Master’s Thesis

Swiss SaltTracker: Predicting Dietary Salt Intake from Retail Loyalty Card Logs

Master’s Thesis in Business Informatics

Autor: Sven Brunner

Gränichen, Aargau, Switzerland

Matriculation Number: 10-933-703

Supervisor: Klaus Fuchs (Auto-ID Labs ETH / HSG)

Supervising Professor: Lorenz Hilty

Informatics and Sustainability Research Group

Department of Informatics

University of Zurich

Date: 31.10.2016

ABSTRACT

Excessive salt intake is a global health concern, since 2.5 million deaths per year are preventable if salt consumption was reduced to healthy levels. In Switzerland, salt intake of 94% of men and 77% of women ranges above the WHO recommendation of 5 g salt per day. To identify people with an increased salt intake, cumbersome 24-hours-urine-samples or paper-based checklists are used today. This thesis provides the technological and statistical foundations to automatically determine a person's salt intake level from his loyalty card logs. To reach this goal, a system of several software services has been developed during the thesis to obtain the loyalty card logs and measure the daily salt intake of a person by using a food-record-checklist which has been validated and calibrated with 24-hours-urine-samples. The paper-based checklist has been translated into a mobile application running on iOS and Android. Because of the small number of participating users in the thesis study, a correlation between the self-reported salt intake in the mobile app and the loyalty card logs could not be validated yet. The application of classification algorithms on the created models and the collected data however yielded very promising results. Given a larger sample size and the technical framework developed during the thesis, the correlation between loyalty card logs and self-reported dietary salt intake is expected to be proven in the near future.

ZUSAMMENFASSUNG

Exzessiver Salzkonsum ist ein globales Gesundheitsproblem, da 2.5 Millionen Tode pro Jahr verhindert werden könnten, würde der Salzkonsum auf ein gesundes Niveau gesenkt werden. Die tägliche Salzaufnahme von 94 % der Männer und 77% der Frauen in der Schweiz liegt über der WHO-Empfehlung von 5 g pro Tag. Um Personen mit erhöhtem Salzkonsum zu identifizieren, werden heute umständliche 24-Stunden Urinproben oder Checklisten auf Papier benutzt. Diese Arbeit hat zum Ziel, die technischen und statistischen Grundlagen zur automatisierten Erkennung des Salzkonsumniveaus einer Person aus deren Kundenkartendaten bereitzustellen. Mehrere Softwaresysteme wurden entwickelt, um die Kundenkartendaten zu erhalten und die Salzaufnahme einer Person unter Zuhilfenahme einer durch 24-Stunden Urinproben validierten und kalibrieten Checkliste zum Lebensmittelverzehr zu messen. Die Checkliste wurde dazu in eine mobile Applikation für Android und iOS überführt. Aufgrund der kleinen Teilnehmeranzahl in der durchgeführten Studie konnte eine Korrelation zwischen des mittels App gemessenen Salzkonsumniveaus und den Kundenkartendaten noch nicht mit Sicherheit festgestellt werden. Die Anwendung von Klassifikationsalgorithmen auf die entwickelten Modelle und die gesammelten Daten lieferte jedoch vielversprechende Resultate. Mit einer grösseren Teilnehmerzahl und dem technischen Gerüst, welches während dieser Arbeit entwickelt wurde, ist zu erwarten, dass eine Korrelation zwischen Kundenkartendaten und Salzkonsumniveau, gemessen durch Einträge in die App, in naher Zukunft festgestellt werden kann.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Application Programming Interface
DRF	Django Rest Framework
FOPH	Federal Office of Public Health
FRCL	Food-Record-Checklist
FSVO	Federal Food Safety and Veterinary Office
ORM	Object-relational mapping
PIID	Product-Item-Ingredient-Database
SDK	Software Development Kit
SGE	Schweizerische Gesellschaft für Ernährung

CHAPTER 1. INTRODUCTION

1.1 Motivation

Cardiovascular diseases, elevated blood pressure and other diet-related diseases are nowadays one of the most frequent causes of mortality worldwide. Diet-related diseases are responsible for more deaths than all non-diet related causes together. (Aburto et al. 2013; Drewnowski and Popkin 1997; Forouzanfar et al. 2015; Popkin 2002). Current estimates suggest that diet-related diseases cause more than 27.8 million deaths per year (Forouzanfar et al. 2015).

A major factor causing and favoring cardiovascular diseases is a person's diet. More concretely, the salt in foods or more precisely the sodium component of salt can be identified as the main driver. (Bachmann et al. 2004; Franco and Oparil 2006; Weinberger 2002). Consequently, reducing salt and subsequently also sodium can significantly decrease blood pressure and derogate the risk of cardiovascular disease (Bachmann et al. 2004; Murielle Bochud et al. 2012; Srinath Reddy and Katan 2004; WHO 2003; Yang et al. 2011). 2.5 million lives could be saved each year if the average salt intake would be decreased below the level of 5 g salt per day.[1] Therefore the reduction of salt and sodium must be a key part of any strategy planning to reduce cardiovascular diseases and diet-related mortality (Aburto et al. 2013; He, Li, and Macgregor 2013; O'Flaherty et al. 2016; Strazzullo et al. 2009).

In 2008 the Federal Food Safety and Veterinary Office (FSVO) and the Federal Office of Public Health (FOPH), both institutions of the Swiss Confederation, created the "Swiss Salt Strategy". The strategy was developed as a consequence of the Swiss Nutrition Policy, which is a "National Program on Nutrition and Physical

[1] http://www.who.int/mediacentre/factsheets/fs393/en/

Activity"[2]. Being aware of the harmful impact too high of a salt intake can have on a person's health, the goal of the "Swiss Salt Strategy" is to reduce the salt intake of the entire population of Switzerland. The strategy, whose implementation took place between 2013 and 2016, is structured into a short-term and a long-term goal. By the end of 2016 an average daily salt intake below 8 grams per person is aspired. The long-term goal is to reach the World Health Organization's (WHO) recommendation of an average of less than 5 grams per day (WHO 2014). The strategy consists of several measures on either the consumer side, for example increasing the personal awareness by campaigns or further data collection and analysis, or the producer or industry side by reducing salt in the manufactured foods.

Approaching the food producers is especially important, as it was shown that in Switzerland 75-80 percent of the salt intake has its source in the consumption of processed food. The addition of salt into meals at home or during cooking only plays a subordinate role (Chappuis et al. 2011b; He and MacGregor 2010). An analysis of the processed foods in Switzerland and their nutrition ingredients revealed which product categories are responsible for the salt intake of the population. The top-five categories were found to be bread (17%), cheese (11%), meat products (8%), soups (7%), and ready meals (5%) (Beer-Borst et al.; Chappuis et al. 2011b; Constanza et al. 2004). It is therefore important to work together with the food industry to lower the salt amount added during the production phase of a food item in order to achieve a nationwide effect on decreasing the population's salt intake (Cobiac, Vos, and Veerman 2010; Zülli, Allemann, and Mitarbeiterin 2011). Equally important is the change of the consumer's eating habits, where only strategies incorporating several measures and actions on both the consumer and producer side can lead to positive long-term effects (Chappuis et al. 2011a).

[2] http://www.bag.admin.ch/themen/ernaehrung_bewegung/index.html?lang=en

Beer-Borst et al. (2009) developed and calibrated a food questionnaire based on data of 24-hour urine collection. The data was gathered between 1993 and 2009 from a sample population in Geneva. The records included data from approximately 13'000 people. The results showed an average salt intake of 10.6 g for men and 8.1 g for women. These values exceed both the recommendation of the WHO of 5 g and the goal of the Swiss Federation of 8 g. The data also revealed that 94% of the male and 77% of the female population have a higher daily salt intake than 5 g (Beer-Borst et al. 2009). This evaluation of the salt intake of the population in Switzerland clearly shows the necessity for intervention to lower the average salt intake, considering the negative impact on health.

Some interventions on the production side have already been conducted successfully, however as already mentioned it is not sufficient to tackle food industry alone (Zülli et al. 2011). The foundation of every intervention on the level of individuals must be an assessment of the personal salt intake. Without this knowledge, any educational consultancy by physicians or nutritionists is not expedient. Traditionally, such an assessment includes the analysis of several 24-hour-urine-samples (Beer-Borst et al. 2004; Brown et al. 2013). These assessments capture an individual's salt intake very accurately but are rather invasive for the person itself and, due to the fact that laboratory equipment is needed for the analysis, also very costly. The resource-insensitivity only allows to analyze a small size of the entire population and only for a very brief timespan. The recording of a person's salt intake over the course of several weeks or months, incorporating even different seasons of the year, is either not manageable or simply too expensive. (Beer-Borst et al. 2009)

As a possible solution to the issues described, this thesis aims to automatically link a person's retail loyalty card logs to his or her (in the following it is referred to the male form only) dietary salt intake. Using Product-Item-Ingredient-Databases (PIID) it is possible to determine the sodium or salt values for food

items appearing on a person's loyalty card log. Governmental regulations, as for example the EU 1169[3], obligate manufactures in the food sector to make ingredient data publicly and electronically available. The resulting PIIDs include trusted data as the data is directly inserted by the respective manufacturer.
Using this data sources, the salt intake of an individual can be measured and monitored efficiently regarding costs and user involvement. By mapping the products from the PIIDs to the products in a consumer's retail log, it could be possible to automatically determine the amount of salt of a purchased product or an entire shopping basket.

The validity of the link between a person's salt intake and his retail log has yet to be verified. The goal of this thesis is to research whether such a correlation between a person's true salt intake and his retail log can be found by developing a mobile application which allows the user to report his salt intake and provide access to his retail log.

1.2 Terminology

Throughout this thesis, several terms will be used referring to a person. When writing about salt intake and salt intake measurement in general the term "person" is used. In the context of loyalty card and retailers the term "customer" is sometimes used. Taking the technical perspective, a person is referred to as "user". All terms are used interchangeably.

[3] http://eur-lex.europa.eu/legal-content/de/ALL/?uri=CELEX%3A32011R1169

CHAPTER 2. RELATED WORK

2.1 Salt Intake Monitoring

In order to create new, non-invasive methods towards dietary salt assessment, a paper-based Food-Record-Checklist (FRCL) (Beer-Borst et al. 2015) (Figure 1) has been suggested aiming to substitute the cumbersome 24-hour-urine-sampling process through a validated paper-based checklist. The checklist is divided into several food item categories, grouping different food items. It is important to note that these categories are different from the food item categories defined by the "Schweizerische Gesellschaft für Ernährung"[4] (SGE), which are also used later on in this thesis. The SGE is a national association concerned with promoting healthy nutrition and nutrition information to the Swiss population and cooperates with FSVO and FOPH. When the term "category" is used throughout the thesis, it is always clarified if the categories from the FRCL or the SGE are meant. The FRCL is a checklist and not a food diary, it therefore only contains categories and products that have been identified as relevant for a person's salt intake.
The structure of the paper based FRCL version is illustrated in Figure 1.

Coffee and tea									
	Reference portion	**Meal**	**Number portions**					**Addition of**	
			1	**2**	**3**	**4**	**5+**	**Cream**	**Milk**
Coffee, Espresso, Tea	1 cup (ca. 200 ml)	Breakfast						○	○
		AM snack						○	○
		Lunch						○	○
		PM snack						○	○
		Dinner						○	○
		Late hour						○	○
Coffee based beverages, warm or cold e.g. Cappuccino, Cafe latte, Latte macchiato	1 cup (ca. 200 ml)	Breakfast							
		AM snack							
		Lunch							
		PM snack							
		Dinner							
		Late hour							

Figure 1 Sample section of the paper-based FRCL

[4] http://www.sge-ssn.ch/

Shown is the category “coffee and tea” with its food items. A food item always has a reference portion. The number of portions consumed for each meal of the day can be selected. Some food items, in the sample category the item “Coffee, Espresso, Tea” have additional information attached. In this example, whether the user also added cream or milk to the beverage. For each food item and its attachments (e.g. cream or milk) a reference portion and a reference sodium value has been assigned. To calculate the salt value in grams (as used in the PIID utilized in this thesis) these values have to be converted to sodium in gram per 100 grams and multiplied with 2.548, as this is the multiplier for converting sodium to salt (Beer 2009).

