this task as been technically described in Sections 6.1, 6.2 and 6.3.

Table 6.8 Prediction of healthiness of the next snack in the Dutch dataset. The frequencies of the target classes were 84% and 16% in 4 and 12 time bins.

a) Accuracy

Encoding	4 time bins		12 time bins	
Model	Train	Test	Train	Test
FFNN	73%	72%	74%	70%
LSTM	70%	70%	67%	70%

b) Precision, Recall and F1 Score

Model	Metric	4 time bins		12 time bins	
		Healthy	Unhealthy	Healthy	Unhealthy
FFNN	Precision	72%	70%	77%	66%
	Recall	76%	66%	65%	77%
	F1 score	74%	68%	71%	71%
LSTM	Precision	75%	66%	72%	68%
	Recall	68%	74%	73%	68%
	F1 score	71%	70%	73%	68%

Table 6.9 The hyper parameters of the best performing model for the Prediction of healthiness of the next snack in the Dutch dataset. The models were optimised with random search.

	4 time bins	12 time bins
FFNN	dense: 320, dense_1: 320, dense_2: 320, dense_3: 320, dense_4: 2	dense: 384, dense_1: 384, dense_2: 384, dense_3: 384, dense_4: 2
LSTM	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 2	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 2

Table 6.8 4 shows the prediction of unhealthy or healthy snack type in the Dutch dataset. It shows that FFNN and LSTM provided reasonable and relatively similar performance in this task. There were small variations in the hyper parameters of the best models performed on the Dutch dataset, as shown in Table 6.9. Additionally, recall and precision scores indicate that the models performed similarly in both target classes. This model will be addressed further below with respect to class imbalance.

6.4.3 Addressing class imbalance and loss asymmetry

Our classifiers assign samples to classes using probabilities, and the probability threshold is usually set to 0.5 by default. Depending on the actual impact of interventions on users and the behaviour of our classifiers, we may require a more customized threshold (Li et al., 2019). There may also be different behaviour for different classes, which is particularly common for imbalanced classes. For instance, this is evident for case (a), which involves 12 time bins, in Table 6.3 and for both (b) cases in Table 6.7 and both cases of (b) in Table 6.8. Changing the classification threshold can improve the recall, however, this is expected to occur at the cost of reducing precision. We use Receiver Operating Characteristics (ROC) and Precision/Recall (PR) curves to understand the behaviour of our classifiers for different decision thresholds and for class imbalance. For a predictive model, a ROC curve summarizes the trade-off between sensitivity and specificity with different probability thresholds. The PR curve summarizes the trade-off between precision and recall. In the ROC curve, the closer the curve to the top left of the plot, the better the model is. In the PR curve, the closer the curve to the top right of the plot, the better the model is.

ROC and PR curves were constructed for the LSTM model introduced in Sections 6.4.1 and 6.4.2. Although LSTM was not the best performing model in Section 6.4.1, the LSTM performance was not substantially different from the best model (XGBoost) in this task. We used the ROC and PR curves to study the effect of using SMOTE (Chawla et al., 2002) for the models with imbalanced classes. The imbalanced classes can be seen in Table 6.7 case (a) with 12 time bins and both cases of (b). Figures 6.2 and 6.4 show the ROC curve for the LSTM model and Figures 6.3 and 6.5 show the PR curves. The optimal threshold values for F1 or J metric are marked with a black point. In Figures 6.2 and 6.3, it can be clearly observed that the ROC and PR curves are improved after oversampling the datasets with imbalanced classes. These results agree with what was reported in Chawla et al. (2002) and priyadarsini (2021) that oversampling improves the ROC curve, F1 score, Precision and Recall.

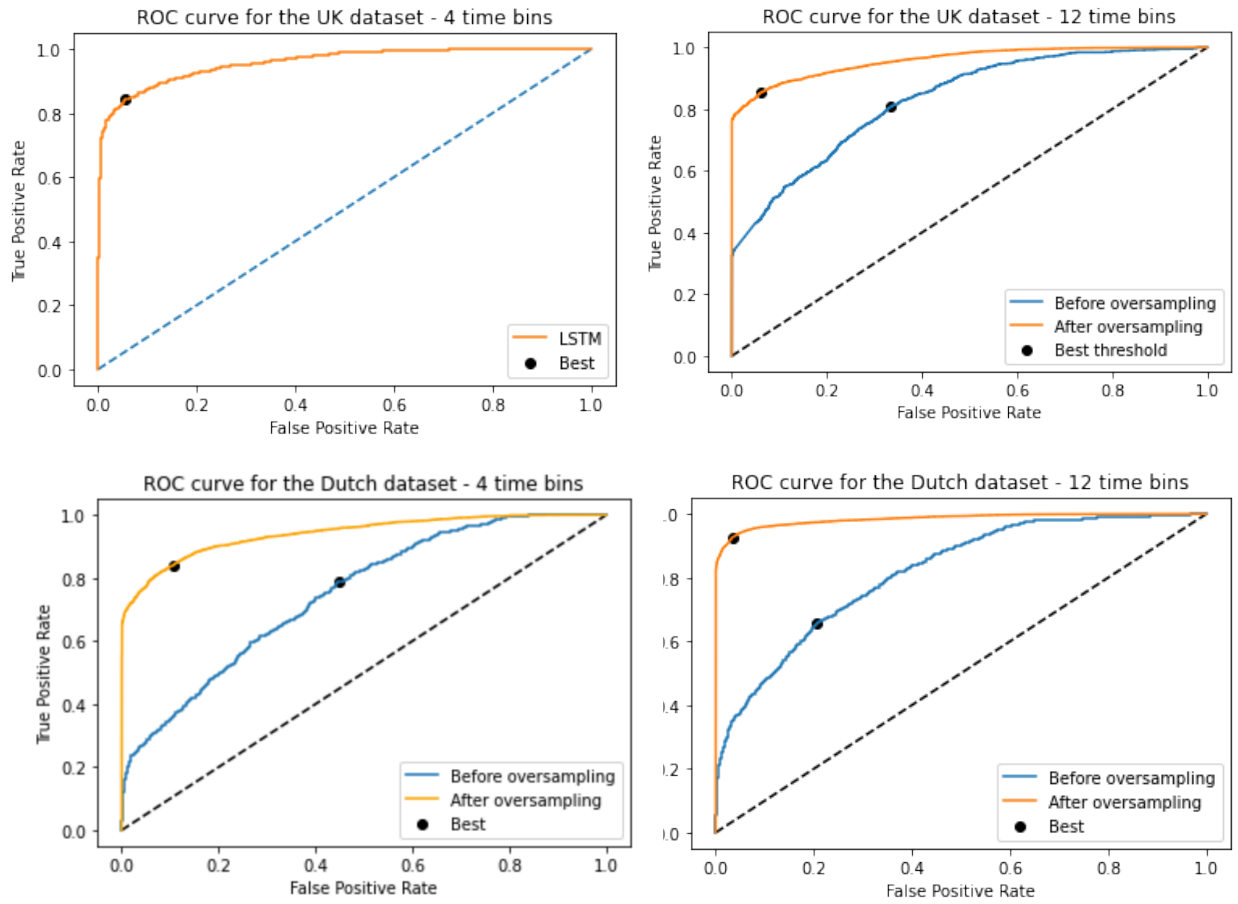


Fig. 6.2 ROC curves for the prediction of unhealthy snack occurrence with 4 and 12 time bins in the two main datasets, UK and Dutch. Oversampling was applied on datasets with imbalanced classes in the target variable.

PR curves are more suitable for predicting an outcome that is relatively rare or in imbalanced classes, because in ROC curves the specificity tends to show small values in imbalanced data, due to large numbers of negatives. Thus, specificity is less informative for the classifier performance, compared to the PR curves (Cook and Ramadas, 2020), if missing actual negatives is not problematic. However, the PR curves still show significant improvement with oversampling, see Figure 6.3.

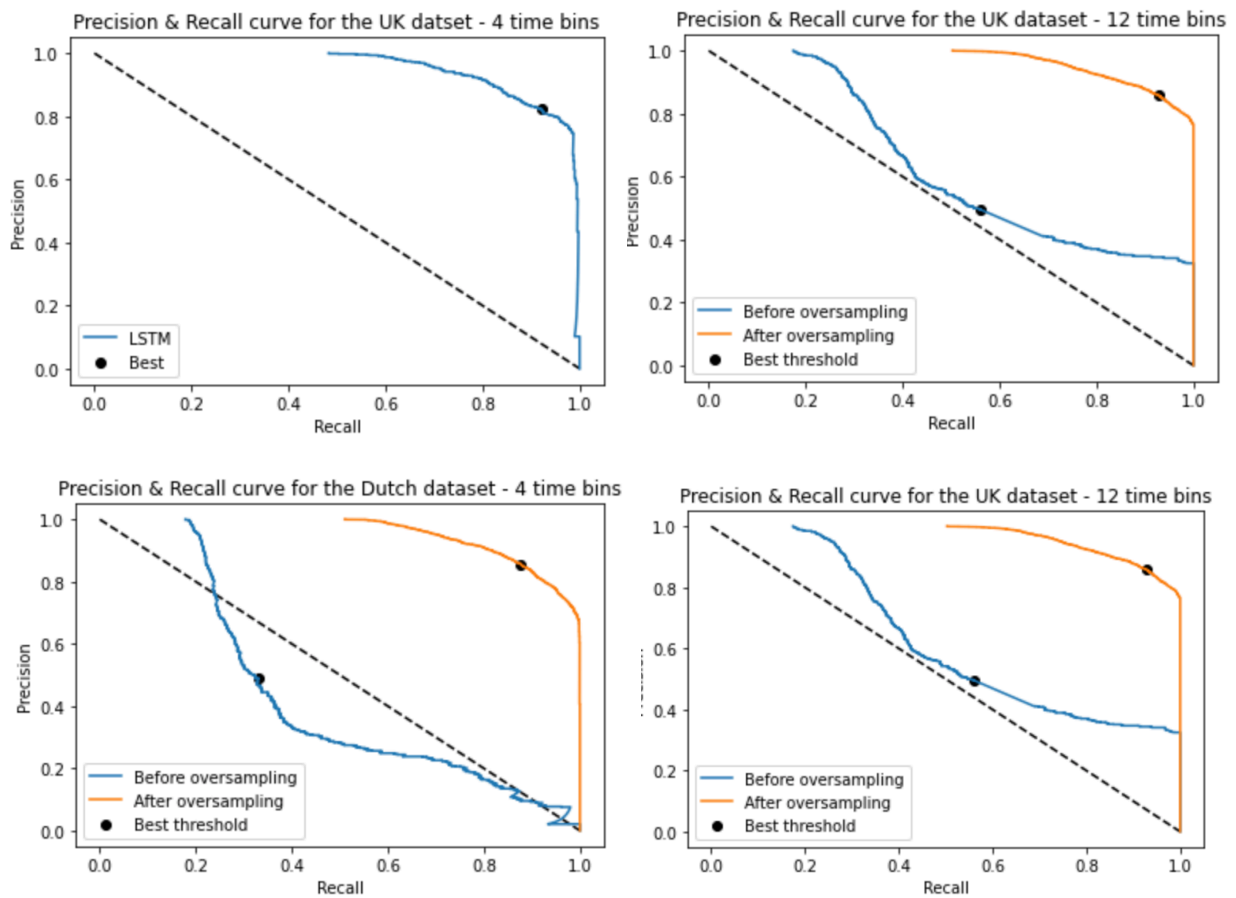


Fig. 6.3 Precision/recall curves for whether a snack is healthy or unhealthy, with 4 and 12 time bins, in the UK and Dutch datasets, predicted with an LSTM model. Oversampling was applied in cases where there were imbalanced classes in the target variable.

Figures 6.4 and 6.5 show the ROC and PR curves for the task of predicting the type of snacking behaviour as healthy or unhealthy conducted in Section 6.4.2. The classes were balanced so no oversampling was performed in this case. The ROC and PR curves in the figures show a moderate performance of the LSTM model in the target classes.

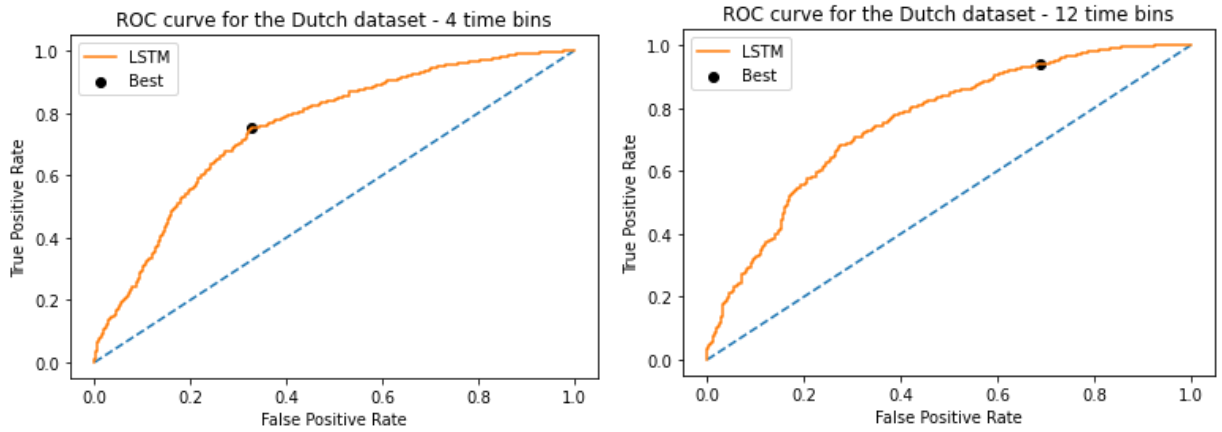


Fig. 6.4 ROC curves for the classification the type of snacking as healthy or unhealthy in time bins of 4 and 12, in the Dutch dataset. The predicted cases were from the LSTM model. No oversampling was required.

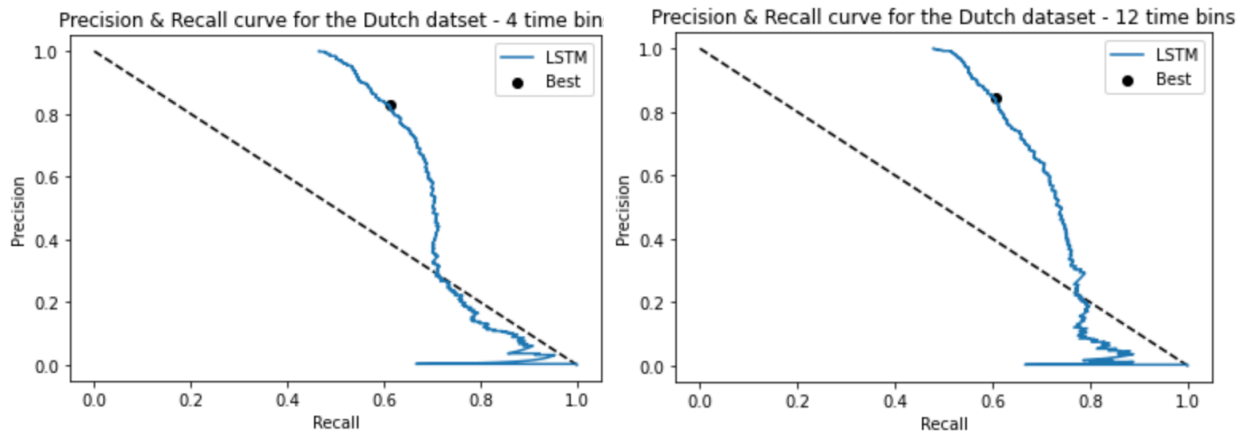


Fig. 6.5 Precision recall curves for the classification of snacks as healthy or unhealthy in time bins of 4 and 12, in the Dutch dataset. The predicted cases were from the LSTM model. No oversampling was required.