Still, the paper-based FRCL approach requires a person to answer a 12-page long document for each time his salt intake level should be assessed, preventing a continuous monitoring of a large-scale, long-term behavioral change of large parts of the population. In food diaries, either in paper-based or digital form, a person needs to log every meal consumed on each day, whereas in the checklist only the specified food items are relevant and must be reported. For both the checklist and the diary, a supervision and external analysis is still necessary. mHealth applications, short for Mobile Health, belong to the digital part of the food diaries. Mobile applications for smartphones and tables allow their users to log the food items directly into the application (e.g. MyFitnessPal 2016; LifeSum 2016; Lose-it! 2016) and often provide feedback and rudimentary analysis, partially replacing a supervising entity. Although it could be shown that the self-reported diet correlates with the actual dietary intake (Ipjian et al. 2016) and can have a positive effect regarding a person’s dietary behavior (Tate, Wing, and Winett 2001; Turner-McGrievy and Tate 2011) there are also severe limitations. The people already interested in health-related topics and maintaining a rather healthy lifestyle and diet are also the ones who actually use such applications in the long term. Studies also have shown that the user-interaction is very limited and the applications are only used for a very short period of time (Block et al.

2015; Casperson et al. 2015; Hebden et al. 2012; Thomas, Leahey, and Wing 2015; Wharton et al. 2014). Especially the high risk population groups for obesity and unhealthy diet tend to show little interest in mHealth applications (Williamson et al. 2006). For many people the manual self-reporting process quickly leads to a lack of motivation and the reporting is discontinued, resulting in low quality reporting results (Williamson et al. 2006). To stimulate and enhance the usage of digital application the entry-barriers must be lowered and approaches must be found to keep up the motivation. Table 1 shows the currently available methods to monitor a person's salt intake.

	Urine-Sampling	Paper-based FRCL	mHealth Diary
Pro	+ Highest accuracy + Objective measurement	+ Validated + Lower Cost	+ No supervision through physician
Contra	- Cumbersome - High costs - One-time - Laboratory Setup - Requires supervision	- Very long paper survey, - Self-reported - One-time - Requires supervision	- Cumbersome manual logging - Very few clinical studies - Underreporting - Only used by highly interested/involved users

Table 1 Overview of contemporary Salt Intake Monitoring Methods

To sum up, it can be said that the state of the art methods to monitor dietary intake in general and salt intake in particular all have significant shortcomings, compromising either on accuracy or cost.

2.2 Predicting Dietary Intake from Loyalty Card Logs

Possibly overcoming the drawbacks of contemporary monitoring approaches, a possible solution to tackle the low retention rate of digital applications where a user can report his dietary intake is the already mentioned automatic linking of the loyalty card logs with the dietary intake: The potential of salt intake monitoring using loyalty card logs has already been shown in various studies (Brinkerhoff et al. 2011; Chidambaram et al. 2013; Coll 2013; Eyles, Rodgers, and Ni Mhurchu 2010; Närhinen, Nissinen, and Puska 1998; Stumbo et al. 2010; Winett et al. 1991) However, a validation of the reliability of using loyalty card log data to predict and approximate a person's true salt intake has not yet been conducted (Illner et al. 2012; Tin, Mhurchu, and Bullen 2007).

The market structure in the sector of food retails in Switzerland seems well-suited to test whether an approximation of salt intake is possible using loyalty card logs. The two largest retailers combine a market share of 80% and both offer a loyalty program which is widely used and accepted within the population, namely Cumulus by Migros (Figure 2) and Supercard by Coop (Accarda 2005; Handelszeitung.ch 2004). A crucial factor is not only the existence of loyalty programs but also the public availability of the items purchased in the context of the loyalty program in the sense that a customer is allowed to review his purchases. Fortunately, this is also the case in Switzerland. Combined with the availability of growing trusted Product-Ingredient-Item Databases (PIID) thanks to regulations like the EU 1169, the situation is promising.

Figure 2 "Cumulus" Loyalty Card
(Source: https://www.migros.ch/de/cumulus.html)

The European Commission has recently agreed on a pan-European data protection regulation (General Data Protection Regulation, GDPR) (European Commission 2016a) which grants European consumers the right for data portability: “A person shall be able to transfer their personal data from one electronic processing system to and into another, without being prevented from doing so by the data controller. In addition, the data must be provided by the controller in a structured and commonly used electronic format.” (European Commission 2016b). EU member states have to transpose the GDRP into national law by May 6th, 2018, which will give consumers more rights to request and process stored personal data from digital services (e.g. loyalty cards, etc.) through digital health applications. Furthermore, the European Commission is actively working on standardizing digital receipts (also e-Invoicing) (European Commission 2014), including recommending interoperability of digital receipt formats (European Commission 2016c), allowing users of digital payment means, e.g. Apple Pay[5], to request and access their digital receipts through further applications, e.g. mHealth applications. These developments, i.e. data portability granted through the GDPR, as well as planned standards on digital receipts and the EU regulation on online food declaration will support the development of mHealth systems.

Already existing research suggests that a single shopping basket of food items is not sufficient to draw any conclusion regarding salt or general nutrient intake for a person or a household (Stumbo et al. 2010). However, data collected over the course of 3-4 months could contain the necessary information (Stumbo et al. 2010). In case of the loyalty card log of Migros, the purchases of the last two years can be obtained.

The mobile application developed in the context of this thesis called “Swiss Salt-Tracker” is aimed to provide a starting point of the automatic linking of loyalty

[5] http://www.apple.com/chde/apple-pay/

card logs with salt intake reported by the user and laying the foundation for further interventions using digital media and applications for reducing salt intake.

2.3 Consumption Classification and Interventions

As already described, various studies (Forouzanfar et al. 2015; WHO 2014) have shown that an increased salt intake correlates with and increases risks of cardiovascular diseases and nephropathy. Therefore, salt intake monitoring practices classify consumption behavior into different risk segments (FSVO 2013, 2016). In this thesis, the salt intake risk classification is defined as depicted in Table 2.

	Healthy	Increased	Excessive
Salt Intake Level	- <= 5g (WHO) - <= 8g (Swiss Federation)	- 8g < x < 12g	- >= 12g
Risk	- Patient´s salt intake considered healthy	- Increased risk	- Strongly increased risk (He and MacGregor 2003)
Health Interventions (Based on (Kumanyika 1991)	- Mass media information or none	- Salt intake reduction - Awareness improvement - Coaching	- Salt intake reduction - Awareness improvement - Coaching

Table 2 Salt Intake Level Classification

People with a salt intake of 8 g and less already comply with the goal of the federal salt strategy and no intervention is needed (at least until the goal of a 5 g salt intake per day is approached more concretely in Switzerland). Between 8 g and

12 g the salt intake is already increased significantly and interventions are necessary (He and MacGregor 2003). Everything above 12 g can be considered excessive (He and MacGregor 2003) and intervention and coaching is strongly advised. The defined salt intake risk levels are mostly based on the goal of the Swiss Salt Strategy, e.g. which distinguished between healthy (8 g and below) and unhealthy (above 8 g) salt intake. The further separation between "Increased" and "Excessive" is not completely arbitrary (He and MacGregor 2003) but also different boundaries and categories would be possible. A more granular categorization however is helpful to provide more individual interventions in the future which are better suited for the respective salt intake risk level.

2.4 Research Gap

The goal of this thesis is to close the research gap which exists between salt or nutrition intake monitoring on the one hand and mHealth applications and loyalty card usage on the other hand (Figure 3). Therefore, using several software systems and the mobile application "Swiss SaltTracker", the connection between an individual's daily salt intake and the loyalty card log of the respective person is established and it is tested if a correlation can be found and if so, its accuracy is assessed. For the measurement of the salt intake the Food-Record-Checklist (FRCL) (Beer-Borst et al. 2015) is used and translated from its original paper-based version to a digital form used in the mobile application. The loyalty card logs will be taken exclusively from the Migros (in the following also referred to as "the retailer") loyalty program "Cumulus" as only these data can be retrieved digitally.

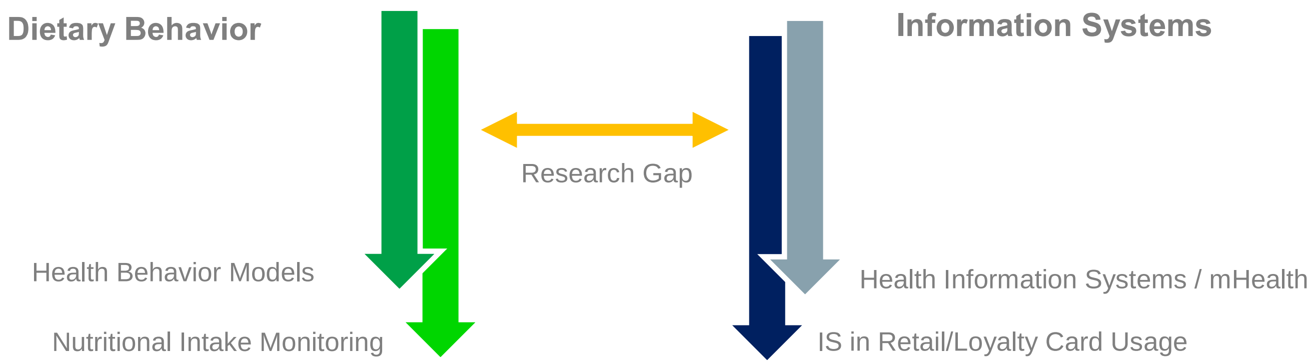


Figure 3 Research Gap

This thesis will not go further into field of interventions nor is the mobile application "Swiss SaltTracker" intended to measure behavioral changes of the user. This thesis however can be a starting point for future research and applications concerned with interventions supported by social-normative theory. It will be important to optimize when the intervention is presented to the user and how its effectiveness is measured.

CHAPTER 3. EXPERIMENT SETUP

3.1 General Setup

The central part of the experiment is the mobile application, which includes the digital version of the Food-Record-Checklist (FRCL). Users are asked to report their diet for four days (not necessarily consecutive) using the FRCL in the mobile app. Implementation details are covered in section 4.4. Further, users can add the credentials of their loyalty card inside the mobile application. The credentials are needed to obtain a user's loyalty card log. It has been shown that undesired or unhealthy eating habits are also evident in a person's purchase choices (Rankin et al. 1998), and subsequently in the respective loyalty card logs as well. As a person's salt intake is not only determined by the products bought, it is necessary to incorporate other factors as well. Examples are the household size or the frequency of eating out-of-home.

In order to link the dietary salt intake and the loyalty card log several linear regression and classification models will be built and tested which are further explained in section 3.4.

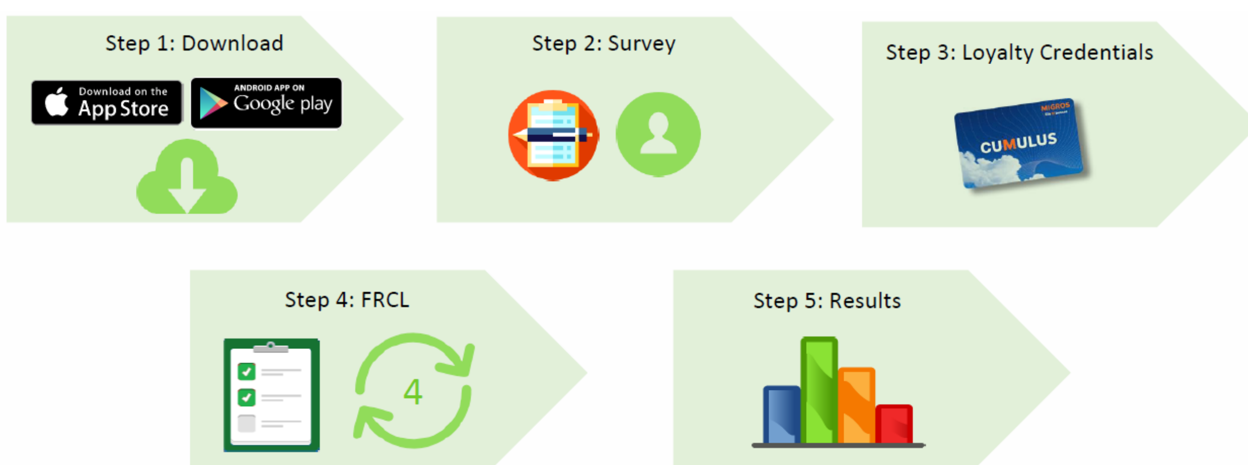


Figure 4 User Journey "Swiss SaltTracker" Study

Figure 4 shows the user journey, e.g. which steps a user passes through during the experiment. In a first step the user needs to download the mobile application from the app store on Android or iOS. Before using the application, users are asked to complete a short survey providing information to their shopping and eating behavior later used as control factors (Figure 4, Step 2) and their loyalty card log credentials (Figure 4, Step 3). Next, users report their nutrition with the help of the digital FRCL over the course of four days (Figure 4, Step 4). In order to motivate users to participate in the study, they receive an autonomously generated dietary assessment upon successful completion of Step 2, 3 and 4 (Figure 4, Step 5).

After completing the study, both the salt intake measured by the FRCL and the loyalty card log of a user is available for further processing. Figure 5 depicts the entire experiment design including the factors which cannot be captured by the research model. Grocery purchases at other retailers than the one the loyalty card log origins from as well as the food consumed in restaurants or canteens is not covered in the loyalty card logs but is reported in the FRCL. The node “Swiss SaltTracker Model” in Figure 5 denotes the regression and classification models developed for linking the salt intake measured by the FRCL and the salt intake calculated from the loyalty card logs.

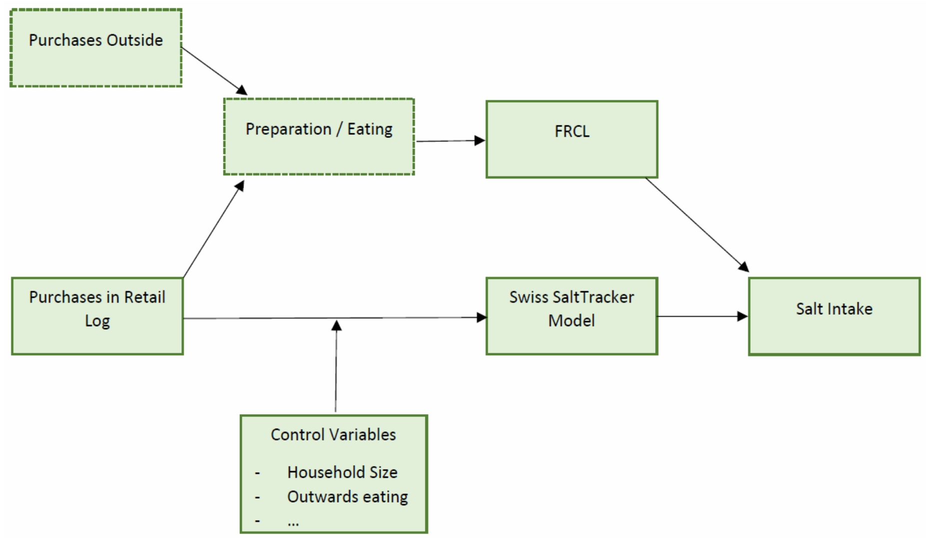


Figure 5 Experiment Design "Swiss SaltTracker"

3.2 Data Sources

To be able to link the loyalty card data of a user to his actual salt intake several data sources need to be integrated and considered. This section presents all relevant data sources in more detail.

3.2.1 Food-Record-Checklist

As mentioned, the FRCL is an integral part of this thesis and also one of the major data sources. In the mobile application to be developed in this thesis, all information of the FRCL concerning categories and food items must also be present in order to be able to compare the salt intake measurements. The selection of how many portions have been consumed however is differently implemented in the mobile application and the user can freely enter the number of portions (allowing to precisely measure the salt intake beyond five portions).

3.2.2 User Survey

It is important to know how representative the loyalty card logs really are for the specific user; therefore, a user survey is conducted during the registration process of the mobile application (Figure 6).

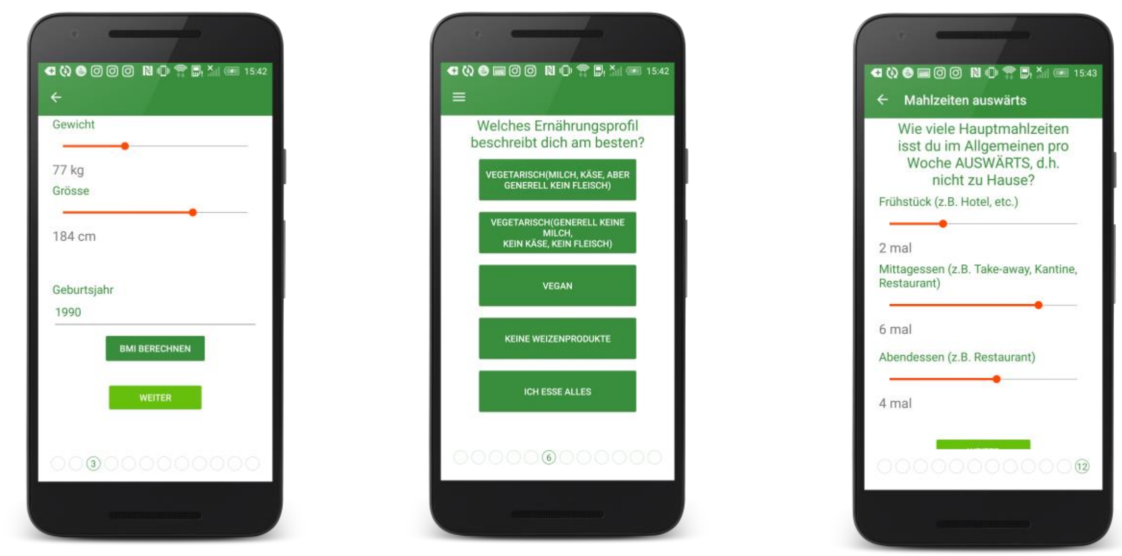


Figure 6 Sample Screenshots of the Registration Process

Besides the specific questions to shopping and eating behavior the user is also asked to provide general personal data such as gender, age, height and weight as this information can also help to construct the models. The most important information collected during the registration process are the credentials for the loyalty card, the username and password of the corresponding user account created in the system of the retailer. Without this credentials it is not possible to retrieve the loyalty card logs for this person. As this step is crucial and the credentials will possibly have to be looked up first, this step can be skipped and the credentials can be entered at any time later in the mobile application.

As a result of the user survey, five subsamples have been selected to serve as input for the regression and classification models. Besides the sample with all users, four additional samples are formed, each having specific properties. Table 3 gives an overview of the user samples and their respective properties.

	All Users	Small Households	Migros Shoppers	Little Outdoor Eating	Ideal Candidates
Household Size	Any	<= 2 people	Any	Any	<= 2 people
Shopping Loyalty	Any	Any	Grocery Shopping >= 60% Migros	Any	Grocery Shopping >= 60% Migros
Cumulus Usage	Any	Any	Use loyalty card at least “often”	Any	Use loyalty card at least “often”
Outdoor Eating Habits	Any	Any	Any	< 10 times per week	< 10 times per week

Table 3 User Samples

The subsamples are all expected to have a better correlation than the complete sample, especially the “ideal candidates” who live in small households (purchased food is likely to be consumed by the buyers), are loyal shoppers, often use their loyalty card and often eat at home (where they presumably prepare and consume the purchased food).

3.2.3 Product Item Ingredient Database

In the middle of 2016 a few databases specifically designed to store nutrition data existed in Switzerland. The main criterion for the Product-Item-Ingredient-Database (PIID) to fulfill, was a high coverage of products manufactured by the retailer of the loyalty card used in this thesis. As Migros is an official partner of the GS1 Trustbox[6], it was the obvious choice to serve as the main source for nutrition data. The Trustbox is also the only nutrition database in Switzerland where the data is directly inserted by the producers of the products, e.g. by Migros itself. The database is operated by Global Standard One (GS1) Switzerland[7], organized as a non-profit association and member of the international GS1 organization, which is, among other things, responsible for the issuing of barcodes and keeper of other international standards in the field of nutrition and supply-chain management.

By the time of writing, the chosen PIID included roughly 3500 products. Despite the data is being automatically added by the producers, the data quality is still not perfect. The nutrition information provided sometimes refer to the serving size and sometimes to 100 g or ml. Further, as the chosen PIID stores all nutrition information as strings, numeric values are not always represented as such or in heterogeneous formats. The most frequent errors detected in connection to numeric values were inconsistent use of "." and "," as floating point delimiter and additional text like range information (e.g. "up to 5") or unit specifications (0.3 g).

The technical integration of the PIID as data source into the application context is further outlined in section 4.3.3.

3.2.4 Loyalty Card Logs

All loyalty card log data utilized in this thesis are loyalty card logs of the "Cumulus" loyalty program. The available information in the loyalty card logs are the name of the product as it is printed on the receipts, the quantity, the price (total and per product) and (although not used in this thesis) the location and time of the purchase. An excerpt of a single loyalty card log record is shown in Figure 7. The

[6] http://www.trustbox.swiss/
[7] https://www.gs1.ch/de/home

process of gathering the retail logs is described in the technical part of this thesis in section 4.3.8.

Menge	Rabatt		
1.000	0.00	NOIR SPECIAL 72% 100G	2.10
2.000	0.60	M-CLAS PENNE 500G	3.00
1.000	0.85	VERMICELLES 2/160G	3.35
2.000	0.00	AB ENDIVIENSALAT 220G	4.80
1.000	0.00	AB PETERLI GESCH. 30G	2.60
1.000	0.00	TOMATENPUREE 440G DS	1.50
1.000	0.00	HEI M-DRINK HOCHPAST 1L	1.70
2.000	0.00	VALFLORA RAHM UHT 1/2L	6.30
2.000	0.00	ENERGYDRINK ERDBEER 330ML	2.60
1.000	0.00	ENERGYDRINK CHOCOLAT 330	1.30
1.000	0.00	GRANA PADANO GERIEB. 250G	4.35
0.797	0.00	APF.GALA I FT H.	2.95
1.300	0.00	CLEMENTINEN LOSE	7.15
1.000	0.00	TR.WEISS KERNLOS AB 500G	2.60
1.000	0.00	CRANBERRY BT S.	5.90
1.000	0.00	Himbeeren ab 250g	5.60
1.000	0.00	KASSENTRAGTASCHE NATIONAL	0.30
Total			**58.10**

Figure 7 Sample Loyalty Card Log Record (partial)

3.3 User acquisition

Before starting the study, the mobile application “Swiss SaltTracker” was made available to the public via iOS and Android (the entry is display in Figure 8) app stores. During the thesis experiment, no incentive was given to the users for their participation, nor was the application advertised in the public.

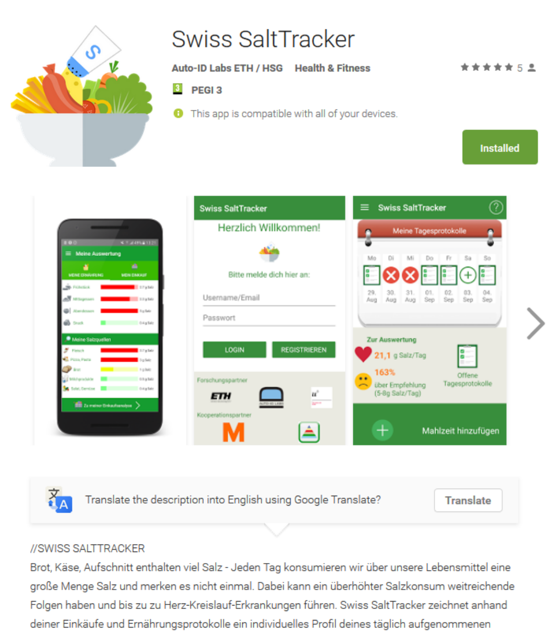


Figure 8 Google Play Store Entry

In the future it is planned to advertise the application on Facebook, which has proved to be a reasonable approach for mobile application user recruitment (Fuchs et al. 2016). Further, the application will be recommended by the "Schweizerische Gesellschaft für Ernährung" (SGE).