Better ROC and PR curves show better results for predicting unhealthy snacking occurrence, after oversampling to correct for imbalanced classes, compared to predicting whether the next snack would be healthy vs. unhealthy.

6.4.4 Predicting the number of unhealthy snacks

This task aims at estimating the number of unhealthy snacks in the time bins where unhealthy snacking was reported. We model this as a classification task, such that the classes are 1, 2, and 3+ (3 or more snacks).

Initially, the number of unhealthy snacks recorded by each participant was calculated in the time bins where the recording/s were made. This variable was used as a target variable, in order to predict the number of unhealthy snacks someone would have in the next time bin, if there is an unhealthy snack. Healthy snacks were excluded from the Dutch dataset in order to be able to compare the Dutch dataset with the UK one. The range of unhealthy snacks was reduced from 6 to 3 in the UK dataset and from 8 to 3 in the Dutch dataset. Reducing the range of unhealthy snacks was due to the fact that consuming three unhealthy snacks or more in a time bin was rare and so, for the cases when this happens, predicting the timing would be less challenging. We use Age, Gender, Employment Status, BMI, CR, UE, EE, IWW, IEHD, Education, Current Location, Day of the Week, and Current Time Bin as input variables. Since CR, UE, EE, IWW, and IEHD variables were only collected in the UK dataset and Education was only collected in the Dutch dataset, these variables were only used in models with their respective dataset.

We then ran another analysis on the Dutch dataset, taking into account the healthy snacks consumed in any time bins, in order to explore if including healthy snacks here would influence model predictions. These healthy snacks were accounted for in the 0 category in 6.13.

Table 6.10 The accuracies for predicting the number of unhealthy snacks consumed in time bins of 4 and 12 in the two main datasets, UK and Dutch. The classes of the target variable are 1,2 and 3+.

Dataset	UK				Dutch			
Model	4 time bins		12 time bins		4 time bins		12 time bins	
	Train	Test	Train	Test	Train	Test	Train	Test
FFNN	68%	65%	87%	87%	68%	65%	79%	79%
LSTM	60%	57%	87%	87%	68%	67%	74%	76%

Table 6.11 The metrics for predicting the number of unhealthy snacks consumed (1,2 and 3+) in time bins of 4 and 12 in the two main datasets, UK and Dutch. a) Multi-class metrics for the UK Dataset b) Multi-class metrics for the Dutch dataset

a) Multi-class metrics for the UK Dataset. The frequencies of the target classes in 4 time bins were 59%, 30% and 11%. The frequencies of the target classes in 12 time bins were 86%, 12% and 2%.

Model	Metric	4 time bins			12 time bins		
		1	2	3+	1	2	3+
FFNN	Precision	72%	45%	66%	87%	67%	61%
	Recall	84%	34%	52%	99%	1%	52%
	F1 score	77%	39%	58%	93%	3%	56%
LSTM	Precision	58%	31%	6%	89%	45%	71%
	Recall	96%	2%	1%	97%	19%	71%
	F1 score	73%	4%	22%	92%	26%	71%

b) Multi-class metrics for the Dutch dataset. The frequency percentages in 4 time bins were 59%, 25%, and 16%. The frequency percentages in 12 time bins were 68%, 19%, and 13%.

Model	Metric	4 time bins			12 time bins		
		1	2	3+	1	2	3+
FFNN	Precision	70%	47%	66%	80%	65%	88%
	Recall	78%	34%	68%	95%	31%	69%
	F1 score	74%	40%	67%	87%	42%	78%
LSTM	Precision	67%	54%	71%	78%	67%	64%
	Recall	92%	7%	72%	91%	18%	75%
	F1 score	77%	13%	72%	84%	29%	69%

Tables 6.10 and 6.11 show the results in estimating the number of unhealthy snacks in both datasets. FFNN and LSTM performed relatively similarly in both datasets. The accuracies when using 12 time bins were higher than when using 4 time bins. For the UK dataset, the LSTM metrics are better with 4 time bins, whereas, FFNN metrics were better with 12 time bins. For the Dutch dataset, the metrics were better for FFNN with 4 and 12 time bins.

The FFNN and LSTM models suffered from under-fitting and over-fitting in the UK dataset, when the time bins were 4 and 12 respectively. This may have been due to the severity of the imbalanced classes in the UK dataset, as this has been suggested to be sometimes the cause of under-fitting and over-fitting (Fatima and Saeed, 2021). The severity of imbalanced

classes was solved by oversampling, as can be seen in Table 6.12. Subsequently, the models' overall accuracy improved, in addition to improved recall in the minority classes.

Table 6.12 Estimating the number of unhealthy snacks consumed when using 4 and 12 time bins in the UK dataset after oversampling. a) Accuracies. b) Multi-class metrics for the UK dataset. The classes of the target variable are 1,2 and 3+. The target classes were balanced.

a) Accuracy

Model	4 time bins		12 time bins	
	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy
FFNN	74%	73%	91%	90%
LSTM	75%	74%	91%	93%

b) Multi-class metrics for the UK dataset.

Model	Metric	4 time bins			12 time bins		
		1	2	3+	1	2	3+
FFNN	Precision	71%	78%	74%	87%	96%	90%
	Recall	83%	44%	94%	96%	75%	99%
	F1 score	76%	56%	83%	91%	84%	94%
LSTM	Precision	60%	84%	93%	86%	96%	99%
	Recall	92%	52%	78%	96%	85%	99%
	F1 score	73%	64%	85%	91%	90%	99%

When the 0 class was included in the Dutch dataset, the accuracies and metrics were better in both FFNN and LSTM. This may suggest that the models performed better due to the increased size of the dataset, when including the 0 class (healthy snacks). It could indicate that healthy snacks positively influence the models' ability to predict the count of unhealthy snacks. The number of units in the dense layer was 256 or 320 most of the times as retrieved from the best performing models presented in Table 6.14.

Table 6.14 The Hyper parameters of the best performing model for predicting the number of unhealthy snacks consumed in time bins of 4 and 12 in the two main datasets, UK and Dutch. The models were optimised with random search after balancing the output classes.

	UK		Dutch	
	4 time bins	12 time bins	4 time bins	12 time bins
FFNN	dense: 320, dense_1: 320, dense_2: 320, dense_3: 320, dense_4: 3	dense: 384, dense_1: 384, dense_2: 384, dense_3: 384, dense_4: 3	dense: 320, dense_1: 320, dense_2: 320, dense_3: 320, dense_4: 4	dense: 320, dense_1: 320, dense_2: 320, dense_3: 320, dense_4: 4
LSTM	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 3	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 3	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 4	Lstm: 320, dense: 256, dense_1: 256, dense_2 = 4

Table 6.13 Estimation of the number of unhealthy snacks consumed as 0,1,2 or 3+ in time bins of 4 and 12 in Dutch dataset. a) Accuracy b) Multi-class metrics for the Dutch dataset. The frequencies of the target classes in 4 time bins were 48%, 31% , 13% and 8%. The frequencies of the target classes in 12 time bins were 52%, 33%, 9% and 6%.

a) Accuracy

Method	4 time bins		12 time bins	
	Train	Test	Train	Test
FFNN	80%	79%	88%	86%
LSTM	81%	79%	85%	86%

b) Multi-class metrics for the Dutch dataset.

Dataset		Dutch							
Model	Metric	4 time bins				12 time bins			
		0	1	2	3+	0	1	2	3+
FFNN	Precision	90%	67%	57%	69%	96%	78%	50%	72%
	Recall	100%	80%	9%	61%	100%	89%	17%	71%
	F1 score	94%	73%	16%	64%	98%	83%	25%	72%
LSTM	Precision	89%	68%	46%	77%	96%	73%	60%	74%
	Recall	100%	79%	16%	56%	100%	93%	2%	54%
	F1 score	94%	73%	24%	65%	98%	82%	5%	63%

6.5 Regression task

In this section, we model the prediction of the duration from one snack until the next unhealthy snack. We only used the common input variables in the UK and Dutch datasets, specifically the dynamic ones (i. e. time bin, day of the week, and location) because these predictors were found to be better predictors of unhealthy snacking in the feature importance analysis in Section 6.4.1. Table 6.15 details the models used in this task and the output and input variables. The size of the training and testing sets were 3,982 and 996 in the UK dataset and 2,778 and 927 in the Dutch dataset.

Table 6.15 Regression task description.

Task	Model	Output variable	input variables
Predicting the time until the next unhealthy snack	LR, RFreg, XGBreg, LSTM, FFNN	Time until the next unhealthy snack, in minutes	Current Location, Current Day of the Week, and Current Time Bin

6.5.1 Predicting the time until the next unhealthy snack

Sending timed interventions requires predicting when an unhealthy snack is going to take place. Since no information was collected from participants between snacks for both the UK and the Dutch dataset, the prediction of time was based on the time of a current unhealthy snack as a predictor for the time of the next snack. The prediction problem in this task was modelled as a regression. We used Current Location, Current Day of the Week, and Current Time Bin as input parameters. The Time Bin variable was treated as an input variable. A fixed input size was fed into the non-recurrent models (i. e. LR, RFreg, XGBreg and FFNN) and into the LSTM, as a sequence of inputs, as described in Section 6.1. The target variable was the time until the next unhealthy snack, in minutes.

The results in Table 6.16 show that the Mean Absolute Error (MAE) was relatively similar for all models. In all cases, LSTM and FFNN seemed to have the lowest MAE, with the exception of the UK dataset when a day was sliced into 12 time bins; RFreg, and XGBreg had the lowest MAE. Table 6.17 shows that the hyper parameters of the best performing models vary between different models and cases, in the UK and Dutch datasets, and for the 4 and 12 time bins.

Table 6.17 The Hyper parameters of the best performing model for or predicting the time of next unhealthy snack in minutes, when using 4 or 12 time bins, in the two main datasets, UK and Dutch. The models were optimised with random search. n-E is the number of estimators and Min_SS is the minimum samples split.

	UK		Dutch	
	4 time bins	12 time bins	4 time bins	12 time bins
RFreg	'n_E': 150, 'min_SS ': 4, 'max_depth': 2	'n_E': 250, 'min_SS': 3, 'max_depth': 3	'n_E': 250, 'min_SS ': 4, 'max_depth': 3	'n_E': 150, 'min_SS ': 2, 'max_depth': 5
XGBreg	'n_E': 200, 'max_depth': 3, 'LR': 0.01, 'gamma': 0.05}	n_E': 300, 'max_depth': 2, ' LR ': 0.1, 'gamma': 0.05	n_E': 150, 'max_depth': 3, 'LR': 0.1, 'gamma': 0.05	'n_E': 250, 'max_depth': 4, 'LR': 0.01, 'gamma': 0.05
FFNN	dense: 128, dense_1: 128, dense_2: 128, dense_3: 128, dense_4: 1	dense: 192, dense_1: 128, dense_2: 192, dense_3: 192, dense_4: 1	dense: 192, dense_1: 128, dense_2: 192, dense_3: 192, dense_4: 1	dense: 192, dense_1: 128, dense_2: 192, dense_3: 192, dense_4: 1
LSTM	Lstm: 320, dense: 192, dense_1: 192, dense_2 = 1	Lstm: 320, dense: 256, dense_1: 512, dense_2 = 1	Lstm: 320, dense: 256, dense_1: 512, dense_2 = 1	Lstm: 256, dense: 512 dense_1: 512 dense_2 = 1

Table 6.16 MAE for predicting the time of next unhealthy snack in minutes, when using 4 or 12 time bins, in the two main datasets, UK and Dutch.

Dataset	UK		Dutch	
Model	4 time bins	12 time bins	4 time bins	12 time bins
LR	20%	17%	70%	67%
RFreg	19%	17%	69%	67%
XGBreg	18%	17%	70%	67%
FFNN	17%	17%	56%	56%
LSTM	18%	16%	56%	57%

6.6 Discussion

Fixed context models outperformed other models in predicting whether an unhealthy snack is going to take place in the next time bin. Additionally, they outperformed the recurrent

model in predicting the time until the next unhealthy snack. The LSTM provided a relatively similar performance to FFNN in predicting healthy or unhealthy snacking, in addition to estimating the number of unhealthy snacks. However, after over-sampling the severity of imbalanced classes in the UK dataset when estimating the number of unhealthy snacks, LSTM performance was slightly better than FFNN performance. Moreover, the time prediction was not very precise. In real applications, we would want to use more information, e.g. from general use of a mobile device or from a fitness tracker or a smartwatch.

We see these two tasks of predicting unhealthy snack occurrence and the time until the next unhealthy snack as essential steps for timing behaviour change interventions, aided by predictive models. In the next chapter, will analyse the models' performance for the tasks for intervention timing.

Chapter 7

Application and Intervention Design Considerations

One of the main objectives of this thesis was to produce predictive models for supporting behaviour change with timed interventions. In this chapter, we propose an analysis more geared towards such applications, based on the specific timing criteria desirable for a putative behaviour change intervention. That is, even though we do not employ an intervention, we offer an analysis concerning timings, as would be necessary to construct a variety of interventions.

In psychology, an influential distinction is to understand human thought as reflecting two types of processes, reflective and impulsive processes . We can consider human behaviour change, in terms of these processes, as discussed in Section 2.3. It is thought that unhealthy snacking may often be an automatic process, as suggested by Tapper (2021). (Tapper, 2021). Many researchers agree on the importance of the delivery time of the intervention, as discussed in Chapter 2. The effectiveness of interventions has been demonstrated to vary depending on the timing of the delivery. Information we have been exposed to more recently tends to be more cognitively accessible. Nudges presented too early, relative to the targeted action, may be ineffective, because of forgetting, while nudges presented late may reduce the amount of time available for processing and executing a change (Gillitzer and Sinning, 2020). In the context of predictive models, there is also the risk of the predicted time being later than the intervention, so that the intervention happens after the targeted event and the opportunity for change has been missed. In general, a nudge offered a short time before the behaviour is desirable, but this has to be balanced against the risk of being too late.

In this chapter, we evaluate the Random Forest Regressor model's performance in addressing the main predictive aim of timing in unhealthy snacking behaviour, by considering

the distribution of residuals concerning the time for the next unhealthy snack and by producing the outcome with different timing parameters.

7.1 Distribution of residuals

In this section we discuss the distribution of residuals for predicted times for the next unhealthy snack, in order to assess the quality of the predictions for our task. In general, we expect a normal residual distribution for many noisy processes and the width of the distribution should depend on the amount of noise and the quality of the model. For instance, a normal narrow distribution would in general go hand in hand with good quality of predictions (Fuller, 1986)).