3.4 Research Questions

The central research question is whether a correlation between the salt intake reported via the FRCL and the loyalty card logs can be found. As a further research focus the necessary prerequisites that a user must fulfill are examined in order to be able to draw a conclusion of a user's salt intake given the respective user´s loyalty card log. It is possible that certain criteria, for example a high share of grocery purchases within the loyalty program or a small household size, must be met in order to successfully correlate the loyalty card logs with the salt intake in the first place. Before further explaining the linear regression and classification models, a very brief description of the two statistical methods is given, covering the basic concepts.

3.4.1 Linear Regression

Linear regression can be used to analyze the correlation between a numeric (in statistics referred to as "scalar") variable, called the dependent variable (denoted with *y*), and several independent variables (δ_i). In statistics, a variable is a property of an object, the object being the entity which is analyzed, for example a person or a country. In this thesis an object is always a user of the mobile app and the variables are certain properties of a user. (Moore et al. 2011)
Linear regression makes use of a linear model which predicts the dependent variable using the independent variables. For each independent variable, a parameter b_i (also called regression coefficients) must be found which describes the relationship between the independent and dependent variable. Additionally, a parameter b_0 is needed. It can be interpreted as the value which the dependent variable is expected to have when all independent variables are zero. A linear model has therefore the following form:

$$\boldsymbol{yi = b_0 + b_1 \times \delta_1 + b_2 \times \delta_2 + \cdots + e}$$

The regression models further explained in the following section are all linear models. The dependent variable is always the salt intake measured by the user's entries to the FRCL in the mobile application. The independent variables are values calculated from the loyalty card log. A positive value for the parameter b_i means that the correlation between the dependent and independent variable is also positive, e.g. when the respective independent variable increases, the dependent variable also increases. For example, a parameter b_i with a value of 0.5 indicates that if the corresponding independent variable increases by one unit the dependent variable would increase by 0.5. The parameter $\boldsymbol{e}$ stands for the error between the observed and predicted response. (Moore et al. 2011)

3.4.2 Classification

In contrast to the linear regression where a numeric value is predicted, classification is used to assign an object to a predefined class (Rokach and Maimon 2005). Classification needs to be distinguished from cluster analysis, which aims to find previously unknown classes or groups from a set of objects (Pretzer 2003). In a first step the classes must be defined. They do not necessarily need to have an intrinsic ordering (e.g. low, medium and high) and classes like "male" and "female" are also allowed. For the concrete calculation of the classification, several algorithms exist. For this thesis, the approach of decision tree learning is used. Decision trees, or when used in the context of classification also called classification trees, are given a set of classes, also called the target or dependent variable, and one or more input variables. The most usual calculation strategy is called top-down induction of decision trees and is performed step-wise. At each step the sample is split into several subsets. The splitting is based on the value of one of the variables of the objects in the sample. The variable with the highest information gain is selected, the split criterion used in this thesis is therefore the "Gain Ratio" (Tan, Steinbach, and Kumar 2006). The process is repeated recursively on all resulting subsets until all objects of a subset have the same class, e.g. the same value of the independent variable. As at each step only one variable is used, not all independent variables are necessarily included in the calculation. (Rokach and Maimon 2005)

3.4.3 Research Question 1: Salt Intake Classification Model

As described in section 2.3, it is important for practitioners in the public health sectors to know if a consumer has a healthy, increased or excessive dietary salt intake. Therefore, three categories, based on the categorization in section 2.3, have been defined which depend on a user's self-reported salt intake through the digital FRCL. Users below and at the target salt value of 8 grams per day defined by the Federal Food Safety and Veterinary Office are categorized as having a "Low" salt intake. Users with an intake between 8 and 12 grams are in the category “Medium”. All users with a higher salt intake than 12 grams per day are labeled “High”. The first research question can therefore be stated as follows:

Research Question 1: Can a user’s purchasing behavior stored in forms of automatically recorded loyalty card logs and respective ingredient data reveal the **risk level (low: <=**8 g NaCl / day, **medium:** 8 – 12 g NaCl / day**, high:** >=12 g NaCl / day**) of his salt intake** derived from his self-reported Food-Record-Checklist?

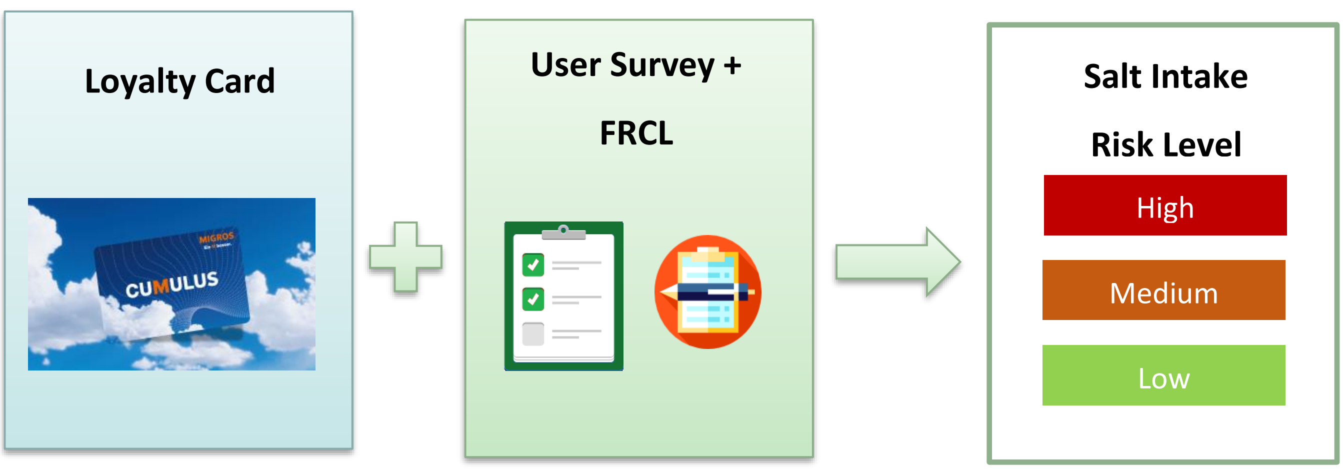


Figure 9 Research Question 1

3.4.4 Research Question 2: Salt Intake Prediction Model

For the second research question the idea is not only to categorize the users into one of three predefined risks levels but the actual salt intake value is approximated. Being able to predict the actual value of the salt intake, even more precise feedback can be provided to the user and interventions can be individualized. Therefore, the second research question can be formulated as follows:

Research Question 2: Can a user's purchasing behavior stored in forms of automatically recorded loyalty card logs and respective ingredient data reveal the **actual value (in g NaCl / day) of his salt intake** derived from his self-reported Food-Record-Checklist?

3.4.5 Model Description

To affirm or negate a correlation and to answer the research questions, several models are built and tested. In this section these models will be described. All models assume that consumers eat what they buy and every person in the household consumes similarly. The dependent variable is always the salt intake measured by the user's entries to the FRCL in the mobile application. Both the classification and the linear regression models are used to find a correlation, differentiating mainly in the target variables. Whereas the classification aims to categorize a user into one of the three predefined categories, the regression approximates the actual value of the user's salt intake.

In the following each model, in total 6, is explained. The input variables created for each model are used for the classification and the regression alike. Each model is tested with all subsamples described in 3.2.2. Because the models will be heavily based on food categories (defined by SGE[8]), it is important to state which categories are known to have a large impact on a person's salt intake and which do not. This has been identified in several studies (Beer-Borst et al.; Chappuis et al. 2011b; Constanza et al. 2004) and is shown in Table 4 (mapped to the SGE categories).

[8] http://www.sge-ssn.ch/ich-und-du/essen-und-trinken/ausgewogen/schweizer-lebensmittelpyramide/

Salt Intake Relevance	
High Impact Categories	- Salty snacks - Milk and dairy products (including cheese) - Fish - Meat and entrails - Meat and sausages - Convenience dishes - Bread and cereals
Low Impact Categories	- Sweets - Fats, oils - Nuts, seeds - Wheat, potatoes, rice - Fruits - Vegetables - Beverages/Drinks

Table 4 Salt Intake Relevance of Food Categories

Model 1: Predicting average daily salt intake from relative spending in different grocery categories (Major Categories)

The first model uses the relative amount of money spent on different product categories. For this purpose, the money spent on products belonging to one of the 13 major categories identified by the SGE is divided by the total money spent on food items. The hypothesis is that people who spend more money on categories known to have many products with a high salt share also have a higher salt intake.

$$\delta_1 = \frac{CHF\ spent\ for\ Category\ 1}{total\ budget\ spent\ on\ Food\ Items\ with\ Loyalty\ Card}$$

$$\delta_1 = Salty\ Snacks$$

$$\delta_2 = Sweets$$

$$\delta_3 = Fats, Oils$$

$$\delta_4 = Nuts, Seeds$$

$$\delta_5 = Milk\ and\ Dairy\ Products$$

$$\delta_6 = Fish$$

$$\delta_7 = Meat\ and\ Entrails$$

$$\delta_7 = Meat\ and\ Sausages$$

$$\delta_8 = Convenience\ dishes$$

$$\delta_9 = Wheat, Potatoes, Rice$$

$$\delta_{10} = Bread, Cereals$$

$$\delta_{11} = Fruits$$

$$\delta_{12} = Vegetables$$

$$\delta_{13} = Beverages/Drinks$$

Model 2: Predicting average daily salt intake from relative spending in different grocery categories (minor categories)

The logic for the second model is exactly the same as for Model 1, except that not only the 13 major but the 111 minor categories are used. Each minor category is assigned to a major category. Using minor categories, more differentiated statements can be made as major categories might be too broad.

$$\delta_1 = \frac{CHF\ spent\ for\ Category\ 1}{total\ budget\ spent\ on\ Food\ Items\ with\ Loyalty\ Card}$$

$$\delta_1 = Salty\ Snacks$$

…

$$\delta_{111} = Asian\ dishes$$

Model 3: Predicting average daily salt intake from relative weights in different grocery categories (major categories)

In Model 1 and 2 only the prices of the products were taken into account. This however bears the risk that costly products are overemphasized. Therefore, Model 3 and 4 (the first with major and the latter with minor categories) incorporate the weights. For each category, the weight of the purchased products is divided by the total weight of all purchased food items and used as independent variable. The weight is the total weight of the product as it was bought and not only the serving size which is defined by the manufacturer and can vary greatly even among similar products. Additionally, the total weight is not only more objective but the data quality in the PIID of the products' total weights is better compared to the serving size. The hypothesis is that consumers buying many products, which is reflected in the weight, from salt intensive categories have a higher salt intake.

$$\delta_1 = \frac{kg\ purchased\ in\ Category\ 1}{total\ kg\ purchased\ on\ Food\ Items\ with\ Loyalty\ Card}$$

$$\delta_1 = Salty\ Snacks$$

$$\delta_n = \cdots$$

Model 4: Predicting average daily salt intake from relative weights in different grocery categories (minor categories)

$$\delta_1 = \frac{kg\ purchased\ in\ Category\ 1}{total\ kg\ purchased\ on\ Food\ Items\ with\ Loyalty\ Card}$$

$$\delta_1 = Salty\ Snacks$$

…

$$\delta_{111} = Asian\ dishes$$

Model 5: Predicting average daily salt intake from salt intake in different grocery categories (major categories)

Model 3 and 4 have the potential shortcoming of overemphasizing heavy products, therefore Model 5 does not use the total weight of the product but only the weight of the included salt. The salt value is calculated by taking the amount per 100 grams and multiplied with the total weight of the product.

$$\delta_1 = \frac{kg\ Salt\ intake\ in\ Category\ 1}{kg\ Salt\ intake\ on\ Food\ Items\ with\ Loyalty\ Card}$$

$$\delta_1 = Salty\ Snacks$$

$$\delta_n = \cdots$$

Model 6: Predicting average daily salt intake from salt intake per product in different grocery categories (major categories)

Within categories the salt values vary a lot. For example, chickens have a much lower salt amount per 100 grams compared to beef or pork, yet all belong to the main category "meat". Model 6 therefore includes the salt intake of the average product bought for each category.

$$\delta_1 = kg\ Salt\ intake\ on\ average\ products\ within\ Category\ 1\ from\ Loyalty\ Card$$

$$\delta_1 = Salty\ Snacks$$

$$\delta_n = \cdots$$

CHAPTER 4. TECHNICAL IMPLEMENTATION

In the previous chapter the research questions and experiment setup including the required data sources such as 1) Food-Record-Checklist (FRCL), 2) user survey, 3) Product-Item-Ingredient-Database (PIID) and 4) loyalty card logs were explained. An integral part of this thesis project is the design and implementation of the technical infrastructure, services, and applications in order to translate the FRCL into a mobile application, provide feedback to the user and, most importantly, assess the relationship between the average salt intake measured by the FRCL and calculated out of the loyalty card logs. This chapter therefore examines the technologies, building blocks and structures used for the technical implementation of the developed software systems. First, a general overview of all systems and their interconnections will be given, followed by a closer examination of each of the major parts.