We plot histograms of the residuals of all models considered in section n 6.5.1 for predicting the time until next unhealthy snack. In Figures 7.1-7.5, the histograms are reduced to a range of -200 – +200 minutes, as larger errors are outliers in the sense that these predictions are not useful for the purpose of timing interventions. The residuals are calculated as predicted minus true values, so that negative residuals indicate that the prediction is earlier than the observed time and positive residuals indicate that the predicted time is later than observed.

Figure 7.1 shows the histogram of residuals for LR for the UK and Dutch datasets. With the UK dataset and 4 time bins, the distribution is left skewed, i.e. there are more errors left of the peak. This is actually a good trait for putative interventions, as earlier predictions are less problematic than late ones, as long as the deviation is not too extreme. The residuals were less skewed with 12 time bins in the UK dataset. On the Dutch dataset, the width of the distributions was large with both 4 and 12 time bins, so that this model in this context would not be suitable for a putative intervention.

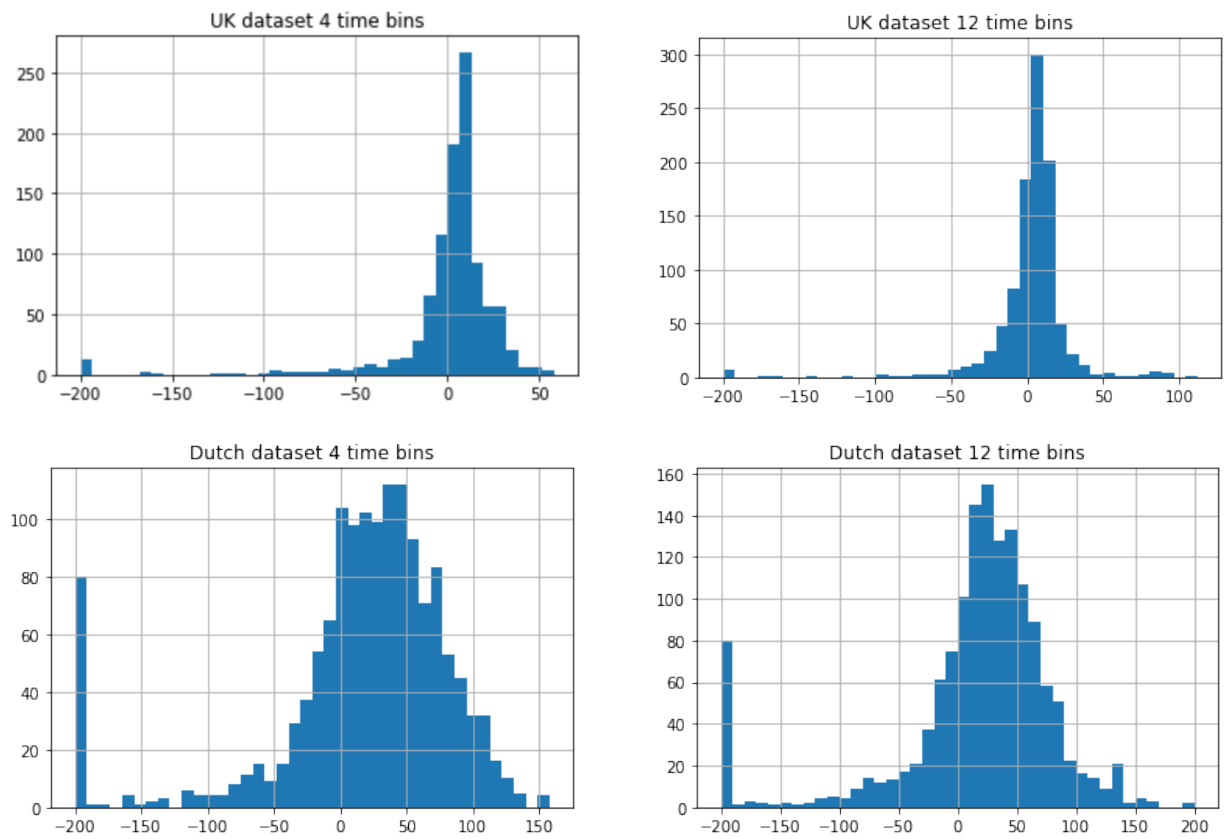


Fig. 7.1 Histogram of residuals for LR with 4 and 12 time bins, for the UK and Dutch datasets.

The histogram of residuals for RFreg, and XGBreg in Figures 7.2 and 7.3 shows close to normal distributions in the UK and Dutch datasets. In most of cases, RFreg residuals were closer to normal distributions in the UK and Dutch dataset.

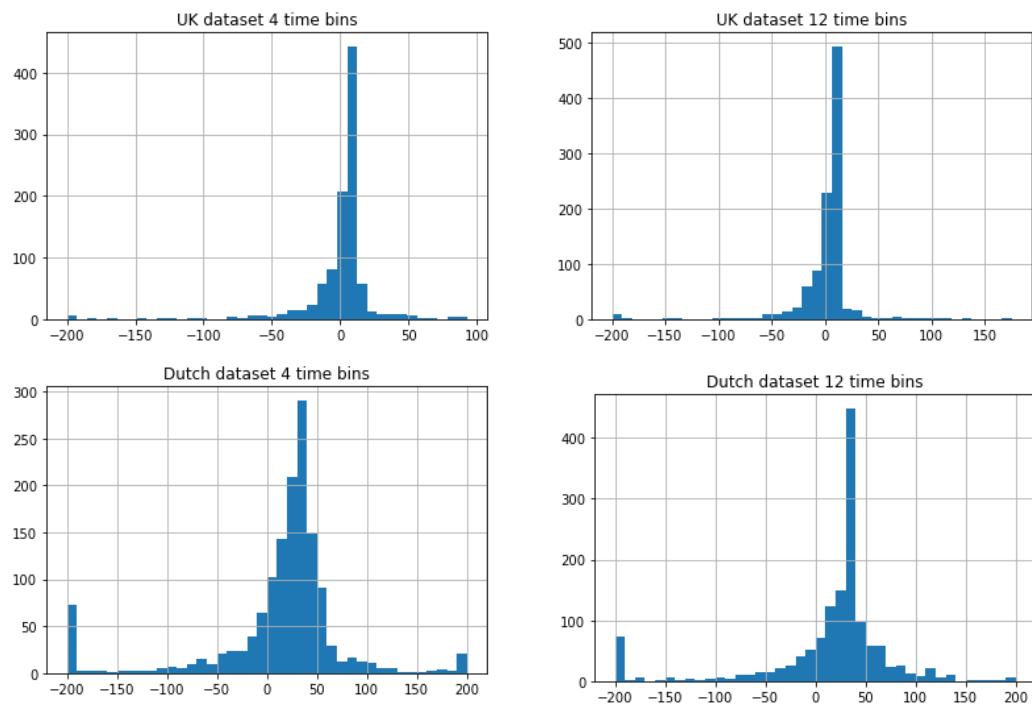


Fig. 7.2 Histogram of residuals for RFreg when using 4 and 12 time bins in the UK and Dutch datasets.

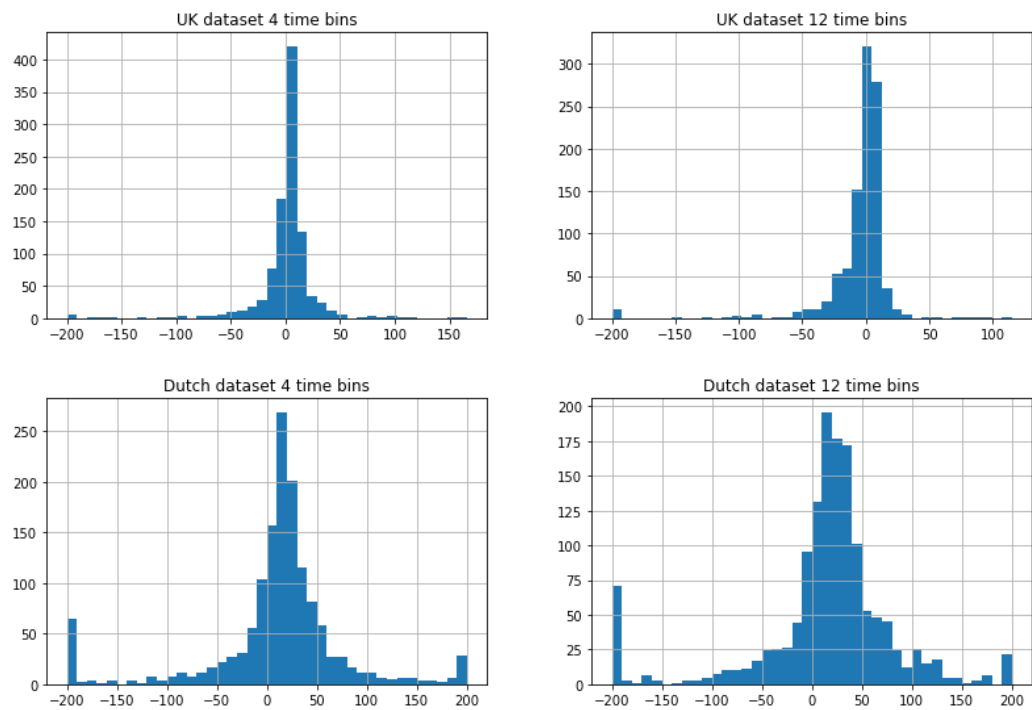


Fig. 7.3 Histogram of residuals for XGBreg with 4 and 12 time bins, for the UK and Dutch datasets.

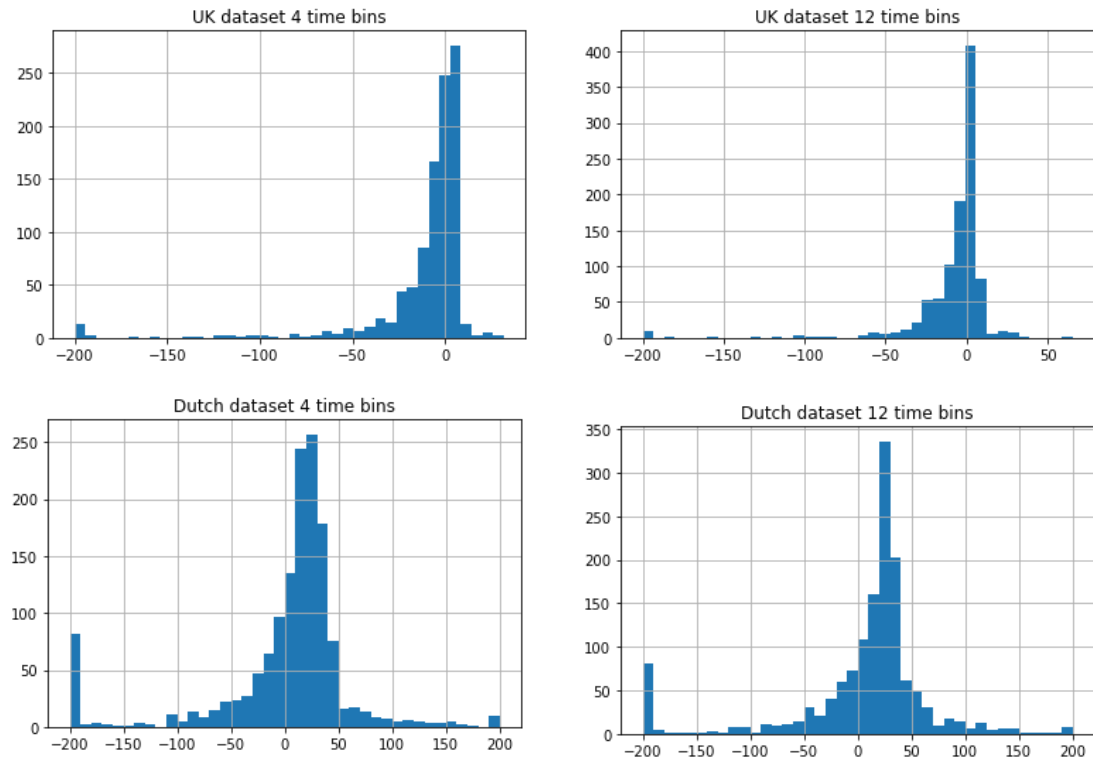


Fig. 7.4 Histogram of residuals for FFNN with 4 and 12 time bins, for the UK and Dutch datasets.

Figure 7.4 shows that the FFNN histogram of residuals was skewed for the UK dataset with 4 time bins. In addition, FFNN residuals for the UK and Dutch dataset with 4 time bins seem to diverge from a normal distribution. However, the residuals distribution was close to normal and less skewed with 12 time bins in both datasets.

The histogram of residuals for LSTM in Figure 7.5 for the UK dataset was less conforming to a normal distribution, compared to other models, including LR. However, the LSTM residuals for the Dutch dataset were clearly normally distributed; the situation here is as good as for FFNN residuals, and possibly ‘better’ compared to the distribution of residuals for LR and XGBreg.

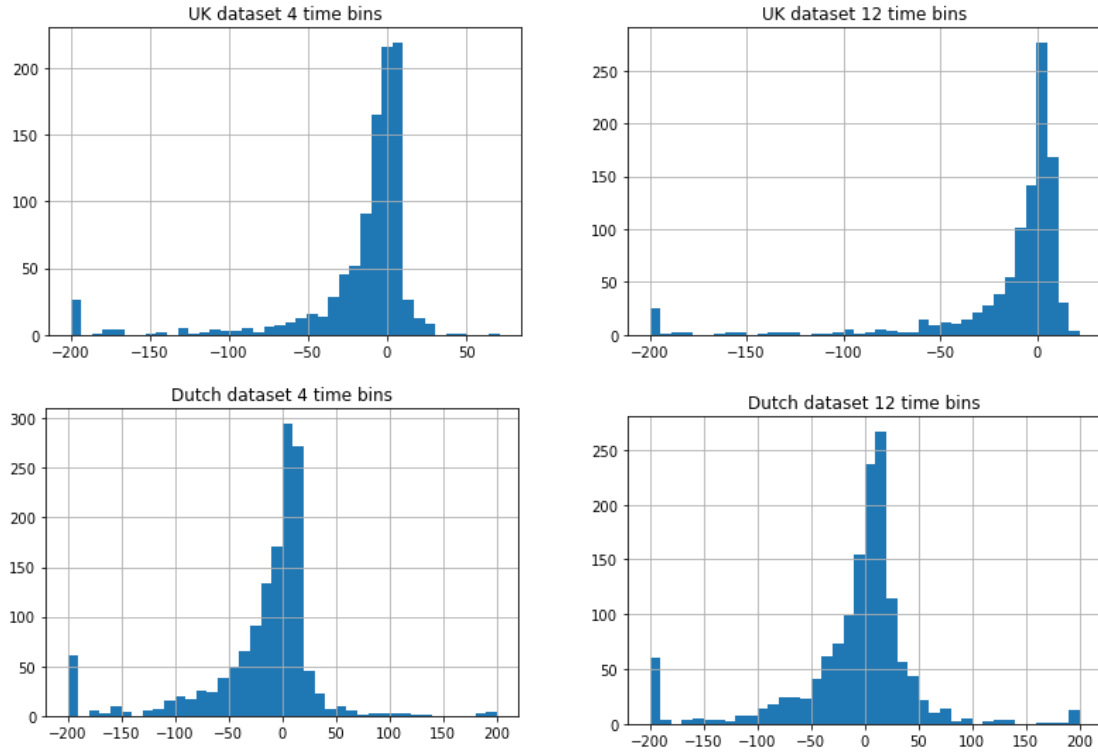


Fig. 7.5 Histogram of residuals for LSTM with 4 and 12 time bins for the UK and Dutch datasets.

In this section, we saw that fixed context models resulted in distributions of residuals with better properties, compared to recurrent models. RFreg residuals were, regarding normality, better compared to other fixed context models (i.e. XGBreg and FFNN).

Overall, the histograms show that with the best models, we can get most predictions within a time window of an hour or less around the actual time, which is encouraging. Interestingly, most models also show a left skew, which again is a desirable property regarding a possible intervention.

7.1.1 Hypothetical Intervention for timing analysis

In this section, we analyse the time of delivering a hypothetical intervention by employing RFreg, as RFreg residuals were, regarding normality, the best (narrower and less skewed). First, the RFreg model constructed in Section 6.5.1 is used for predicting the time until the next unhealthy snack. Then, we determine the Proposed Intervention Time (PIT), calculated as Predicted unhealthy Snacking Time (PST) minus Lead Time (LT), as shown in Equation 6.5.1 is used for predicting the time until the next unhealthy snack. Then, we determine the Proposed Intervention Time (PIT), calculated as Predicted unhealthy Snacking Time (PST)

minus Lead Time (LT) as shown in Equation 7.1. In order to analyse the effect of different intervention timings, we use LT values of 0, 20, 30, 45, 60, 75 and 90 minutes. We then calculate the difference between the Actual Unhealthy snacking Time (AUT) and PIT as seen in Equation 7.2

$$PIT = PST - LT \quad (7.1)$$

$$TD = AUT - PIT \quad (7.2)$$

TD is time difference between the Actual Unhealthy snacking Time (AUT) and PIT. TD is time difference between the Actual Unhealthy snacking Time (AUT) and PIT. Next, we test if the intervention time PIT is within a time window of width W, before the AUT. We use values of width W = 30, 60 or 120 minutes and categorise as follows:

- **Early:** the intervention is delivered **before** the time window ($TD > W$).
- **Hitting:** the intervention is delivered **within** the time window ($TD > 0$ and $TD < W$).
- **Late:** the intervention is delivered **after** the actual unhealthy snacking time ($TD < 0$).

Each of the figures below shows the results of the hypothetical intervention for a specific time window W (i.e. 120, 60 or 30 minutes) and the timing of the hypothetical intervention being **Early**, **Hitting** or **Late**, with respect to the time window, when using different values for LT. Each bar chart in each figure shows the results for a dataset using a different number of time bins, in order to analyse the influence of time bin approach.

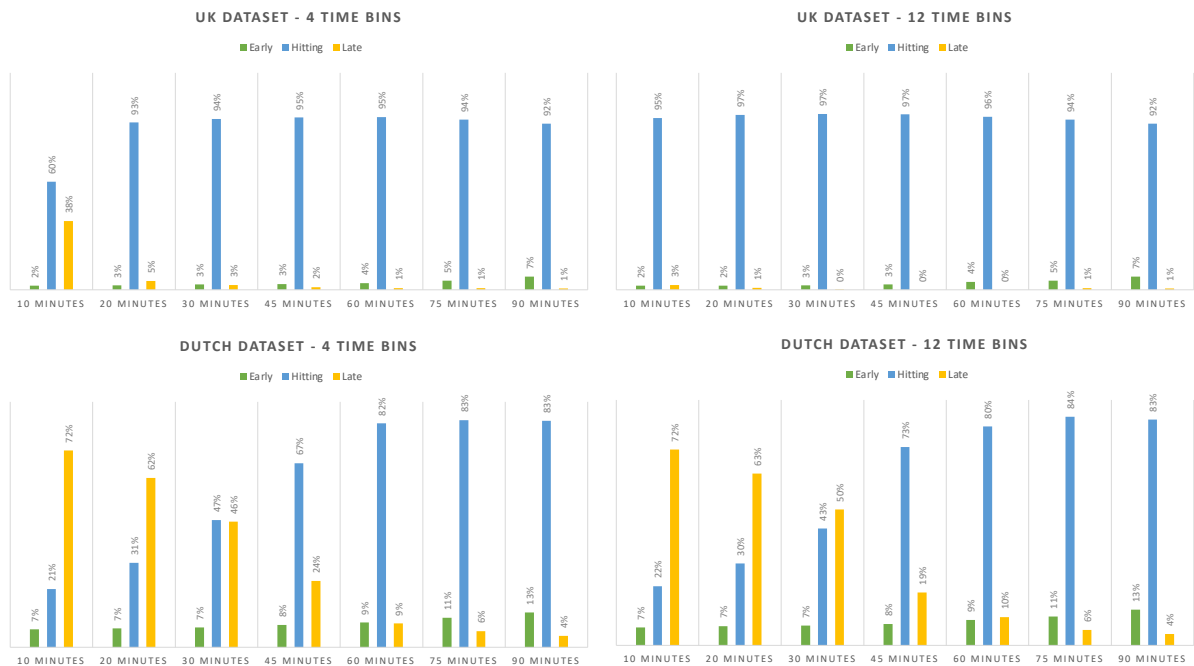


Fig. 7.6 Case distribution of a 120-minute window for accepting the intervention in 4 and 12 time bins in the UK and Dutch datasets. The model used was RFreg.

Figure 7.6 shows the evaluation of the hypothetical interventions using a time window of 120 minutes. For the UK dataset, the percentage of the interventions classified as **hitting** was high, with most LT values, when using both 4 and 12 time bins. In the Dutch dataset, it was high for LT = 60 minutes, with a percentage around 80%, when using both 4 and 12 time bins.



Fig. 7.7 Case distribution for a 60-minute window for accepting the intervention with 4 and 12 time bins in the UK and Dutch datasets. The model used was RFreg.

Figure 7.7 shows the evaluation of the hypothetical interventions in a time window of 60 minutes. The percentage of Hits for the hypothetical interventions was most of the time high in the Dutch dataset with larger LT values, while it was high in the UK with smaller LT values. The **Hitting** percentage reached 97% and 70% in the UK and Dutch dataset respectively.

Figure 7.8 shows the evaluation of the hypothetical interventions in a time window of 30 minutes. It can be observed that the highest percentage of **Hitting** for the UK dataset reached 81%, when X was 10 minutes, when using 12 time bins. By contrast, the highest percentage of **Hitting** for the Dutch dataset was 49%, when X was 45, when using 12 time bins.

These results for a hypothetical intervention show that the **Hitting** percentage varied between the UK and Dutch datasets, with different times X prior to the predicted time of an unhealthy snack. The percentage of **Hitting** was slightly higher when using 12 time bins, except in the Dutch dataset with time window W of 60 minutes, whereas the **Hitting** percentage with 4 time bins was better. The **Hitting** percentage with 4 time bins was better. The **Hitting** percentage for the UK dataset was high in all time windows, although it was higher when the time window was 120 and 60 minutes. For the Dutch dataset, the **Hitting** percentage was the highest when the time window was 120 minutes. There was a trend for this percentage to reduce, as the time window was reduced. The **Hitting** percentage for the

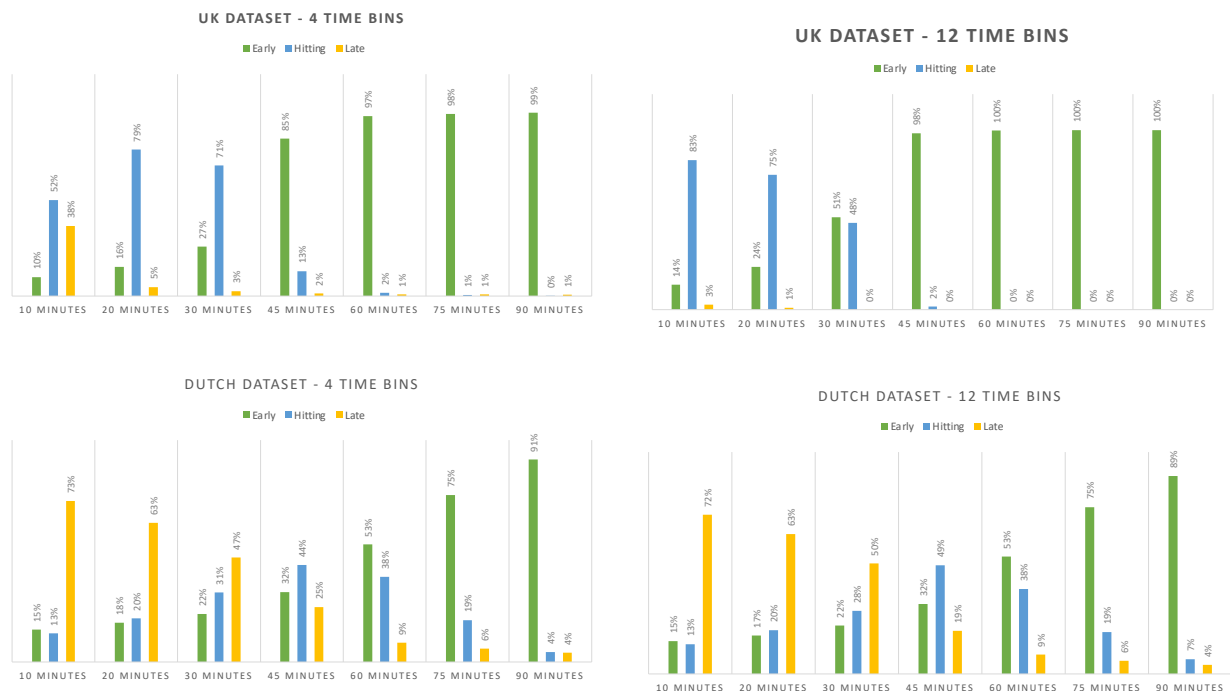


Fig. 7.8 Case distribution with a 30 minutes window for accepting the intervention with 4 and 12 time bins, for the UK and Dutch datasets. The model used was RFR.

Dutch dataset reached 49% when the time window was 30 minutes. This indicates that the number of accurate hypothetical interventions within 30 minutes prior to the consumption of unhealthy snacks is reasonable.

7.2 Discussion

This chapter evaluated the use of one of the explored models for the timing of hypothetical interventions to reduce unhealthy snacking behaviour. With some certainty we can say that, whatever the exact format of an intervention, it would benefit from accurate timing information, regarding unhealthy snacks. The shortest window time we investigated was 30 minutes. For this time window, we think that we were able to deliver a reasonable number of accuracy for hypothetical interventions, before the actual time of unhealthy snacks, with 83% accuracy for the UK and 49% for the Dutch dataset. With shorter time windows (e. g. 20 or 15 minutes), a hypothetical intervention could not be timed reliably enough, especially for the Dutch dataset. Note, the best time for delivering a nudge-based interventions is an unresolved matter, as discussed in the Introduction Interventions very close to or at the same time as unhealthy snacks might come with an increased risk of being delivered after an

unhealthy snack. Therefore, the trade-off between minimising the time gap and delivering an intervention after an unhealthy snack should be considered and regularly evaluated, when building predictive models and designing interventions.

RFreg delivered the best performance in predicting the timing of unhealthy snacks and so this approach is best suited for a hypothetical intervention. This may indicate that fixed context models are generally a good choice in such applications, as they are also easier to train and use. With a larger dataset, LSTM might perform better, which should be examined in future studies, if and when more data becomes available.

Chapter 8

Summary and Conclusion

8.1 Summary

Interventions based on Machine Learning (ML) models can be made more effective by delivering messages at the right time, before a decision for the targeted behaviour is taken. The overarching goal of this thesis was to design an app and use it to collect data on unhealthy snacking behaviour, then to analyse and model the data, including in relation to timing a hypothetical intervention that supports behaviour change.

Initially, two different online datasets, comprising human activities and human room occupancy, were modelled, in order to gain a general understanding of human behaviour and evaluate our methods for predicting human behaviour. In these experiments, the models applied were Hidden Markov Model (HMM), Nonlinear Autoregressive Network with eXogenous inputs (NARX), Neural Networks (NN), Decision Trees (DT) and Random Forest (RF). The results showed that DT, RF and NARX NN performed well in the corresponding tasks, reaching an accuracy of 99%.

Our 'Snack Tracker' app was designed and developed specifically for collecting the snacking data with time and location. Over a period of 4 weeks, 111 participants completed the study in three groups at different periods. All tools used in collecting our 'UK dataset' satisfied all the ethics, privacy, and security considerations of General Data Protection Regulation (GDPR) and/ or City, University London. Most of the messages and feedback received from participants in the survey-based or app-based parts were positive about participation in the study. We collected our own data (the 'UK dataset') the UK. The data collection study was run in two parts: a survey-based part and an app-based part. Initially, 200 participants, aged 18 to 60 years old, took part in the survey-based part, as a self-administered survey, which included parts for demographics, healthy eating motivation, and Three-Factor Eating Questionnaire (TFEQ). 184 participants completed the survey and 170 out of 184 participants

were selected and invited to the app-based part of the study. Participants then used the app for four weeks to record their unhealthy snacks and we obtained their records for our UK dataset.

In addition to the UK dataset, we obtained access to another dataset (the 'Dutch dataset') encompassing general eating behaviour, including unhealthy snacks, from another project in the Netherlands.

The Dutch and UK datasets were pre-processed in order to ensure the quality of the collected data, achieve reliable analysis, and transform the data to a format suitable for machine learning (ML) models. The pre-processing included data cleaning, translation to English, encoding survey questions into scores, one-hot encoding, feature engineering and scaling, and over-sampling. We considered slicing a regular day into 4 and 12 time bins, in the data pre-processing stage, in order to explore the effect of predictive models with small and larger time bins. The pre-processing stage also involved the extraction of healthy vs. unhealthy snacks from the more numerous food categories recorded in the Dutch dataset.

The datasets were then analysed for understanding the behaviour in both datasets and identifying any patterns and trends. The total number of unhealthy snacks recorded in the UK dataset was more than in the Dutch dataset. The mean number of unhealthy snacks per day and per person in each dataset were 2.4 and 0.93 for the UK and Dutch datasets, respectively. This may have been due to the criteria set in selecting the participants in the UK study, whereby we selected subjects both consumed a high number of unhealthy snacks.

In both datasets, female participants showed more variation than male participants, in the number of unhealthy snacks consumed daily. More employees took part in the study than participants with other employment status (e.g., students or home/keepers) in both datasets. Employed, job seekers, overweight, and obese participants consumed more unhealthy snacks than other demographic groups in the UK dataset.

Participants were found to have unhealthy snacks most of the time at home and more in the afternoon than other day time. Moreover, daily number of unhealthy snacks did not vary considerably between weekdays and weekends in both datasets. Tracking behaviour may raise the awareness of unhealthy snacking and subsequently lead to behaviour change even without interventions, and indeed we been found a reduction of snacking over time in both datasets.