4.1 General System Overview

The most visible part for the user is the mobile application by the name "Swiss SaltTracker" which was made available to participants of the thesis study via the Swiss Apple iOS app and Google Android Play stores. The mobile application serves as the interface between the user and the remaining system setup. Within the application, the user can track his salt intake by entering all consumed food items in the respective observation period, i.e. four completed FRCL questionnaires equaling four days. The app is also responsible for collecting relevant user data through a mandatory user survey, e.g. as personal anthropological data (age, weight, etc.) and eating and shopping behavior (User survey questionnaire 7.1 in the appendix) and the user's credentials for the loyalty card login of his consumption records. After completing the required four daily FRCL protocols, the user is offered the possibility to receive an analysis of his salt intake. The analysis consists of two independent assessments: 1) The first part reveals a user's salt intake from eating behavior and comprises all data which were inserted by the user directly by logging his meals via the FRCL within the app (Figure 17). 2) The second analysis assesses a user's purchasing behavior and uses the data of the loyalty card logs. The provided analysis is primarily meant as an incentivization towards users to participate in the thesis study and to log their diet.

The mobile application represents the client side of the entire system and is connected to two independent backend services: 1) A service responsible for all PIID related data and logic and 2) a second service specifically designed for the "Swiss Salt Tracker" app which handles all user related operations. The two services could have been merged into one from a technical point of view, but as the data imported from the PIID is very likely to be used in future projects and is not related to the dietary intake measurement of the mobile application, a clear separation was the more future-oriented approach and increases the long-term benefit of this thesis. Both backend services are connected to a separate database each. The mobile application does not interact with the databases directly but exclusively communicates with the services. Figure 10 shows a high-level overview of the entire system.

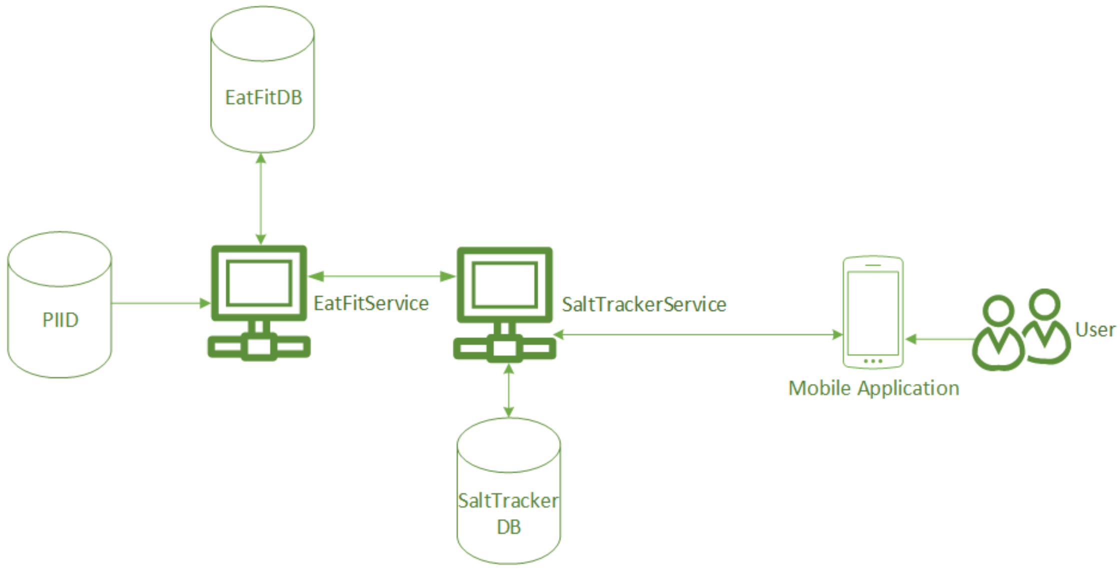


Figure 10 General System Overview

4.2 System Requirements

Before explaining the system components in detail, it is important to state the requirements the system must fulfil.

Mobile App

The mobile application must be able to let the user add records to the digital FRCL and provide a way to collect user-specific data such as loyalty card credentials or personal information like age or gender. Further it is intended to present the user a short summary of his salt intake over the course of the experiment. The transformation from the validated paper-based version of the FRCL to its digital counterpart must be performed in a way so that it is still convenient for the user to add records, commonly used user interface design patterns in health and fitness apps must be taken into account and the content of the FRCL must remain exactly the same. Otherwise any conclusion based on the FRCL would no longer be valid.

SaltTrackerService

The SaltTrackerService must act as the backend to the mobile application and manage the connection to the database. It must allow the user to easily switch devices, without having to manually migrate data. Further, the integration of a solution to import the loyalty card logs is an integral part of the SaltTrackerService.

EatFitService

The main purpose of the EatFitService is to import and manage the data from the PIID. It must further be possible to categorize the imported products according to the food categories defined by the "Schweizerische Gesellschaft für Ernährung" (SGE). The entire service must be written in a generic way and must be completely decoupled from the mobile application.

4.3 Backend Services

4.3.1 Frameworks

Many frameworks and technologies exist for developing web services in various programming languages. For this thesis, it was important to use a flexible and easily extensible framework. The web development framework Django[9] was chosen which is written in Python. More precisely, the Django Rest Framework (DRF)[10] was used which is, as the name suggest, a collection of libraries built ontop of Django to support the development of REST services in Django. REST (Representational State Transfer) is a common architectural type to write web ser-

[9] https://www.djangoproject.com/
[10] http://www.django-rest-framework.org/

vices and is well suited for stateless communication to clients and allows for a clear and strict separation between server and client (Fielding 2000).

The communication is handled over the HTTP protocol and JSON was chosen as data exchange format (as it causes less overhead compared to XML and nicely fits the DRF). Django also offers an object-relational mapping (ORM) framework which facilitates and abstracts the development of the database and allows the programmer to keep the business logic inside the server application almost entirely. To reduce technology breaks, both services are developed in Python and Django.

4.3.2 Hosting

All services and databases are hosted on Microsoft Azure[11]. Azure is a cloud computing platform developed by Microsoft and helps to alleviate the deployment process and aim to reduce the server configuration and maintenance efforts. There exist several hosting possibilities in Azure, which mostly address the trade-off between reducing configuration and maintenance and keeping as much control over the server and the application as possible. For the two services, the hosting product best suited was chosen. The EatFitService, a more generic service acting as the bridge to the PIID, is hosted on a dedicated virtual machine running Ubuntu. This allows having full control over the machine, which is important as future projects might have additional or different requirements. The SaltTracker-Service, which is fully dedicated to the "Swiss SaltTracker" mobile application is hosted as "Azure App Service"[12]. The App Service hosting plan greatly facilitates deployment and the underlying hardware is completely abstracted. It is only possible to deploy the application and no direct access to the physical (or virtual) machine is possible. In exchange, many features like load balancing or request monitoring are already built in.

[11] https://azure.microsoft.com/
[12] https://azure.microsoft.com/de-de/services/app-service/

The databases for the two services are also hosted on Azure in a "database-as-a-service" plan. This again entirely removes the necessity to setup the database server. As a drawback, the only database management system supported is Microsoft SQL Server.

The main advantages to host the services and databases on Azure over a traditional hosting approach are the reduced maintenance and configuration complexity and the dynamic scaling of the instances. Every product on Azure (virtual machines, databases etc…) can be scaled-up by simply selecting a different instance type with different hardware. This allows to start with relatively weak hardware and increase memory, storage or CPU power when more resources are needed and more traffic is generated. The change from one hardware configuration to anther is only a matter of seconds and is performed automatically.

4.3.3 Data Import

The most integral part of the EatFitService is the connection to the PIID. Although the PIID does provide an API, it is not intended to be queried directly in order to avoid high traffic on the PIID servers. The data must rather be imported from the PIID into a self-managed database. The import of the data can be performed either manually by downloading a JSON-dump from an FTP-server or it can be fetched from a SOAP API, which supports both JSON and XML. The latter approach was chosen, as this allows to import the data automatically and regularly. To perform the import, the data structure of the PIID first has to be recreated in a database. The data is distributed over 11 tables; the exact schema can be found in the appendix (7.2). Besides the nutrition facts, data for protein, carbohydrate, salt etc. and other product related information like images, package size or allergens are stored. As the PIID is a nutrition database for the entire country of Switzerland, many text strings are available in the three national languages German, French and Italian (but not Rhaeto-Romance) and English. To ensure that changes in the PIID after the initial data import are also reflected in the EatFit database, a periodically executed task fetches updates from the PIID, parses the JSON-response and inserts or updates the data in the EatFit database.

The schema of the EatFit database is similar to one of the original PIID with additional tables and columns for user management, categorization, logging, and periodic tasks and is included in the appendix (Figure 34 and Figure 35).

4.3.4 Data Preparation and Correction

As already mentioned, the data quality in the PIID is certainly high but has still room for improvement. For the thesis project, it is crucial to have as many products with an accurate salt value as possible. Instead of a simple data cleansing, where invalid data would be removed, faulty entries are tried to be corrected. As only salt values are relevant for this thesis, the data quality improvement process was only applied to incorrect salt entries. By “incorrect” a wrong number format is meant, the correctness with regards to the actual content was not verified at a large scale. For improving the data, the following rules were applied:

- Additional text was completely removed, only leaving the numeric amount. Regular expression matching was used to extract the numeric values. Example: *up to 0.5 g -> 0.5*
- Mismatches between comma and decimal point were harmonized. Example: *0,5 -> 0.5*
- When the salt value was completely missing but the sodium value was available, the salt value was calculated using the sodium value and the relevant conversion multiplier of 2.548 (Beer 2009).

This correction of values is also performed automatically on newly imported data by the periodic task.

4.3.5 Categorization

After importing the products, they need to be categorized. The categorization cannot be performed automatically but is conducted manually by the SGE and also manually applied to the EatFit database via the REST interface.

4.3.6 Missing Items

First tests with real loyalty card log data showed that approximately 30-40%[13] of the bought and relevant products could be found in the PIID. Products are only considered relevant, if they can be counted as nutrition, items like shipping bags or toiletries as well as any other non-food products were not utilized. The share of products found in the PIID can of course vary drastically from one user to another. However, a 30-40% coverage is very likely too low for a significant analysis, especially if the share of products bought at the retailer is also relatively low. Therefore, missing items in the PIID had to be added manually. The possibility to add these items into the already existing data structure was not further pursued as it would result in conflicts or duplicates when a formerly missing item would be added to the PIID and imported into the EatFit database.
Thus, an additional database table was created, only consisting of the most relevant product data. Table 5 shows the information stored in the table.

[13] Based on tests with two real loyalty card logs and data from 2 years

Missing Product Items Table
Name of the product
Total weight in g
GTIN
SGE Category
Image icon URL
Price
Salt per 100 g
Sodium per 100 g
Energy per 100 g
Fat per 100 g
Saturated Fat per 100 g
Carbohydrate per 100 g
Sugar per 100 g
Protein per 100 g
Fibers per 100 g

Table 5 Missing Product Items Content

To provide a convenient way to enter this data, a rudimentary, web-based administration interface was created (a screenshot is shown in Figure 11). Despite adding missing items to the database, it cannot be guaranteed that all bought products of a new user are present in either the PIID or the manually added data, as new products are added regularly by the retailer and a new user might have simply purchased different products than the test users. This means that the process of adding missing products should be repeated periodically in order to guarantee a high coverage.

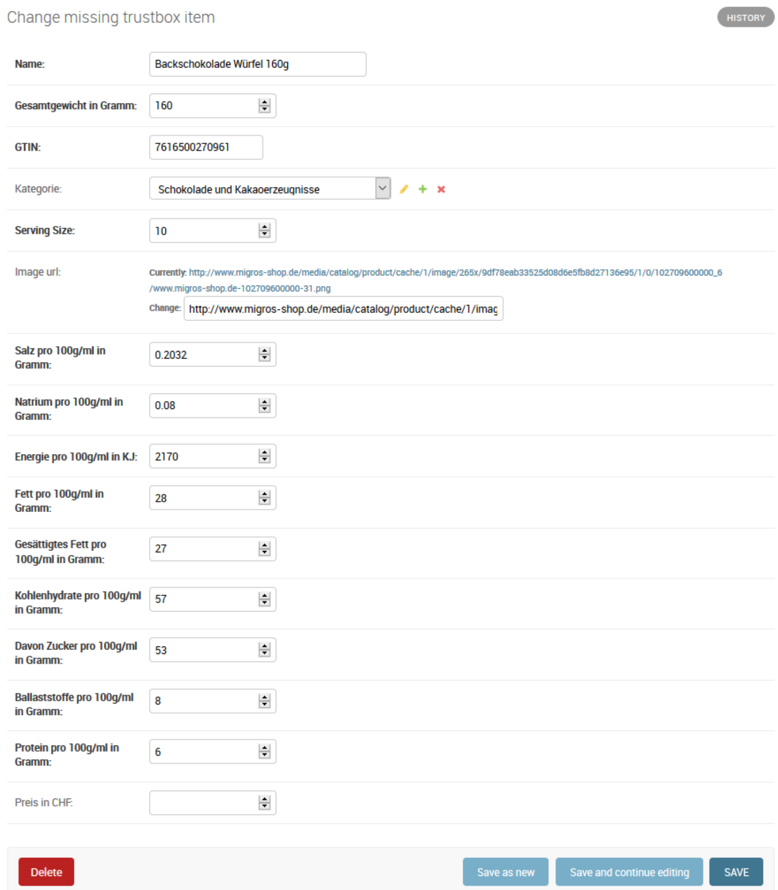


Figure 11 Administration Tool Missing Items

4.3.7 App Connection

The main purpose of the SaltTrackerService is to serve as the backend for the mobile application. All information entered into the mobile app is immediately transferred to the service and stored in the SaltTracker database. This architecture has many advantages over a traditional approach where the data would be stored locally on the device.

- **"Thin" Client**

  The main part of the business logic and computationally complex calculations remains on the server. The mobile app (e.g. the client) can be kept "thin", meaning that only light-weighted tasks are directly performed on the device and no complex logic is included.

- **Synchronization and Concurrent Usage**
  As all information is present on the server at any time, it is very easy for the user to switch between devices. Only the user credentials have to be reentered in the mobile app and the entire personal data is available on any device. Further, it is even possible to use the application concurrently on multiple devices without the danger of conflicting states.

- **Updates**
  To provide a new version of a mobile application, a new build has to be submitted to Apple's or Google's app stores, which further distribute updates to the users. This process takes several hours (in case of the Google Play Store) or even several days (Apple App Store). As some parts of the mobile app uses data which is directly loaded from the server (often with a caching mechanism) rather than being hard-coded, these parts can be updated without changing the actual mobile application's source code and a submission to the app stores is therefore not necessary in some cases.

Almost every functionality in the mobile application has its counterpart in the SaltTrackerService. To preclude explanatory redundancy, the description of the functionality is only covered in section 4.4.3 together with the implementation of the mobile application.

4.3.8 Import of Loyalty Card Logs

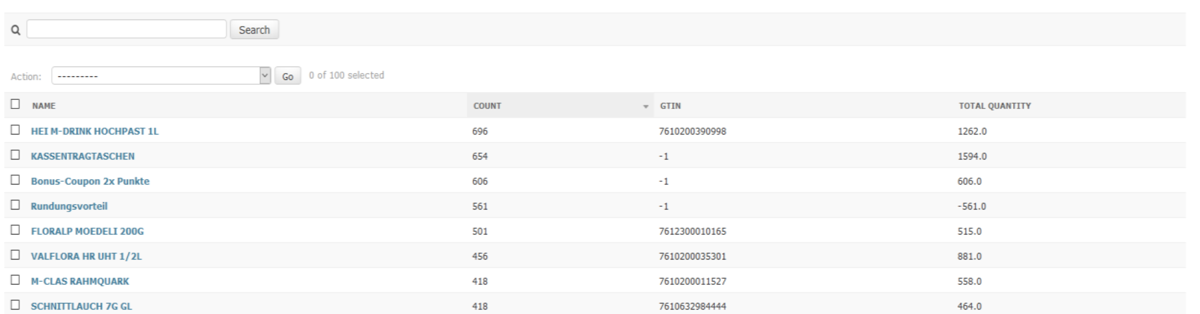
The loyalty card logs used for the analysis in this thesis origins solely from one retailer. The retailer persists all products of a purchase when the loyalty card is used when paying. Unfortunately, the retailer does not provide a public API to query this loyalty card logs. Instead, there is a web portal on the official webpage where a registered user can review his purchased products of the last two years. Therefore, the only possibility to receive the data is to login to the web portal, download the purchase history and extract the products. To perform these tasks automatically, an independent tool to import the data was developed. First, the tool has to login with the credentials provided by the user. After navigating to the relevant webpage, the last 20 purchases are shown. To obtain all purchases of the last 2 years, the date range must be changed which yields the paginated results (again 20 per page). For every purchase, the corresponding link has to be followed and the resulting HTML needs to be downloaded. From the HTML code, the actual product data can be extracted and saved to the database. The tool to import the loyalty card logs was developed by the Auto-ID Labs.

4.3.9 Matching

Using the tool for importing the loyalty card logs, the log data for a specific user can be obtained. Unfortunately, the logs only show an abbreviated name of the product (together with the price and the quantity) in the web portal. These names are the same as the ones printed on the physical receipt in the store. In the PIID however, the complete product names and the Global Trade Item Number (GTIN)[14], which identifies a product uniquely worldwide, are stored. The linkage from a product in the loyalty card log to a product from the PIID (and also the manually added products) is therefore nontrivial and cannot be achieved with a simple string equality test. As the matching of products names on receipts to actual products names is not a core topic of this thesis, it was decided to conduct the matching manually. Although first tests with measuring string similarity (for example with the Levenshtein distance (Hicham, Abdallah, and Mustapha 2012)) looked promising, it was not enough to render manual checks unnecessary. Further work could include a matching based on several product sources (similar to (Karpischek, Michahelles, and Fleisch 2014)) or to use machine learning to find out how names are abbreviated and which parts of the names should be exclud-

[14] https://www.gs1.ch/gs1-system/das-gs1-system/barcodes-identification/gtin

ed (for example quantity declaration) or given more weight. Like for the manual addition of products missing in the PIID, a web-based administration tool was created to support the manual matching process (Figure 12).



NAME	COUNT	GTIN	TOTAL QUANTITY
HEI M-DRINK HOCHPAST 1L	696	7610200390998	1262.0
KASSENTRAGTASCHEN	654	-1	1594.0
Bonus-Coupon 2x Punkte	606	-1	606.0
Rundungsvorteil	561	-1	-561.0
FLORALP MOEDELI 200G	501	7612300010165	515.0
VALFLORA HR UHT 1/2L	456	7610200035301	881.0
M-CLAS RAHMQUARK	418	7610200011527	558.0
SCHNITTLAUCH 7G GL	418	7610632984444	464.0

Figure 12 Administration Tool Matching

4.3.10 Calculation of Results

As described in section 4.1 the user is provided with an elementary assessment of his salt intake over the course of the experiment after tracking his diet over the course of four days. The four-day timeframe should be sufficient to collect a user's salt intake and still be manageable without losing the motivation for the self-reporting. The original paper-based FRCL is filled over the course of three days but with a personal introduction, therefore an extra day was added in the app version to become familiar with the concept.

The first part of the analysis, the results of the salt intake directly measured by the user's entries, is calculated on the fly. For the second part, the analysis of the data from the loyalty card log, the calculation is performed asynchronous to avoid a too large waiting time for the user in the mobile app. The relatively long execution time for this part of the results is due to the fact that data from two databases is involved and that more application logic is necessary. A periodic task calculates the results of the analysis of the loyalty card logs for every user and stores them in the database. Upon a request by the mobile app, the SaltTrackerService serves the pre-calculated data. As all other features related to the mobile app, the concrete content of the assessment presented to the user is described in section 4.4.3.

4.3.11 Backend System Overview

To summarize the implementation of the backend services and processes Figure 13 shows all relevant backend components and the processes needed to obtain and transform the raw data from the PIID and the loyalty card logs in order to make the data useful for this thesis and the mobile application.

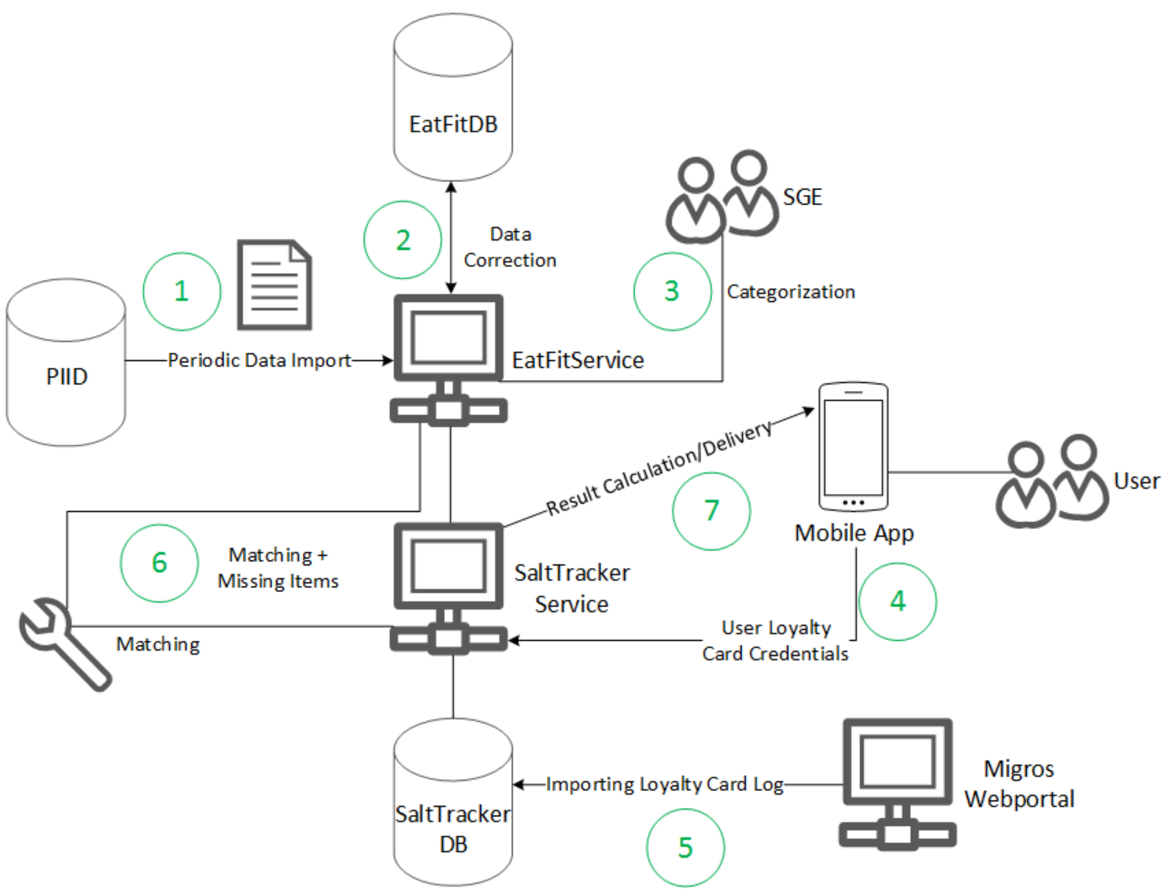


Figure 13 Process Overview

Step (1) denotes the periodic data import from the PIID to the EatFitService. The data is corrected (2) and stored to the EatFit database and categorized by the SGE (3). Using the loyalty card credentials entered by the user (4), the loyalty card log can be imported (5). The products in the loyalty card log have to be matched to the products of the PIID and missing items are created if necessary (6). After completing the study, the SaltTrackerService calculates the nutrition assessment and delivers the results to the mobile application and the user (7).