An important aspect to consider when delivering a hypothetical intervention for some target behaviour is the time of the intervention. Delivering timed interventions requires a predictive model and ML has been the most successful approach for predictive modelling. We applied Linear Regression (LR), Bayesian Network (BN), Random forest (RF), Extreme Gradient Boosting (XGBoost), Feed Forward Neural Networks (FFNN), Long short-term

memory (LSTM), Random forest regressor (RFreg) and Extreme Gradient Boosting regressor (XGBreg) to the UK and Dutch datasets. The models were employed for predicting an unhealthy snack as present or not in a time bin, predicting the property of the next snack as healthy or unhealthy (for Dutch dataset), and estimating the number of unhealthy snacks in the next time bin. Furthermore, the models were constructed for predicting the time until the next unhealthy snack, as a regression task. Modelling unhealthy snacking behaviour led to the following conclusions:

- The accuracy of predicting the unhealthy snacking (yes or no) when using 4 and 12 in time bins were found to be between 87% and 95% for both datasets (UK and Dutch). XGBoost has provided the best performance in both cases: before and after oversampling.
- The feature importance analysis showed that the dynamic predictors time bin, day of the week, and location were better predictors of unhealthy snacking, than static ones, such as age, gender, employment status and BMI.
- The accuracy of predicting the class of the next snack (healthy or unhealthy) was 70% for LSTM and FFNN models. This analysis was only possible with the Dutch dataset, as in the UK dataset we did not collect information about healthy snacks.
- The accuracy of estimating the number of unhealthy snacks someone would have in the next time bin reached 93% and 86% in the UK and Dutch dataset respectively. Furthermore, including the healthy snacks in the Dutch dataset led to better model performance in estimating the number of unhealthy snacks in the next time bin.
- Oversampling resolved problems with imbalanced classes in the UK dataset. It also reduced some apparent underfitting, when the time bins were 4 and over-fitting when the time bins were 12. Additionally, model accuracy improved from 57% to 74% and between 87% and 93%, with 4 and 12 time bins, respectively.
- Our models were able to predict the time until the next unhealthy snack with a mean absolute error of 17 minutes in the UK dataset and 53 minutes in the Dutch dataset. Additionally, the residuals distribution suggests that some non-linearity was required in making the predictions, in order to obtain good results on both datasets; this was evident by corresponding models associated with more narrow and non-skewed residual distributions. Fixed context models provided better predictions than the recurrent model (i.e. LSTM) on our datasets, although this might change with larger datasets.

Our models were evaluated on the task of timing for the delivery of a hypothetical intervention. The tasks were: predicting an unhealthy snack occurrence in a time bin, predicting the class of the next snack (healthy or unhealthy), and predicting the time until the next unhealthy snack. As the two former tasks were classification, the models were evaluated using several standard metrics, including ROC and precision recall curves. The models for predicting the time until the next unhealthy snack were analysed with hypothetical interventions, by testing whether predicted snacking times fell into time windows of 120, 60 and 30 minutes, before the actual unhealthy snack time. The evaluation of the models for the delivery of a hypothetical unhealthy snacking intervention led to the following conclusions:

- Handling imbalanced classes and determining the best probability thresholds should be taken into account, when building the predictive models.
- Using our models, we were able to achieve timing precision (e.g., for a hypothetical nudge) of within an hour prior to the next unhealthy snack of up to 96% of cases for the UK dataset and up to 70% of cases for the Dutch dataset. We were also able to achieve timing precision of within 30 minutes prior to the next unhealthy snack of up to 83% for the UK dataset and 49% of the Dutch dataset.
- The trade-off between minimising the time gap between a hypothetical intervention and the consumption of an unhealthy snack, and delivering an intervention after an unhealthy snack should be considered and evaluated in any application of timed interventions for behaviour change, as more data becomes available and models are further optimised.

8.2 Thesis contribution

The study of human eating behaviour is of great practical importance, and psychologists have devoted enormous effort to understanding it. An interdisciplinary challenge here is that psychologists, while well-versed in corresponding psychological theory, may lack the technical tools to formally analyse relevant data. Why is formal analysis needed? For once, any theory of eating behaviour assumes that there must be some corresponding structure (e.g., in the pattern of instances of eating). These simple, perhaps obvious, facts have been the starting points of my PhD project.

A surprising realisation is that there has been extremely little work in formal analysis of basic eating behaviour data, such as instances of eating. Spanakis et al. (2017) is the notable exception we could find, and it is worth noting that the research team is a multi-author across many disciplines. The lack of relevant literature may allude to the interdisciplinary challenge

just mentioned above. Therefore, my objective is to collect data for one of the simplest kinds of eating behaviour and rigorously and extensively explore the available structure using a variety of ML algorithms. Spanakis et al. (2017) employed a more elaborate data collection method, requiring participants to offer answers to several questions (e.g., regarding hunger). Their approach is quite valuable. However, we also think that there is merit in examining structures in terms of the distribution of relevant instances without additional variables that might aid prediction. Therefore, we primarily studied just the occurrence of instances of eating unhealthy snacks.

Specifically, we showed the practicality of the data collection adopting our app and providing a rich dataset on individual unhealthy snacking behaviour. We explored several ML models to predict unhealthy snacking from the sparse data, limited, as said, to only events of unhealthy snacking. We identified which ML model works best for predicting unhealthy snacking as the main objective of the thesis. Additionally, we presented a reasonable methodology for how the present approach could be used, in principle, in conjunction with hypothetical interventions.

8.3 Future work

The modelling results are encouraging, and we are confident that the approach explored in this thesis can be useful for many corresponding applications. However, this thesis is only a step towards delivering timed nudges by predictive models. Hence, we discuss many areas to extend this work below.

8.3.1 Data collection

Larger datasets often lead to better predictive models. One way to obtain a larger dataset is to run another app-based data collection period, ideally for more than four weeks, and by recruiting more participants. This would allow the researcher to validate the results across larger time scales, which is clearly a key consideration, and if possible, eventually realise interventions on the back of ML methods for predicting eating behaviour. We could consider gathering more data from individuals because this may also assist in building more personalised predictions and interventions. Additionally, we could include more features about daily eating behaviour in order to improve prediction capability. However, we need to be mindful that more data may reduce the practicality of any app-based method. Additional features could involve other eating behaviours, like main meals, and drinks. In addition, automatically captured geolocation data could improve the predictions of time until the

next unhealthy snacking instance. Moreover, we could try to predict the time of unhealthy snacking just from time bins, rather than from one unhealthy snack to the next. Furthermore, we could capture sensor data from smartwatches (e.g., blood sugar or blood pressure, if and when they become available) as additional features for unhealthy snacking predictions. It would also be useful to model data specifically for participants who consume a high number of daily unhealthy snacks because this group particularly needs behaviour change, and there may also be specific patterns associated with their behaviour.

Another aspect to consider is recruiting more participants with various relevant personal characteristics (such as sensitivity to reward or high external eating) in a way that sampling balances different groups and explores their patterns of unhealthy snacking patterns, as well as model prediction performance for each subset. For example, the Dutch and Uk datasets had few participants in even basic demographic subsets, such as students and the unemployed. This practice (diversity and equal share of different personal characteristics) might lead to the precise modelling of eating behaviour patterns and a better prediction from the models investigated.

8.3.2 Data analysis and predictive models

Despite the general success of neural networks in recent years, we found that more traditional models like XGBoost performed very competitively in predicting next snack class and the number of unhealthy snacks. These models performed better than FFNN and LSTM. This picture may change when the number of data points increases, as neural networks, particularly deep ones, benefit greatly from large datasets. Note, however, that any practical application along such lines would be inevitably limited to modest datasets at best because there is a limit to how many observations we can expect from any individual participant. An alternative to LSTM is Gated Recurrent Unit (GRU) because it is faster, requires less memory, and sometimes produces better performance than LSTM (ArunKumar et al., 2022).

Although, our current work does not seem to suggest age and gender as important factors, more analysis needs to be conducted concerning these factors to investigate their roles in unhealthy snacking analysis and prediction. Imbalanced classes may need to be handled by oversampling because this was shown to improve model performance regarding sensitivity, specificity, and precision thresholds. Because our current work lacks an independent validation set in the predictive models, due to the limited size of the datasets, this could be an area of importance to be investigated and addressed in the future.

One could apply transfer learning by employing both datasets (UK and Dutch datasets) together to optimise model performance. This might be useful, especially when predicting the time until the next unhealthy snack because it is expected to lead to benefits, such as

shortening the gap between a hypothetical intervention and target behaviour. Transfer learning could also assist in developing personalised, predictive models by building models on data from multiple individuals and training them to predict the unhealthy snacking behaviour of a particular individuals.

Another interesting approach is online incremental ML, which refers to situations in which continuous data streams are used to update a model. Online ML becomes essential in interactive situations where training examples are obtained based on human responses over time (He et al., 2020). Furthermore, online ML is typically beneficial in application scenarios such as learning in changing settings, as well as providing an efficient scheme for big data processing due to its sequential processing (He et al., 2020). Thus, online ML has the potential to substantially improve the predictions of unhealthy snacking.

Multi-task learning could be an alternative approach in which a single model is built to simultaneously learn these multiple goals and tasks. Multi-task learning has been successfully employed in many real-world applications, for instance, recommender systems. It could progress the performance of the multiple predictive tasks by learning them jointly. However, the prediction quality of the multi-task models could be influenced sometimes by the relationships between tasks (Ma et al., 2018).

8.3.3 Interventions

One could extend the app development to add some nudges/ interventions based on the predictive models developed in this project. Specific interventions for unhealthy snacking behaviour could take a variety of forms (e.g., written or audio messages). We should evaluate the trade-off between timing interventions close to unhealthy snacks and the risk of sending late interventions. Our current approach could be implemented as three persuasive technology strategies from Cialdini (2009): tracking/self-monitoring, reminders, and suggestions. We suggest implementing other strategies of persuasive technologies, such as, goal setting, rewards, and social support strategies, and then evaluate their effectiveness on unhealthy snacking behaviour. Additionally, we could try designing our interventions to prompt people to shift towards engaging in reflective processes (i.e. exercise self-control). Of course, more generally, we should evaluate any specific sort of intervention within the context of a randomised control trial, as is standard in health and clinical applications. Moreover, users' opinions and evaluations need to be considered with specific interventions for unhealthy snacking behaviour, as such feedback might affect the future design of the interventions.

8.3.4 Other relevant human behaviours

It could be an interesting future direction to extend the present approach to other human behaviours, with the potential for improvements from small changes or prompts, such as, smoking cessation, reducing alcohol intake, exercising regularly, and driving safely. Our proposed method are adaptable to other human behaviours with small or moderate changes.

We could also explore external validation in various settings to assess the reproducibility and generalisability of the models. It is worth noting that ensuring the reproducibility of predictive models should be done with some care. This is because, in principle, the predictive models could be used for promoting, rather than reducing, negative behaviours (i.e. promoting unhealthy snacking or smoking behaviour).

Overall, exploring app development, data collection and analysis, and ML for understanding a target behaviour has been an exciting journey and will hopefully lead to applications that will help improve people's lives.

Bibliography

- Albon, C. (2018). Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. O'Reilly.
- Aldenaini, N., Alqahtani, F., Orji, R., and Sampalli, S. (2020). Trends in persuasive technologies for physical activity and sedentary behavior: a systematic review. Frontiers in artificial intelligence, 3:7.
- Ali, J., Khan, R., Ahmad, N., and Maqsood, I. (2012). Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5):272.
- Almoraie, N. M., Saqaan, R., a. Alharthi, R., Alamoudi, A., Badh, L., and Shatwan, I. M. (2021). Snacking patterns throughout the life span: potential implications on health. Nutrition research, 91(7):81–94.
- Alpaydin, E. (2013). Introduction to Machine Learning, 2nd Edition. The MIT Press.
- Alshurafa, N., Kalantarian, H., Pourhomayoun, M., Liu, J. J., Sarin, S., Shahbazi, B., and Sarrafzadeh, M. (2015). Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor. IEEE sensors journal, 15(7):3909–3916.
- Altmeyer, M., Lessel, P., Hosseini, S., and Krueger, A. (2018). Silvercycling: evaluating persuasive strategies to promote physical activity among older adults. In Proceedings of the 2018 acm conference companion publication on designing interactive systems, pages 45–50.
- Anderson-Bill, E. S., Winett, R. A., Wojcik, J. R., and Winett, S. G. (2011). Web-based guide to health: relationship of theoretical variables to change in physical activity, nutrition and weight at 16-months. Journal of medical Internet research, 13(1):e27.
- Aoun, C., Nassar, L., Soumi, S., El Osta, N., Papazian, T., and Rabbaa Khabbaz, L. (2019). The cognitive, behavioral, and emotional aspects of eating habits and association with

- impulsivity, chronotype, anxiety, and depression: a cross-sectional study. Frontiers in behavioral neuroscience, 13(6):204.
- Armitage, C. J. (2015). Time to retire the theory of planned behaviour? a commentary on sniehotta, presseau and araújo-soares. Health Psychology Review, 9(2):151–155.
- Arno, A. and Thomas, S. (2016). The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. BMC public health, 16(1):1–11.
- ArunKumar, K., Kalaga, D. V., Mohan Sai Kumar, C., Kawaji, M., and Brenza, T. M. (2022). Comparative analysis of gated recurrent units (gru), long short-term memory (lstm) cells, autoregressive integrated moving average (arima), seasonal autoregressive integrated moving average (sarima) for forecasting covid-19 trends. Alexandria Engineering Journal, 61(10):7585–7603.
- Ashbrook, D. and Starner, T. (2002). Learning significant locations and predicting user movement with GPS. In 6th International Symposium on Wearable Computers (ISWC 2002), 7-10 October 2002, Seattle, WA, USA, pages 101–108. IEEE Computer Society.
- Asimakopoulos, S., Asimakopoulos, G., and Spillers, F. (2017). Motivation and user engagement in fitness tracking: Heuristics for mobile healthcare wearables. Informatics, 4(1):5.
- Aswani, A., Kaminsky, P., Mintz, Y., Flowers, E., and Fukuoka, Y. (2019). Behavioral modeling in weight loss interventions. European journal of operational research, 272(3):1058–1072.
- Banna, J. C., Panizza, C. E., Boushey, C. J., Delp, E. J., and Lim, E. (2018). Association between cognitive restraint, uncontrolled eating, emotional eating and bmi and the amount of food wasted in early adolescent girls. Nutrients, 10(9):1279.
- Berry, K. J., Mielke Jr, P. W., and Iyer, H. K. (1998). Factorial designs and dummy coding. Perceptual and motor skills, 87(3):919–927.
- Bhandari, M. and Joensson, A. (2011). Clinical Research for Surgeons. Thieme electronic-book library. Thieme.
- Biglan, A., Hayes, S., and Wilson, D. (2015). The Nurture Effect: How the Science of Human Behavior Can Improve Our Lives and Our World. New Harbinger Publications.
- Bilinski, P. and Bremond, F. (2011). Evaluation of local descriptors for action recognition in videos. In International Conference on Computer Vision Systems, pages 61–70. Springer.

- Bishop, C. M. (2007). Pattern recognition and machine learning, 5th Edition. Information science and statistics. Springer, 1 edition.
- Biswas, M., Kuppili, V., Saba, L., Edla, D. R., Suri, H. S., Cuadrado-Godia, E., Laird, J. R., Marinho, R. T., Sanches, J. M., Nicolaides, A., et al. (2019). State-of-the-art review on deep learning in medical imaging. Front Biosci (Landmark Ed), 24(3):392–426.
- Bochud, M., Chatelan, A., Blanco, J.-M., and Beer-Borst, S. (2017, publisher=The Federal Office of Public Health and the Food Safety and Veterinary Office). Anthropometric characteristics and indicators of eating and physical activity behaviors in the swiss adult population: results from menuch 2014-2015.
- Bonaccorso, G. (2018). Machine Learning Algorithms: Popular algorithms for data science and machine learning. Packt Publishing Ltd, 2 edition.
- Bregler, C. (1997). Learning and recognizing human dynamics in video sequences. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 568–574. IEEE.
- Brownlee, J. (2020). Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python. Machine Learning Mastery.
- Cai, C. W. (2020). Nudging the financial market? a review of the nudge theory. Accounting & Finance, 60(4):3341–3365.
- Cappelleri, J., Bushmakin, A., Gerber, R., Leidy, N., Sexton, C., Lowe, M., and Karlsson, J. (2009). Psychometric analysis of the three-factor eating questionnaire-r21: results from a large diverse sample of obese and non-obese participants. International journal of obesity, 33(6):611–620.
- Castmo, M. and Persson, R. (2018). The alliance of digital nudging & persuasive design: The complementary nature of the design strategies. Master's thesis, Economics and Management.
- Cerda, P., Varoquaux, G., and Kégl, B. (2018). Similarity encoding for learning with dirty categorical variables. Machine Learning, 107(8):1477–1494.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16(1):321–357.

- Cheng, G., Wan, Y., Saudagar, A. N., Namuduri, K., and Buckles, B. P. (2015). Advances in human action recognition: A survey. CoRR, abs/1501.05964.
- Chew, H. S. J., Ang, W. H. D., and Ying, L. (2021). The potential of artificial intelligence in enhancing adult weight loss: a scoping review. Public health nutrition, 24(8):1–59.
- Choujaa, D. and Dulay, N. (2010). Predicting human behaviour from selected mobile phone data points. In Proceedings of the 12th ACM international conference on Ubiquitous computing, pages 105–108.
- Cialdini, R. (2009). Influence: The Psychology of Persuasion. Collins Business Essentials. HarperCollins e-books.
- Cleobury, L. and Tapper, K. (2014). Reasons for eating ‘unhealthy’ snacks in overweight and obese males and females. Journal of Human Nutrition and Dietetics, 27(4):333–341.
- Cohen, M. (2021). Hello heart makes it simple and fun to manage your blood pressure. <https://www.helloheart.com/for-individuals>. Retrieved 15 Aug 2021.
- Consolvo, S., Everitt, K., Smith, I., and Landay, J. A. (2006). Design requirements for technologies that encourage physical activity. In Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 457–466.
- Cook, J. and Ramadas, V. (2020). When to consult precision-recall curves. The Stata Journal, 20(1):131–148.
- de la Torre Díez, I., Garcia-Zapirain, B., López-Coronado, M., Rodrigues, J. J., and del Pozo Vegas, C. (2017). A new mhealth app for monitoring and awareness of healthy eating: development and user evaluation by spanish users. Journal of medical systems, 41(7):1–7.
- Dix, A., Finlay, J., and Beale, R. (1992). Analysis of user behaviour as time series. PEOPLE AND COMPUTERS, pages 429–429.
- Dunford, E. K. and Popkin, B. M. (2017). Disparities in snacking trends in us adults over a 35 year period from 1977 to 2012. Nutrients, 9(8):809.
- Dunneram, Y., Ramasawmy, D., Pugo-Gunsam, P., and Jeewon, R. (2013). Determinants of eating habits among pre-retired and post-retired mauritians. Int J Food Sci Nutr, 2(3):109–115.

- Edney, S., Plotnikoff, R., Vandelanotte, C., Olds, T., De Bourdeaudhuij, I., Ryan, J., and Maher, C. (2017). “active team” a social and gamified app-based physical activity intervention: randomised controlled trial study protocol. BMC public health, 17(1):1–10.
- Fang, H., Xu, H., Yuan, H., and Zhou, Y. (2017). Discretization of continuous variables in bayesian networks based on matrix decomposition. In 2017 International Conference on Computing Intelligence and Information System (CIIS), pages 184–187. IEEE.
- Fatima, G. and Saeed, S. (2021). A novel weighted ensemble method to overcome the impact of under-fitting and over-fitting on the classification accuracy of the imbalanced data sets. Pakistan Journal of Statistics and Operation Research, 17(2):483–496.
- Fenicio, A. and Calvary, G. (2015). Persuasion through an ambient device: proof of concept and early evaluation of cregrette, a smoking cessation system. In European Conference on Ambient Intelligence, pages 252–267. Springer.
- Finger, J. D., Tylleskär, T., Lampert, T., and Mensink, G. B. (2013). Dietary behaviour and socioeconomic position: the role of physical activity patterns. PLoS One, 8(11):e78390.
- Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. Morgan Kaufmann.
- Fogg, B. J. (2009a). The behavior grid: 35 ways behavior can change. In Proceedings of the 4th international Conference on Persuasive Technology, pages 1–5.
- Fogg, B. J. (2009b). Creating persuasive technologies: an eight-step design process. In Proceedings of the 4th international conference on persuasive technology, pages 1–6.
- Ford, A. L., Bergh, C., Södersten, P., Sabin, M. A., Hollinghurst, S., Hunt, L. P., and Shield, J. P. (2010). Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. Bmj, 340(7740):250–259.
- Forman, E. M., Goldstein, S. P., Crochiere, R. J., Butryn, M. L., Juarascio, A. S., Zhang, F., and Foster, G. D. (2019a). Randomized controlled trial of ontrack, a just-in-time adaptive intervention designed to enhance weight loss. Translational behavioral medicine, 9(6):989–1001.
- Forman, E. M., Goldstein, S. P., Zhang, F., Evans, B. C., Manasse, S. M., Butryn, M. L., Juarascio, A. S., Abichandani, P., Martin, G. J., and Foster, G. D. (2019b). Ontrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. Translational behavioral medicine, 9(2):236–245.

- Franklin, V. L., Waller, A., Pagliari, C., and Greene, S. A. (2006). A randomized controlled trial of sweet talk, a text-messaging system to support young people with diabetes. Diabetic medicine, 23(12):1332–1338.
- Fuller, W. A. (1986). Measurement Error Models. John Wiley Sons, Inc., USA.
- Fuller-Tyszkiewicz, M., Richardson, B., Klein, B., Skouteris, H., Christensen, H., Austin, D., Castle, D., Mihalopoulos, C., O'Donnell, R., Arulkadacham, L., et al. (2018). A mobile app-based intervention for depression: End-user and expert usability testing study. JMIR mental health, 5(3):e9445.
- Gambs, S. (2018). Privacy and ethical challenges in big data. In International symposium on foundations and practice of security, pages 17–26. Springer.
- Ganesan, A. N., Louise, J., Horsfall, M., Bilsborough, S. A., Hendriks, J., McGavigan, A. D., Selvanayagam, J. B., and Chew, D. P. (2016). International mobile-health intervention on physical activity, sitting, and weight: the stepathlon cardiovascular health study. Journal of the American College of Cardiology, 67(21):2453–2463.
- Genena, D. M. and Salama, A. A. (2017). Obesity and eating habits among university students in alexandria, egypt: a cross sectional study. World journal of nutrition and health, 5(3):62–68.
- Gillitzer, C. and Sinning, M. (2020). Nudging businesses to pay their taxes: Does timing matter? Journal of Economic Behavior & Organization, 169(C):284–300.
- Giudici, P. (2005). Applied Data Mining: Statistical Methods for Business and Industry. Wiley.
- Glanz, K. and Bishop, D. B. (2010). The role of behavioral science theory in development and implementation of public health interventions. Annual review of public health, 31(1):399–418.
- Goldstein, S. P., Brick, L. A., Thomas, J. G., and Forman, E. M. (2021). Examination of the relationship between lapses and weight loss in a smartphone-based just-in time adaptive intervention. Translational behavioral medicine, 11(4):993–1005.
- Goldstein, S. P., Zhang, F., Thomas, J. G., Butryn, M. L., Herbert, J. D., and Forman, E. M. (2018). Application of machine learning to predict dietary lapses during weight loss. Journal of diabetes science and technology, 12(5):1045–1052.

- Gordon, M. and Kochen, M. (1989). Recall-precision trade-off: A derivation. Journal of the American Society for Information Science, 40(3):145–151.
- Gupta, R. and Pathak, C. (2014). A machine learning framework for predicting purchase by online customers based on dynamic pricing. Procedia Computer Science, 36(10):599–605.
- Haque, M. S., Isomursu, M., Kangas, M., and Jämsä, T. (2018). Measuring the influence of a persuasive application to promote physical activity. In PPT@ PERSUASIVE, pages 43–57.
- Haththotuwa, R. N., Wijeyaratne, C. N., and Senarath, U. (2020). Worldwide epidemic of obesity. In Obesity and obstetrics, pages 3–8. Elsevier.
- He, J., Mao, R., Shao, Z., and Zhu, F. (2020). Incremental learning in online scenario. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13926–13935.
- Heer, J., Hellerstein, J. M., and Kandel, S. (2015). Predictive interaction for data transformation. In Seventh Biennial Conference on Innovative Data Systems Research, CIDR 2015, Asilomar, CA, USA, January 4-7, 2015, Online Proceedings. www.cidrdb.org.
- Hellerstein, J. M. (2008). Quantitative data cleaning for large databases. United Nations Economic Commission for Europe (UNECE).
- Hofmann, W., Friese, M., and Wiers, R. W. (2008). Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. Health Psychology Review, 2(2):111–137.
- Hu, Y., Cao, L., Lv, F., Yan, S., Gong, Y., and Huang, T. S. (2009). Action detection in complex scenes with spatial and temporal ambiguities. In 2009 IEEE 12th International Conference on Computer Vision, pages 128–135. IEEE.
- Huang, Y.-P. and Yen, M.-F. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. Applied Soft Computing, 83(10):105663.
- Hunter, C. M., Peterson, A. L., Alvarez, L. M., Poston, W. C., Brundige, A. R., Haddock, C. K., Van Brunt, D. L., and Foreyt, J. P. (2008). Weight management using the internet: a randomized controlled trial. American journal of preventive medicine, 34(2):119–126.

- Hurling, R., Catt, M., De Boni, M., Fairley, B., Hurst, T., Murray, P., Richardson, A., and Sodhi, J. (2007). Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. Journal of medical Internet research, 9(2):e7.
- Hyndman, R. and Athanasopoulos, G. (2014). Forecasting: principles and practice. OTexts.
- Iliou, T., Anagnostopoulos, C.-N., Nerantzaki, M., and Anastassopoulos, G. (2015). A novel machine learning data preprocessing method for enhancing classification algorithms performance. In Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS), pages 1–5.
- Intille, S. S., Kukla, C., Farzanfar, R., and Bakr, W. (2003). Just-in-time technology to encourage incremental, dietary behavior change. In AMIA Annual Symposium Proceedings, volume 2003, page 874. American Medical Informatics Association.
- Jebara, T. (1998). Action-reaction learning: Analysis and synthesis of human behaviour. PhD thesis, Massachusetts Institute of Technology.
- Jensen, F. V. (1998). An introduction to bayesian networks, UCL press, 1996, pp 178, ISBN 1-85728-332-5. Knowl. Eng. Rev., 13(2):201–208.
- Jin, D., Halvari, H., Maehle, N., and Olafsen, A. H. (2022). Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. Behaviour & Information Technology, pages 1–20.
- John, P., Cotterill, S., Richardson, L., Moseley, A., Stoker, G., Wales, C., and Smith, G. (2011). Nudge, Nudge, Think, Think: Experimenting with Ways to Change Civic Behaviour. Bloomsbury Publishing.
- Joshi, A., Kale, S., Chandel, S., and Pal, D. K. (2015). Likert scale: Explored and explained. Current Journal of Applied Science and Technology, 7(4):396–403.
- Jović, A., Brkić, K., and Bogunović, N. (2012). Decision tree ensembles in biomedical time-series classification. In Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium, pages 408–417. Springer.
- Kant, A. K. and Graubard, B. I. (2015). 40-year trends in meal and snack eating behaviors of american adults. Journal of the Academy of Nutrition and Dietetics, 115(1):50–63.

- Karim, M. R., Cochez, M., Beyan, O., Decker, S., and Lange, C. (2019). Onconetexplainer: explainable predictions of cancer types based on gene expression data. In 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pages 415–422. IEEE.
- Karim, N. S. A., AlHarbi, A., AlKadhi, B., and AlOthaim, N. (2017). Mobile application on smoking cessation based on persuasive design theory. In Alias, R. A., Ling, P. S., Bahri, S., Finnegan, P., and Sia, C. L., editors, 21st Pacific Asia Conference on Information Systems, PACIS 2017, Langkawi, Malaysia, July 16-20, 2017, page 179.
- Karki, M., Basu, S., DiBiano, R., Mukhopadhyay, S., Weltman, J., and Stagg, M. (2016). A symbolic framework for recognizing activities in full motion surveillance videos. In 2016 IEEE Symposium Series on Computational Intelligence (SSCI), pages 1–7. IEEE.
- Karlsson, J., Persson, L.-O., Sjöström, L., and Sullivan, M. (2000). Psychometric properties and factor structure of the three-factor eating questionnaire (tfeq) in obese men and women. results from the swedish obese subjects (sos) study. International journal of obesity, 24(12):1715–1725.
- Kasl, S. V. and Cobb, S. (1966). Health behavior, illness behavior and sick role behavior: I. health and illness behavior. Archives of Environmental Health: An International Journal, 12(2):246–266.
- Kelly, M. P. and Barker, M. (2016). Why is changing health-related behaviour so difficult? Public health, 136:109–116.
- Kerr, D. A., Harray, A. J., Pollard, C. M., Dhaliwal, S. S., Delp, E. J., Howat, P. A., Pickering, M. R., Ahmad, Z., Meng, X., Pratt, I. S., et al. (2016). The connecting health and technology study: a 6-month randomized controlled trial to improve nutrition behaviours using a mobile food record and text messaging support in young adults. International Journal of Behavioral Nutrition and Physical Activity, 13(1):1–14.
- Kim, P. (2017). Deep learning. In MATLAB Deep Learning, pages 103–120. Springer.
- Kondracki, N. L., Wellman, N. S., and Amundson, D. R. (2002). Content analysis: Review of methods and their applications in nutrition education. Journal of nutrition education and behavior, 34(4):224–230.
- Kong, Y. and Fu, Y. (2018). Human action recognition and prediction: A survey. Int J Comput Vis, page 1366–1401.