4.4 Mobile Application

4.4.1 Framework

When developing a mobile application, there exist several frameworks to choose from. The first decision to take is whether the app should be developed "native", e.g. using the official SDKs (software development kit) from Google and Apple or with a third-party framework. Most third-party frameworks abstract the underlying native frameworks and use different programming languages (for example HTML in combination with JavaScript). These frameworks have the disadvantage that the resulting GUI (graphical user interface) does not have the native look and feel and often suffers from inferior performance compared to native implementations (Sommer and Krusche 2013).

To reach a wide range of users it is important to support all major mobile operating systems. Especially in Switzerland, where both iOS and Android are almost equally distributed, an app supporting both operating systems is advantageous.
As the development of two native applications would exceed the scope of the thesis project, it was decided to use a cross-platform framework. To avoid the shortcomings of non-native development, Xamarin[15] was chosen as the development framework.

[15] https://www.xamarin.com/

Xamarin, in May 2016 acquired by Microsoft[16], is a company developing a framework of the same name to support the creation of mobile applications running on Android, iOS and Windows 10. Using Xamarin, there are two separate possibilities to develop mobile apps. Using the first possibility, the application logic is written in the programming language C# (and can therefore be shared among all platform) but the graphical user interface is written in the respective native language and tools (e.g. axml for Android). The second possibility is to use Xamarin Forms, which does not only allow to share the logic but also the frontend code. In contrast to most other cross-platform frameworks, the user interface code is not placed in a layer above the native user interface but all graphical controls and views are renderer to the corresponding native controls and views. For example, when a "search- bar" is created in Xamarin Froms, it is renderer to a "UISearch-Bar" on iOS, a "SearchView" on Android and a "AutoSuggestBox" on Windows 10[17].

This allows to create apps with a single codebase; without having to create several distinct app projects in the respective native languages. Of course, this approach does also have some disadvantages. As C# is used for development, the Mono Runtime (which allows to use C# on other platforms than Windows) needs to be started-up when the mobile app starts which results in extended launch times. Further, for the development of the graphical user interface the markup language XAML is used. Compared to the native tools to create user interfaces, the Xamarin Forms approach is certainly not yet as advanced and takes more time to create comparable layouts (in terms of performance and design). Nevertheless, it was decided that the benefit of being able to create an iOS and an Android app with a single codebase outweighs these shortcomings.

[16] https://blogs.microsoft.com/blog/2016/02/24/microsoft-to-acquire-xamarin-and-empower-more-developers-to-build-apps-on-any-device/#sm.0000nqtoj4u2eejtvbj1c4liun3vd
[17] See https://developer.xamarin.com/guides/xamarin-forms/custom-renderer/renderers/

Further, it is possible to inject native code where the layout capabilities of Xamarin Forms are too limited. This however means, that this code has to be written for all targeted platforms and a solid understanding of the respective platforms is necessary. This concept is called "custom renderers" in Xamarin and was also used several times while developing the "Swiss SaltTracker".

4.4.2 Architecture

A very important part of an architectural concept for any application which lets the user interact graphically, is the definition of the interplay between the logic, data and the graphical user interface. As general architectural design, the Model-View-Viewmodel (MVVM) pattern was applied for the "Swiss SaltTracker" app. A design pattern in software engineering can be best thought of as a "best practice approach", which has proven to be an ideal solution to an architectural problem. The MVVM pattern is a modified version of the well-known Model-View-Controller (MVC) pattern which is a so-called compound pattern, consisting of the Stategy, Observer and Composite patterns. The goal of this pattern is a strict separation between the view, e.g. the user interface, and the business logic. (Freeman and Freeman 2013)

The use of the MVVM pattern, introduced by John Gossman (Gossman 2005) and based on the Presentation Model by Martin Fowler (Fowler 2006), is often the recommended way of developing applications when using Microsoft frameworks. The MVVM pattern can also be considered to be a Presentation Model specifically applied to the Microsoft frameworks and architectural design guidelines (Smith 2009).

The "View" component in the MVVM pattern is meant to be kept as "thin" as possible, which means that only the structure and the design of the user interface should be defined, when possible in a declarative language like XAML. The "Model" represents the pure data and the logic. So far this is very close to the

MVC pattern. However, the newly introduced component, the “Viewmodel”, is intended to be used differently than the “Controller” in MVC. In MVVM the “Model” and the “View” should not be connected in any way, the entire communication is carried out to the “Viewmodel”. The “Viewmodel” is responsible to retrieve the data from the “Model” (potentially by triggering functionality in the “Model”) and hand it to the “View” or change the structure of the “View”. The MVVM pattern further relies heavily on data binding. With data binding, a graphical control in the “View” is bound to a property or attribute in the “Viewmodel”. Updates, either in the “View”, triggered by the user or in the “Viewmodel” are directly synchronized. Using this design, the “View” can be created completely separated from the “Model” and can even be switched out with ease. Further a “Viewmodel” can be connected to several “Views”, reducing code duplication.

4.4.3 Design and Functionality

The design of the mobile application is inspired by other mobile apps in the health or fitness category. The primary application color used is green and the icons and images are mostly comic-style. The goal was to create a clean and friendly design which encourages the user to use the app. For the main navigation, a master-detail[18] interface is used (also known as hamburger menu) which enables to user to reach any functionality in the app from any screen. A this is a common way for handling navigation in an app, most users should be familiar with the concept.

Registration and Login

The first important part of the mobile application is the registration process. The purpose of the registration is, besides the obvious creation of a user account

18 https://developer.apple.com/library/content/documentation/Cocoa/Conceptual/CocoaBindings/Tasks/masterdetail.html

(which allows the user to use the app across different devices), to gather additional information from the user.

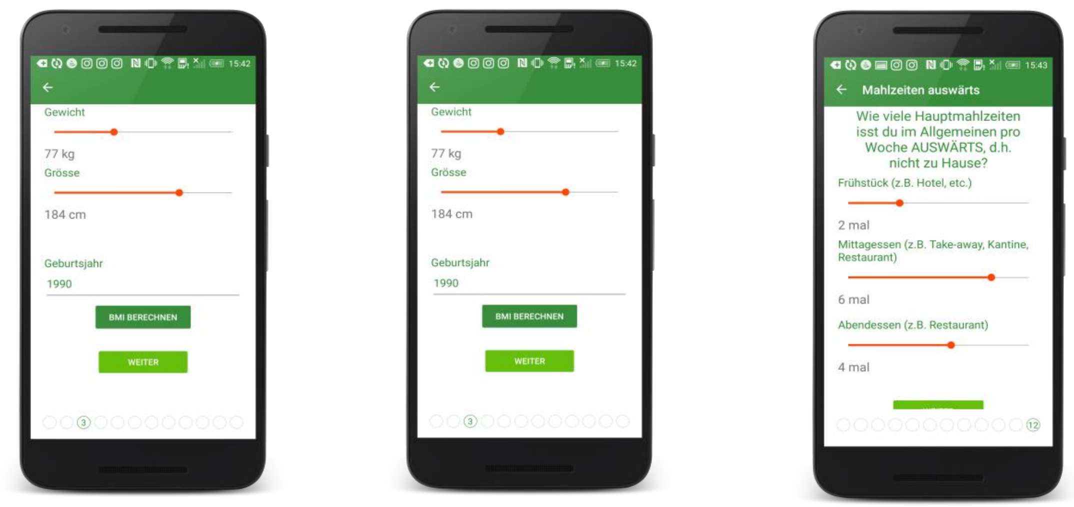


Figure 14 App - Registration

The content of the registration is described in 3.2.2. Figure 14 shows parts of the registration process and the questionnaire screen. The registration is started from the welcome or login screen (Figure 15).

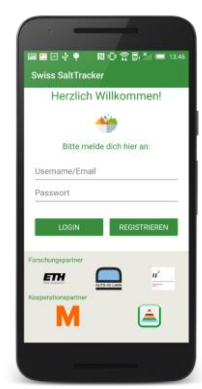


Figure 15 App - Login Screen

Home Screen

The home screen (Figure 16, a) is the entry point for registered users as the app makes use of single sign-on, which means that once a user is authenticated (e.g. has provided the credentials and the server has successfully verified them), the app automatically signs the user in when the app is started. In the upper half of the screen the overview over the last week is shown, indicating for which days, food records have already been entered. Food records can only be entered three days back as the quality of entries for days further in the past is expected to be relatively poor. By clicking the respective day, the user is prompted to insert a food record for the chosen day. The middle part of the screen is state dependent. It always provides a short link to the not already closed records. In the first four days, it displays the number of days for which records need to be entered until the assessment of the salt intake records and the loyalty card log is available. After completing four days (Figure 16, b), a brief summary of the average salt intake over the last seven completed days as well as a relative comparison to the upper limit of 8 grams of salt is displayed.

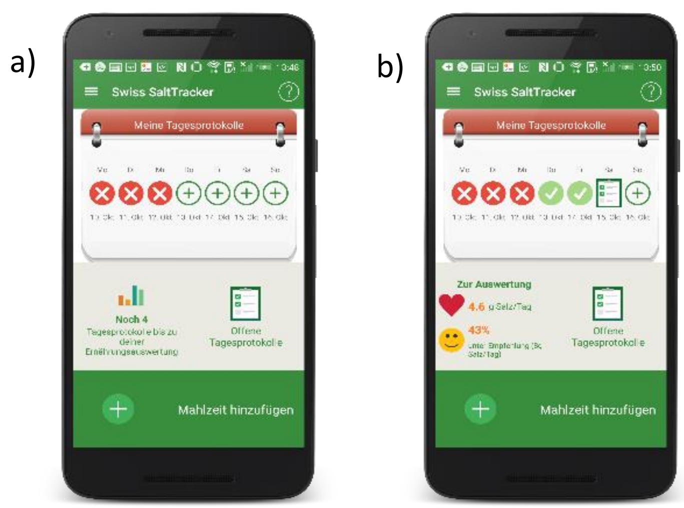


Figure 16 App - Home Screen

Food-Record-Checklist Screen

When adding a new meal or food, a screen for choosing from the categories of the FRCL is presented as a grid. This allows the user to choose all categories out of which the consumed meal consisted. Using a grid, the user has immediately the overview over all existing categories and does not need to scroll a list. By clicking the “record” button, a new screen with the food items for the chosen category is displayed. When multiple categories are selected, the user is guided through several screens with food items, each belonging to one of the selected categories. Besides the grid view, there is also a screen for adding recently recorded items (as a shortcut) and a view for a text search, all included in a tabbed page (Figure 17).

Figure 17 App - FRCL Screen

Food Items Screen

Subsequent to the food-record-checklist screen, the food items can be selected (Figure 18). All food items belonging to a category are displayed as a list and the quantity can be specified with a numeric stepper. The additional information that is also present in the paper version of the FRCL, e.g. information about the reference portion or more detailed examples, is also displayed next to the food item.

Figure 18 App - Food Items Screen

Open Records Screen

When adding a record for a day, this day appears on the open records screen (Figure 19). The user can either delete or “close” the day or view (by clicking the “details” button) all previously recorded food items for the day. When the user closes the day, it is not possible to add any food items for that day, forcing the user to commit to the recorded food items.

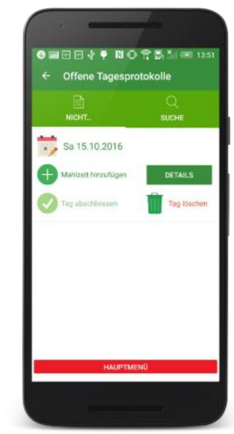


Figure 19 App - Open Records Screen

Assessment Screen

The assessment is only reachable if the user has already recorded food items for four days and entered the loyalty card credentials. The results are divided into two parts, the first part incorporating the results from the FRCL and the second part related to the loyalty card log, both embedded in a tabbed page. As the results only have informational character and are not meant to act as an intervention, a disclaimer stating exactly such is displayed on both pages.