- Kuhn, M. and Johnson, K. (2013). Data pre-processing. In Applied predictive modeling, pages 27–59. Springer.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553):436–444.
- Lee, L. T., Bowen, P. G., Mosley, M. K., and Turner, C. C. (2017). Theory of planned behavior: social support and diabetes self-management. The Journal for Nurse Practitioners, 13(4):265–270.
- Li, T., Haines, J. K., de Eguino, M. F. R., Hong, J. I., and Nichols, J. (in press). Alert now or never: Understanding and predicting notification preferences of smartphone users. ACM Trans. Comput.-Hum. Interact.
- Liciotti, D., Bernardini, M., Romeo, L., and Frontoni, E. (2020). A sequential deep learning application for recognising human activities in smart homes. Neurocomputing, 396(26):501–513.
- Liu, H., Fang, T., Zhou, T., and Wang, L. (2018). Towards robust human-robot collaborative manufacturing: multimodal fusion. IEEE Access, 6(1):74762–74771.
- Liu, H. and Wang, L. (2021). Collision-free human-robot collaboration based on context awareness. Robotics Comput. Integr. Manuf., 67(1):101997.
- Liu, X., Zheng, C., Xu, C., Liu, Q., Wang, J., Hong, Y., and Zhao, P. (2017). Nighttime snacking is associated with risk of obesity and hyperglycemia in adults: A cross-sectional survey from chinese adult teachers. Journal of Biomedical Research, 31(6):541.
- Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., and Chi, E. H. (2018). Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1930–1939.
- Mahmoud, S., Lotfi, A., and Langensiepen, C. (2013). Behavioural pattern identification and prediction in intelligent environments. Applied Soft Computing, 13(4):1813–1822.
- Mazhar, O., Navarro, B., Ramdani, S., Passama, R., and Cherubini, A. (2019). A real-time human-robot interaction framework with robust background invariant hand gesture detection. Robotics and Computer-Integrated Manufacturing, 60(1):34–48.
- McBride, C. M., Emmons, K. M., and Lipkus, I. M. (2003). Understanding the potential of teachable moments: the case of smoking cessation. Health education research, 18(2):156–170.

- Medsker, L. and Jain, L. C. (1999). Recurrent neural networks: design and applications. CRC Press, Inc., USA, 1st edition.
- Mekruksavanich, S. and Jitpattanakul, A. (2021). Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. Sensors, 21(5):1636.
- Miller, A. D. and Mynatt, E. D. (2014). Stepstream: a school-based pervasive social fitness system for everyday adolescent health. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 2823–2832.
- Monekosso, D. N. and Remagnino, P. (2009). Synthetic training data generation for ADL modeling. In Intelligent Environments (Workshops), volume 4 of Ambient Intelligence and Smart Environments, pages 137–144. IOS Press.
- Moore, G., Galway, L., and Donnelly, M. (2017). Remember to smile: design of a mobile affective technology to help promote individual happiness through smiling. In Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, pages 348–354.
- Muneer, A., Fati, S. M., and Fuddah, S. (2020). Smart health monitoring system using iot based smart fitness mirror. Telkomnika, 18(1):317–331.
- Murad, A. and Pyun, J.-Y. (2017). Deep recurrent neural networks for human activity recognition. Sensors, 17(11):2556.
- Murakami, K. and Livingstone, M. B. E. (2016). Associations between meal and snack frequency and diet quality in us adults: National health and nutrition examination survey 2003-2012. Journal of the Academy of Nutrition and Dietetics, 116(7):1101–1113.
- Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., and Turaga, D. S. (2017). Learning feature engineering for classification. In Ijcai, pages 2529–2535.
- NHS (2020). Statistics on obesity, physical activity and diet, england. NHS. Retrieved 27 Dec 2021.
- Nuzzi, C., Pasinetti, S., Pagani, R., Ghidini, S., Beschi, M., Coffetti, G., and Sansoni, G. (2021). Meguru: a gesture-based robot program builder for meta-collaborative workstations. Robotics and Computer-Integrated Manufacturing, 68(2):102085.
- Obermayer, J. L., Riley, W. T., Asif, O., and Jean-Mary, J. (2004). College smoking-cessation using cell phone text messaging. Journal of American College Health, 53(2):71–78.

- Ockene, J. K., Ma, Y., Zapka, J. G., Pbert, L. A., Goins, K. V., and Stoddard, A. M. (2002). Spontaneous cessation of smoking and alcohol use among low-income pregnant women. American journal of preventive medicine, 23(3):150–159.
- Oinas-Kukkonen, H. and Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. Commun. Assoc. Inf. Syst., 24(1):28.
- Oyebode, O., Graham-Kalio, B., and Orji, R. (2020). Hearthealth: A persuasive mobile app for mitigating the risk of ischemic heart disease. In International Conference on Persuasive Technology, pages 126–138. Springer.
- Oyebode, O., Ndulue, C., Mulchandani, D., A. Zamil Adib, A., Alhasani, M., and Orji, R. (2021). Tailoring persuasive and behaviour change systems based on stages of change and motivation. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–19.
- Paay, J., Kjeldskov, J., Skov, M. B., Lichon, L., and Rasmussen, S. (2015). Understanding individual differences for tailored smoking cessation apps. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 1699–1708.
- Paldán, K., Simanovski, J., Ullrich, G., Steinmetz, M., Rammos, C., Jánosi, R. A., Moebus, S., Rassaf, T., Lortz, J., et al. (2019). Feasibility and clinical relevance of a mobile intervention using trackpad to support supervised exercise therapy in patients with peripheral arterial disease: study protocol for a randomized controlled pilot trial. JMIR research protocols, 8(6):e13651.
- Papandreou, G., Zhu, T., Chen, L.-C., Gidaris, S., Tompson, J., and Murphy, K. (2018). Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. In Proceedings of the European conference on computer vision (ECCV), pages 269–286.
- Peng, L. and Lei, L. (2005). A review of missing data treatment methods. Intell. Inf. Manag. Syst. Technol, 1:412–419.
- Pentland, A. and Liu, A. (1999). Modeling and prediction of human behavior. Neural computation, 11(1):229–242.
- perfect Diet Tracker (n.d.). perfect diet tracker. <https://perfectdiettracker.com/>. Retrieved 15 Aug 2021.

- Piernas, C. and Popkin, B. M. (2010). Snacking increased among us adults between 1977 and 2006. The Journal of nutrition, 140(2):325–332.
- Pinder, C., Vermeulen, J., Cowan, B. R., and Beale, R. (2018). Digital behaviour change interventions to break and form habits. ACM Transactions on Computer-Human Interaction (TOCHI), 25(3):1–66.
- priyadarsini, P. I. (2021). Abc-bsrf: Artificial bee colony and borderline-smote rf algorithm for intrusion detection system on data imbalanced problem. In Proceedings of International Conference on Computational Intelligence and Data Engineering: ICCIDE 2020, pages 15–29. Springer.
- Purohit, A. K. and Holzer, A. (2019a). Functional digital nudges: Identifying optimal timing for effective behavior change. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pages 1–6.
- Purohit, A. K. and Holzer, A. (2019b). Functional digital nudges: Identifying optimal timing for effective behavior change. In Mandryk, R. L., Brewster, S. A., Hancock, M., Fitzpatrick, G., Cox, A. L., Kostakos, V., and Perry, M., editors, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019. ACM.
- Purpura, S., Schwanda, V., Williams, K., Stubler, W., and Sengers, P. (2011). Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In Tan, D. S., Amershi, S., Begole, B., Kellogg, W. A., and Tungare, M., editors, Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, BC, Canada, May 7-12, 2011, pages 423–432. ACM.
- Razmjoo, A., Xanthopoulos, P., and Zheng, Q. P. (2017). Online feature importance ranking based on sensitivity analysis. Expert Systems with Applications, 85(19):397–406.
- Reitermanova, Z. (2010). Data splitting. In WDS, volume 10, pages 31–36.
- Renner, B., Sproesser, G., Strohbach, S., and Schupp, H. T. (2012). Why we eat what we eat. the eating motivation survey (tems). Appetite, 59(1):117–128.
- Robinson, S., Perkins, S., Bauer, S., Hammond, N., Treasure, J., and Schmidt, U. (2006). Aftercare intervention through text messaging in the treatment of bulimia nervosa—feasibility pilot. International Journal of Eating Disorders, 39(8):633–638.

- Rothert, K., Strecher, V. J., Doyle, L. A., Caplan, W. M., Joyce, J. S., Jimison, H. B., Karm, L. M., Mims, A. D., and Roth, M. A. (2006). Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity, 14(2):266–272.
- Russell, S. J. and Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th Edition). Pearson.
- Saba, L., Biswas, M., Kuppili, V., Godia, E. C., Suri, H. S., Edla, D. R., Omerzu, T., Laird, J. R., Khanna, N. N., Mavrogeni, S., et al. (2019). The present and future of deep learning in radiology. European journal of radiology, 114(5):14–24.
- Salim, M. H. M., Ali, N. M., and Noah, S. A. M. (2017). Mobile application on healthy diet for elderly based on persuasive design. International Journal on Advanced Science, Engineering and Information Technology, 7(1):222–227.
- Samsung (n.d.). Samsung health. Samsung. Retrieved 15 Aug 2021.
- Shah, T., Purohit, G., Nair, S. P., Patel, B., Rawal, Y., and Shah, R. (2014). Assessment of obesity, overweight and its association with the fast food consumption in medical students. Journal of clinical and diagnostic research: JCDR, 8(5):CC05.
- Singh, Y. and Chauhan, A. S. (2009). Neural networks in data mining. Journal of Theoretical & Applied Information Technology, 5(1):101–108.
- Solomatine, D., See, L. M., and Abrahart, R. (2009). Data-driven modelling: concepts, approaches and experiences. Springer.
- Spanakis, G., Weiss, G., Boh, B., Lemmens, L., and Roefs, A. (2017). Machine learning techniques in eating behavior e-coaching. Personal and Ubiquitous Computing, 21(4):645–659.
- Tao, W., Leu, M. C., and Yin, Z. (2020). Multi-modal recognition of worker activity for human-centered intelligent manufacturing. Engineering Applications of Artificial Intelligence, 95(C):103868.
- Tapper, K. (2021). Health Psychology and Behaviour Change: From Science to Practice. Macmillan Education UK, 1 edition.
- Tauber, S., Navarro, D. J., Perfors, A., and Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. Psychological review, 124(4):410.

- Teeriniemi, A.-M., Salonurmi, T., Jokelainen, T., Vähäniikkilä, H., Alahäivälä, T., Karppinen, P., Enwald, H., Huotari, M.-L., Laitinen, J., Oinas-Kukkonen, H., et al. (2018). A randomized clinical trial of the effectiveness of a web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. Journal of internal medicine, 284(5):534–545.
- Toscos, T., Faber, A., Connelly, K., and Upoma, A. M. (2008). Encouraging physical activity in teens can technology help reduce barriers to physical activity in adolescent girls? In 2008 Second International Conference on Pervasive Computing Technologies for Healthcare, pages 218–221. IEEE.
- van der Lubbe, L. M. and Klein, M. C. (2019). Designing a system with persuasive communication to improve diet compliance for elderly users. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare, pages 234–241.
- van Rossum, G. and Drake Jr, F. L. (1995). Python tutorial. Centrum voor Wiskunde en Informatica Amsterdam.
- Vasquez, D., Fraichard, T., Aycard, O., and Laugier, C. (2008). Intentional motion on-line learning and prediction. Machine Vision and Applications, 19(5):411–425.
- Verhoeven, A. A., Adriaanse, M. A., de Vet, E., Fennis, B. M., and de Ridder, D. T. (2015). It’s my party and i eat if i want to. reasons for unhealthy snacking. Appetite, 84(2):20–27.
- Vu, T. H., Dang, A., Dung, L., and Wang, J.-C. (2017). Self-gated recurrent neural networks for human activity recognition on wearable devices. In Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pages 179–185.
- Wan, X. (2019). Influence of feature scaling on convergence of gradient iterative algorithm. In Journal of physics: Conference series, volume 1213, page 032021. IOP Publishing.
- Wang, J., Ma, Y., Zhang, L., Gao, R. X., and Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. Journal of manufacturing systems, 48(C):144–156.
- Wang, Y., Li, J., Liu, Q., and Ren, Y. (2015). Prediction of purchase behaviors across heterogeneous social networks. The Journal of Supercomputing, 71(9):3320–3336.
- Wang, Y. and Reiterer, H. (2019). The point-of-choice prompt or the always-on progress bar? a pilot study of reminders for prolonged sedentary behavior change. In Extended

- Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pages 1–6.
- Weber, L. and Fore, M. E. (2007). Race, ethnicity, and health: An intersectional approach. In Handbooks of the sociology of racial and ethnic relations, pages 191–218. Springer.
- West, R. and Brown, J. (2013). Theory of Addiction. Addiction Press. Wiley.
- WHO, W. H. O. (2021). Noncommunicable diseases. <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases/>. Retrieved 29 Dec 2021.
- Williamson, D. F., Parker, R. A., and Kendrick, J. S. (1989). The box plot: a simple visual method to interpret data. Annals of internal medicine, 110(11):916–921.
- Yang, J., Xu, Y., and Chen, C. S. (1997). Human action learning via hidden markov model. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 27(1):34–44.
- Yang, W., Sun, S., Hao, Y., and Wang, S. (2022). A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. Energy, 238(3):121989.
- Yu, Y., Cao, J., and Zhu, J. (2019). An LSTM short-term solar irradiance forecasting under complicated weather conditions. IEEE Access, 7(1):145651–145666.
- Zhang, Y. and Pennacchiotti, M. (2013). Predicting purchase behaviors from social media. In Proceedings of the 22nd international conference on World Wide Web, pages 1521–1532.
- Zheng, H., Wang, H., and Black, N. (2008). Human activity detection in smart home environment with self-adaptive neural networks. In 2008 IEEE International Conference on Networking, Sensing and Control, pages 1505–1510. IEEE.

Appendix A

Ethics approval



Psychology Research Ethics Committee
School of Arts and Social Sciences
City University London
London EC1R 0JD

10 April 2018

Dear Emmanuel Pothos, Shaima Alghamdi
Tillman Weyde, Katy Tapper

Reference: PSYETH (S/L) 17/18 87

Project title: Tracking snacking

I am writing to confirm that the research proposal detailed above has been granted approval by the City University London Psychology Department Research Ethics Committee.

Period of approval

Approval is valid for a period of three years from the date of this letter. If data collection runs beyond this period you will need to apply for an extension using the Amendments Form.

Project amendments

You will also need to submit an Amendments Form if you want to make any of the following changes to your research:

- (a) Recruit a new category of participants
- (b) Change, or add to, the research method employed
- (c) Collect additional types of data
- (d) Change the researchers involved in the project

Adverse events

You will need to submit an Adverse Events Form, copied to the Secretary of the Senate Research Ethics Committee [redacted] in the event of any of the following:

- (a) Adverse events
- (b) Breaches of confidentiality
- (c) Safeguarding issues relating to children and vulnerable adults
- (d) Incidents that affect the personal safety of a participant or researcher

Issues (a) and (b) should be reported as soon as possible and no later than 5 days after the event. Issues (c) and (d) should be reported immediately. Where appropriate the researcher should also report adverse events to other relevant institutions such as the police or social services.

Should you have any further queries then please do not hesitate to get in touch.

Kind regards

[redacted]
Ethics committee Secretary
Email: [redacted]

[redacted]
Chair
Email: [redacted]

Fig. A.1

Appendix B

Participants comments about the study

- I've been having a lot of fruit recently, so haven't snacked as much as I normally would, particularly unhealthy snacks.
- I am surprised at how easy to use the app is.
- Recording each unhealthy snack made me more aware of what I was eating and I think reduced the number of unhealthy snacks I ended up eating, instead of just mindlessly consuming them.
- Hi I can confirm that I have been logging all my unhealthy snacks. There have indeed been some days when I haven't had any which will be the days that I have missed. I think the app made me more aware of the amount of unhealthy snacks I was having!
- I work four days on then four days off, and generally don't snack during my working days, we work 12 hour days with three breaks when I have regular food, so no time for snacks. Hope this makes sense.
- Yes I am still receiving the SMS reminders. I have been trying to cut down on unhealthy snacks recently.
- It all depends what shift I am working as to how many I have! Tuesday 18th would have been none, Friday 21st I had 2 and yesterday 22nd I had 1 but totally forgot to mark them down (the reminders were very helpful!)
- On days with no recordings I did not consume any unhealthy snacks. Thank you so much.
- Hi Shaima. I am trying to be good on my diet but on the days I have more unhealthy are the days I have struggled. Those where I haven't are those I have managed to stay on track. Hope this helps.
- Thank You for your msg :) yes I am enjoying taking part in your study and yes all my details/recordings are correct to the current moment!
- The days when they weren't recorded were because I was too busy with work and I ended up not eating unhealthy food during the day, so I thought I shouldn't be recording, but you

can see I'm recording all the snacks I make between healthy meals. This study is helping me to realize the amount of harmful food in my health that I am eating daily. Thank you.

- after payment of £16 I would be willing to continue for more research; although I'd have to ask you to resent the 'app' details for me to use my mobile phone as the app takes up far too much space on my iPad!
- Hi Shaima, Thank you for your message! I have been receiving the SMS reminders, and was really enjoying participating in the study. I have to be perfectly honest and tell you that my phone charger stopped working throughout the study, so I wasn't always able to record the snacks when I had them, but I was still recording them when I switched my phone on
- tracking my habits had made me more aware and mindful of what I am eating, although I think having a less stressful week has also made a difference.
- Was very different but interesting
- Food often makes me hungry but I enjoy fasting and when I have done it once in the week it is much easier on other days. I am taking medication at the moment which has to be with food and eating makes me much hungrier.
- A nice interesting and highly relevant study, thanks.
- It was very interesting.
- Super interesting. This particular subject is quite a big impact of my life.
- Good concept.
- The first part of the study has made me realise how much snacking I do!!
- Good study
- Interesting and easy. Thanks!
- This sounds like a fun and interesting study.
- great survey, feels really relevant to me and i am sure for everyone else and their lifestyles nowadays.
- Its made me think about what im eating
- Great study. Enjoyed taking part
- A very good study I like it I,Ãm on a keto diet and I have lost 20 kg of weight. I eat a lot of snacks which are keto friendly and does not put in any weight. Even if I do I get back on track next day.
- I'm managing to record snacks, I'm just trying to be a little bit healthier.

Appendix C

Snacks consumed in the Netherlands

Snack translated	Freq.
Cakes	265
Candeproof, yogurt	2
Candy	169
Candy bar	558
Candy bar, cookies	26
Candy bar, nuts	3
Candy bar, pastries, cookies	1
Candy bar, pastry	2
Candy hand, chips	1
Candy, candy hand	3
Candy, chips	1
Chips, cookies	6
Chips, nuts	7
Chips, pastries	1
Cookies	691
Cookies, nuts	10
French fries, candy, chips	3
French fries, salad	30
Fries	550
Fries, ice cream	1
Fruit	2270
Fruit Salad	73

Continued on next page

Snack translated	Freq.
Fruit, candy	3
Fruit, candy hand	11
Fruit, candy strap, yogurt	1
Fruit, chips	7
Fruit, Cookies	37
Fruit, cookies, nuts	1
Fruit, nuts	23
Fruit, pastries	3
Fruit, salad, chips	1
Fruit, salad, yogurt	7
Fruit, yogurt	199
Fruit, yogurt, nuts	29
Fruit, yogurt, pastries	1
Ice	126
Ice cream	3
Ice cream muffin	3
Ice cream yogurt	1
Ice, fruit	3
Muffin	342
Muffin, candy	1
Muffin, candy bar, cookies	1
Muffin, candy hand	3
Muffin, chips	1
Muffin, cookies	6
Muffin, Fruit	2
Muffin, Fruit, Yogurt, Nuts	1
Muffin, nuts	1
Muffin, Pastry	5
Nuts	184
Pastry, cookies	5
Potato chips	356
Potatoes, candy	1
Potatoes, candy strap, yogurt	1
Potatoes, fries	20

Continued on next page

Snack translated	Freq.
Potatoes, fries, salad, chips	1
Potatoes, ice cream	3
Potatoes, pastries	1
Potatoes, salad	37
Potatoes, salad, pastry	1
Potatoes, yogurt	1
Potatoes, yogurt, pastries	1
Salad	769
Salad, cookies	3
Salad, yogurt	13
Sweets, cookies	6
Sweets, yogurt	1
Yogurt, cookies	1
Yogurt, nuts	218
Yogurt, Pastry	2
Total	7120

Appendix D

UK Participants snacking trend throughout the study

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
1	4	1	1	0	u
2	4	2	2.7	1	i
3	4	2	1.9	0	u
4	4	2	1.5	-1	d
5	4	2	2.6	1	i
6	4	1	1.3	0	u
7	4	2	2.5	1	i
8	4	5	3.9	-1	d
9	4	1	1.7	1	i
10	4	1	1.3	0	u
11	4	1	1.8	1	i
12	4	3	1.6	-1	d
13	4	1	1.7	1	i
14	4	1	2.3	1	i
15	4	2	1.2	-1	d
16	4	1	1.7	1	i
17	4	1	2	1	i
18	4	3	2.5	-1	d
19	4	3	1.8	-1	d

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
20	4	3	1.3	-2	d
21	4	1	1.2	0	u
22	4	1	1	0	u
23	4	1	1.2	0	u
24	4	1	1.6	1	i
25	3	2	0	-2	d
26	4	1	1.5	1	i
27	4	1	1.2	0	u
28	4	3	3.7	1	i
29	4	1	1.3	0	u
30	4	5	2.7	-2	d
31	4	1	1	0	u
32	4	3	2	-1	d
33	4	2	1.7	0	u
34	4	3	2	-1	d
35	4	2	2.5	1	i
36	4	5	3.4	-2	d
37	4	7	2.2	-5	d
38	4	2	1.2	-1	d
39	4	2	2.7	1	i
40	4	1	1.2	0	u
41	4	3	1	-2	d
42	4	2	1.6	0	u
43	4	2	1.5	-1	d
44	4	2	1.3	-1	d
45	4	1	1.8	1	i
46	4	1	1.7	1	i
47	4	1	2.3	1	i
48	3	4	0	-4	d
49	4	2	2.6	1	i
50	4	1	1.3	0	u
51	4	2	1.3	-1	d
52	4	4	5.9	2	d
53	4	2	2.2	0	u

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
54	4	1	1.4	0	u
55	4	1	1.3	0	u
56	4	2	1	-1	d
57	4	1	1	0	u
58	4	5	1.7	-3	d
59	4	4	3.7	0	u
60	4	4	1.6	-2	d
61	4	2	1.5	-1	d
62	4	2	1	-1	d
63	3	2	0	-2	d
64	4	2	2.6	1	i
65	4	4	3.4	-1	d
66	4	2	2.4	0	u
67	4	2	2	0	u
68	4	1	1.8	1	i
69	4	2	3.3	1	i
70	4	2	3.3	1	i
71	4	2	1	-1	d
72	4	5	1.5	-4	d
73	4	2	9.5	8	i
74	4	3	2.8	0	u
75	4	1	1.7	1	i
76	4	2	1.5	-1	d
77	4	2	1.8	0	u
78	4	1	2	1	i
79	4	2	3.1	1	i
80	4	3	4.9	2	i
81	4	4	1.6	-2	d
82	4	7	3.8	-3	d
83	4	5	2.7	-2	d
84	4	3	3	0	u
85	4	2	2.1	0	u
86	4	1	3.3	2	i
87	4	5	1.3	-4	d

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
88	4	4	2	-2	d
89	4	2	2	0	u
90	3	1	0	-1	d
91	4	2	1.8	0	u
92	4	1	2.2	1	i
93	4	3	1.8	-1	d
94	4	2	1.5	-1	d
95	4	1	2.9	2	i
96	4	1	1.2	0	u
97	4	4	2.3	-2	d
98	4	3	3.8	1	i
99	4	4	3	-1	d
100	4	1	3.6	3	u
101	4	2	1	-1	d
102	4	2	1.8	0	u
103	4	4	2.8	-1	d
104	4	3	2.1	-1	d
105	4	2	4	2	i
106	4	1	1.6	1	d
107	4	3	3	0	u

Appendix E

Dutch participants snacking trend throughout the study

*participants with only 1 week of recording (28 participants) were excluded from Dutch progress tables. Healthy snacking was excluded too.

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
1	3	2	1	-1	d
3	2	2	2.5	1	i
4	4	1	1	0	u
6	4	1	1	0	u
8	2	1.4	1.5	0	u
10	4	1	2	1	i
11	2	3.5	2.5	-1	d
12	2	1.8	1.3	-1	d
13	2	2	1.8	0	u
14	4	1	1.5	1	i
19	3	2	1.5	-1	d
20	3	4	2.8	-1	d
21	3	2	1.8	0	u
23	2	2	3	1	i
24	2	1	1	0	u
26	4	5	1.3	-4	d
27	2	1	1	0	u

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
28	2	1	1.5	1	i
32	3	1	1.4	0	u
34	3	1	1	0	u
35	2	1.3	1	0	u
36	4	2	3.3	1	i
38	2	1	1	0	u
39	3	1	2	1	i
40	3	2	2.1	0	u
43	3	4.4	2.7	-2	d
45	4	4	1.3	-3	d
47	3	4	2	-2	d
50	4	1.3	1	0	u
52	2	1	1	0	u
53	2	1	1.3	0	u
57	4	1	1.2	0	u
60	2	1	1.5	1	i
61	3	1	1.3	0	u
65	2	1	1.5	1	i
71	4	2	1	-1	d
74	3	2	4.5	3	i
75	2	2	1	-1	d
78	4	1	1	0	u
83	4	1	2	1	i
86	2	1.5	1	-1	d
87	3	1	1	0	u
89	4	5.3	5	0	u
93	3	1.5	1.5	0	u
95	2	2	1	-1	d
97	2	1	1.3	0	u
100	2	2	1.3	-1	d
101	2	1	1	0	u
102	3	1	1	0	u
106	4	2	1	-1	i
107	2	2	2.5	1	u

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
108	2	4	2	-2	d
111	4	2	2	0	u
114	4	1.3	1	0	u
116	4	1	1.5	1	i
119	2	2	1	-1	d
121	4	4	2	-2	d
122	2	1	1	0	u
123	4	1.3	1.2	0	u
125	2	1	1	0	u
126	4	2	1	-1	d
127	4	1	1	0	u
129	2	3	4	1	i
130	2	1	1	0	u
133	2	1	1	0	u
134	4	1	1	0	u
136	3	1	1	0	u
137	2	2	1	-1	d
141	4	2.5	1	-2	d
143	2	1	1.8	1	i
149	2	1	1.3	0	u
150	3	1	1	0	u
152	4	6	4	-2	d
154	2	8	1	-7	d
155	2	2	1.3	-1	d
157	2	1	1	0	u
161	3	1	1	0	u
163	4	2	1	-1	d
164	3	1	1	0	u
166	3	1	1	0	u
167	4	1.1	5	4	i
168	3	1	1	0	u
169	3	1.2	1	0	u
170	4	2	1	-1	d
171	4	1	1	0	u

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
175	4	2	1.7	0	u
179	2	1	1.5	1	i
181	4	1	1	0	u
184	2	1	1	0	u
186	2	1	1.8	1	i
188	3	2	1	-1	d
202	3	1	1	0	u
204	3	3	1.6	-1	d
302	4	1	2.5	2	i
303	3	1	1	0	u
304	4	1	1	0	u
306	4	1	1	0	u
307	2	1.5	1	-1	d
308	3	3	2	-1	d
310	3	3	1	-2	d
312	2	1.7	1.7	0	u
314	3	1	1	0	u
399	3	2	1.3	-1	d
401	2	1	1	0	u
403	3	1.2	1	0	u
406	2	1	1.5	1	i
407	4	1	1	0	u
408	4	1	2	1	i
412	3	1.4	1	0	u
415	2	1	2	1	i
416	4	3	1.5	-2	d
420	4	1	2.3	1	i
501	4	1	1	0	u
503	3	1	1	0	u
504	2	1.5	1.6	0	u
506	3	1	1	0	u
507	4	1	2	1	i
508	4	3	1	-2	d
510	3	2	1.2	-1	d

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
511	2	2.3	1	-1	d
512	2	2	1.3	-1	d
515	2	2	1.2	-1	d
516	3	1	1.3	0	u
599	3	1	1.5	1	i
602	2	1	1	0	u
603	4	1	1	0	u
605	3	2	1.2	-1	d
606	3	1	1.4	0	u
607	3	1	1	0	u
608	3	2	1.5	-1	d
609	2	1	1.5	1	i
610	3	1	2	1	i
699	4	2	1	-1	d
701	4	2	1	-1	d
702	2	1	1	0	u
704	3	1	2.3	1	i
707	4	2	1	-1	d
709	2	2	1	-1	d
710	3	2	1.8	0	u
712	3	1	1.6	1	i
714	3	1	1	0	u
716	4	2	1	-1	d
798	3	3	1.9	-1	d
801	2	1.3	1.5	0	u
802	2	1.8	1.6	0	u
803	4	1	1	0	u
804	4	1	1.7	1	i
805	4	1	1	0	u
806	2	1.7	2	0	u
807	3	1	1	0	u
808	2	1.8	1	-1	d
809	2	1.9	1.7	0	u
810	3	3	2.2	-1	d

Continued on next page

ID	No of weeks	1st week snacks	Last week Snacks	Differences	Snacking Progress
811	3	3	2.2	-1	d
812	3	1.5	2	1	i
814	2	1.5	1.2	0	u
815	4	1	1	0	u
818	2	1.5	2	1	i
819	3	1	1.2	0	u
820	3	1	1	0	u
822	3	1	1	0	u
825	2	1.5	1	-1	d
827	3	3.1	2	-1	d
829	3	1.2	1.7	1	i
831	4	2	1	-1	d
834	2	1.3	1	0	u
898	3	1	1.3	0	u
899	3	1	1	0	u