The first part of the results informs the user about his salt intake using different dimensions and metrics. To give a general overview, the average salt intake per day as well as how many days the user remained below the critical value of 8 grams are displayed (Figure 20, a). Below, a line chart is presented showing the salt intake over time with the salt in grams on the y-axis and the time in days on the x-axis (Figure 20, b). The line chart provides the information whether the salt intake is equally distributed or varies from day to day. Further, the salt intake is broken down to an average day, displaying the average salt intake for breakfast, lunch, dinner and in-between meals, helping the user to understand when he consumes salt throughout the day. The last chart presented shows the top-five categories from the FRCL with the largest impact to the user's salt intake averaged per day (Figure 20, c).

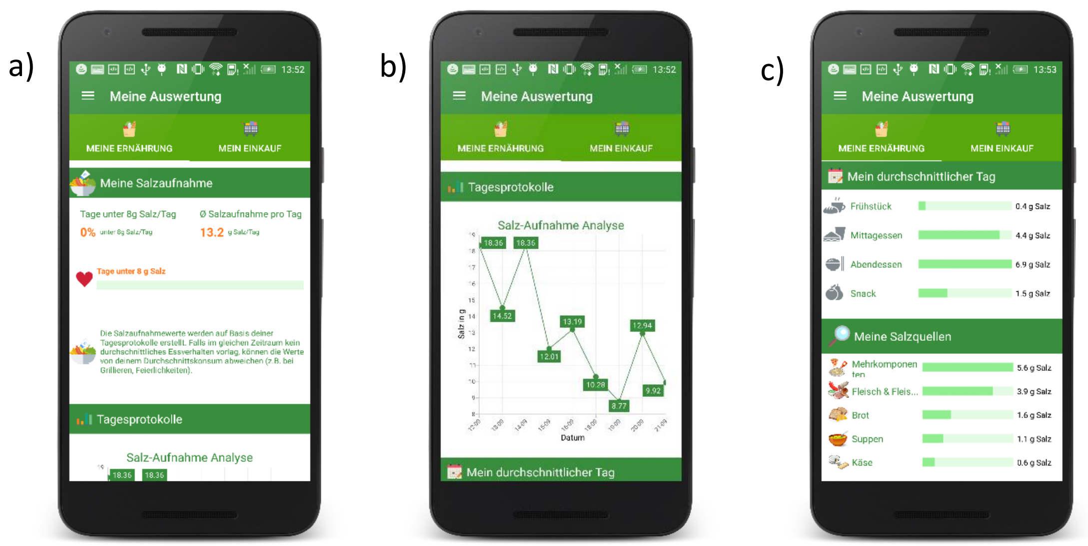


Figure 20 App - Assessment Screen Part 1

The second part of the assessment is similarly structured, also starting with general information, namely the number of baskets or purchases that were taken into account for the results and the salt included as an average per basket (Figure 21, a). Using a bar chart, the salt included in the different baskets together with the date of purchase is presented (Figure 21, b). If several purchased were conducted on the same day, they are combined into one in the graph. The chart's intention is to give the user an idea of how much salt the purchased baskets included. As the chart displays purchases up to two years back, it is also possible to spot seasonal differences. To give insights where the salt in the baskets comes from, an average basket divided into categories, again the top-five, is presented (Figure 21, c). As the PIID data is categorized by the SGE, these categories are used. The final part of the assessment is called “shopping tips” (Figure 21, d). For each of the five categories (only from the loyalty card log) presented, a healthier product alternative is suggested. Healthier in this context means, that a product must have less salt but at most one-third more sugar or fat than the average product in the respective category. In other terms, for each category the products are ordered ascending according to their salt value. Then the average sugar and fat of all products belonging to this category is computed. Starting with the product with the lowest salt amount per 100 grams, the first product whose sugar and fat values do not exceed the average sugar and fat values of the category by one-third is chosen as a “shopping tip” for the category. This procedure is repeated for all categories displayed to the user. Additionally, for each category a hand-written tip with additional salt-related information is also displayed. The “shopping tips” are a first step into a more personal and accurate analysis of an individual’s shopping behavior and the starting point for interventions.

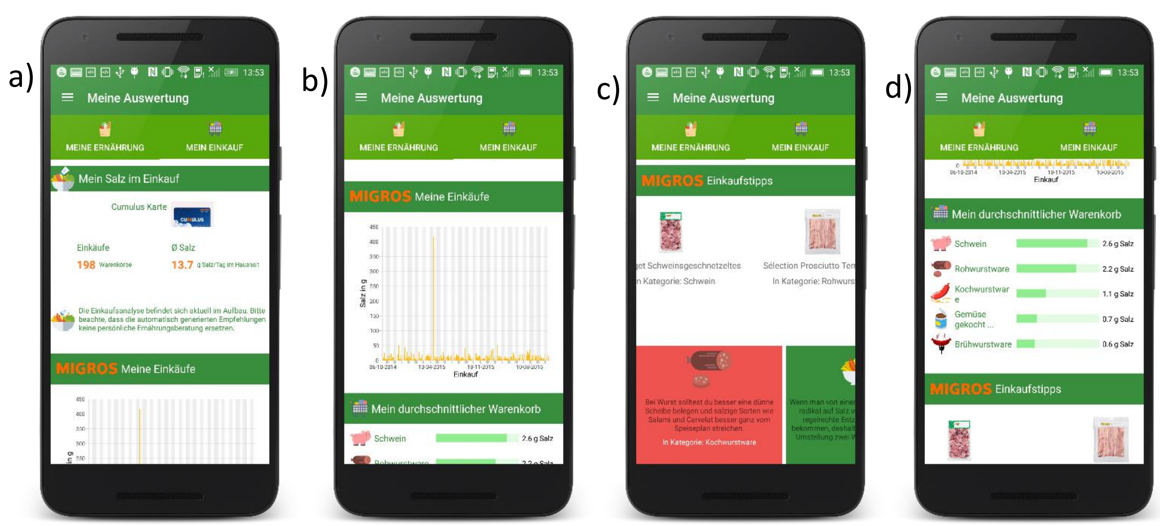


Figure 21 App - Assessment Screen Part 2

From a technical perspective, the computation of the results of the loyalty card log is much more involved compared to the results for the FRCL. First the loyalty card log has to be imported from the retailer's web interface which is currently performed in a stand-alone periodic task and the results are written to the database. A second periodic background task fetches these data from the database and maps the products from the loyalty card log to PIID products (or the products that have been manually added). As previously described in section 4.3.9, a matching table was created in the database and records were manually added. As the PIID products and the purchased data are on different databases (because the PIID database is to be used for different projects as well), several SQL queries are needed to gather the data and tables cannot be automatically joined (as this is not supported by Django and Azure). Therefore, for each purchased item the corresponding PIID item must be found by making a look-up with the name to the matching table which grants the product's GTIN. A second look-up is required to find the product via GTIN in the PIID data. Having found the product, the nutrition facts can be collected. What makes the procedure even more costly is the structure of the PIID data, which are often stored as key-value pairs, for example the nutrition fact names as keys and the actual amounts as values, all in one table. Such a structure is not ideal for relational database queries via SQL. Because this process must be repeated for every product and for every user it is outsourced in a periodic task and the results are stored in the database from which they are served to the mobile application.

Other Screens

In addition to the presented screens, there are various other screens for supporting or informative functionality. Examples are screens for changing the personal information and one-time question entered during registration or an overview of the associate partners and sponsors of the project. Further, in the first week after the installation of the app, the user receives a brief, daily notification which remembers him to record his diet.

CHAPTER 5. RESULTS

This chapter starts with the description of the users participating in the study followed by the presentation of the data analysis and the results. The thesis study was conducted over the course of five weeks starting with the launch of the mobile application on September 4th and finished on October 10th 2016.

5.1 User Distribution

In this section the properties of the participating users will be outlined. All "Swiss SaltTracker" users were invited to participate in the study. To complete the study, a user was asked to complete four Food-Record-Checklists (FRCL) on four different days and enter the loyalty card credentials.

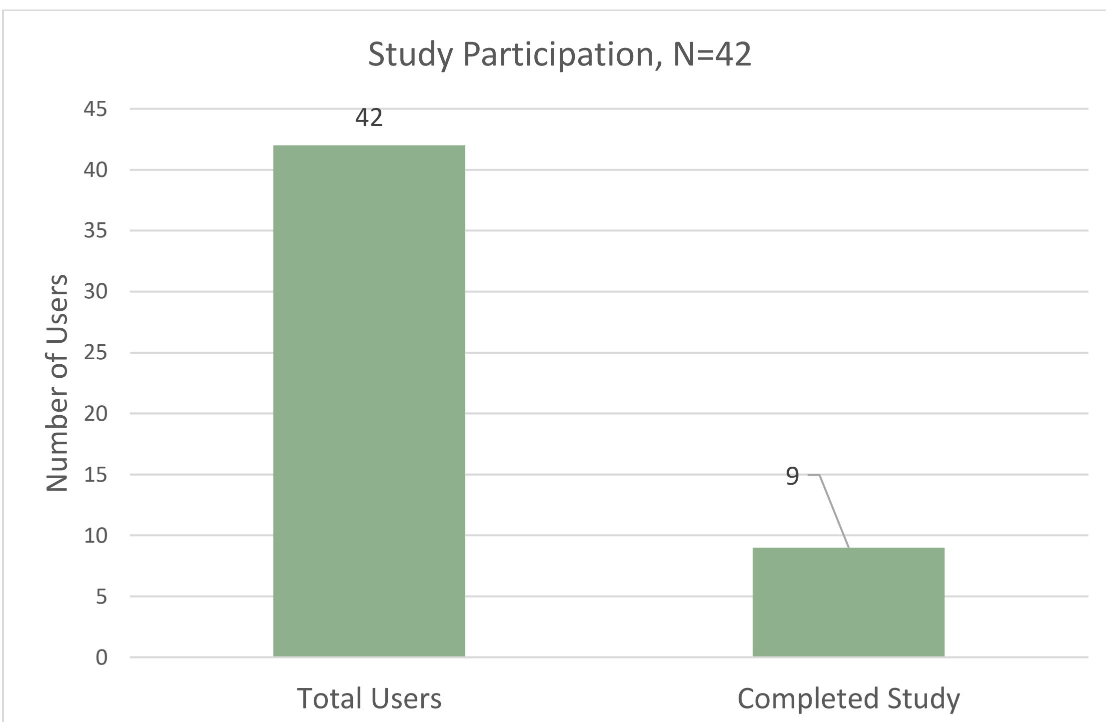


Figure 22 Study Participation

The study participation is illustrated in Figure 22. 42 users downloaded the app until October 10th 2016 and 9 (21.4%) also completed the entire study. Out of all users, 25 (59.5%) completed at least one FRCL and 17 users (40.4%) completed at least four FRCL but not necessarily also entered their loyalty card credentials.

The completion rate and the sample size might seem relatively low at first. Considering the time and effort for a user participating in the study, both the completion rate and the sample size are acceptable. The first obstacle are the four days a user must report his diet, the second are the credentials for the loyalty card. As it was learned during the course of the experiment, the larger part of the users did not know their loyalty card credentials and did not have an account on the web portal of the retailer. The only possibility to receive the loyalty card credentials is by asking the retailer to send the credentials by mail (physical). For many users, this step was too much of an effort and they decided not to participate in the study or not even to download the app. A further reason can also be privacy concerns or a user simply does not want to share his data. For these issues, however, a solution can be found in the future. More time and an intensified advertising of the mobile application (as proposed in 3.3) will eventually lead to a larger sample size which allows to repeat and refine the analysis in this thesis.

Another interesting property is how many FRCLs the users entered in average and how the distribution of entered FRCLs looks. Among all users the average number of recorded FRCLs is 4.35 with a standard deviation of 1.68. The mean is just slightly above the required four days to complete the study and enables the user to receive his nutrition assessment. Figure 23 shows the (kernel) density function of the completed FRCLs for all users who completed at least one FRCL. Like the average indicated, most users did not complete four FRCL or stopped shortly afterwards, very few users completed 10 or more FRCLs.

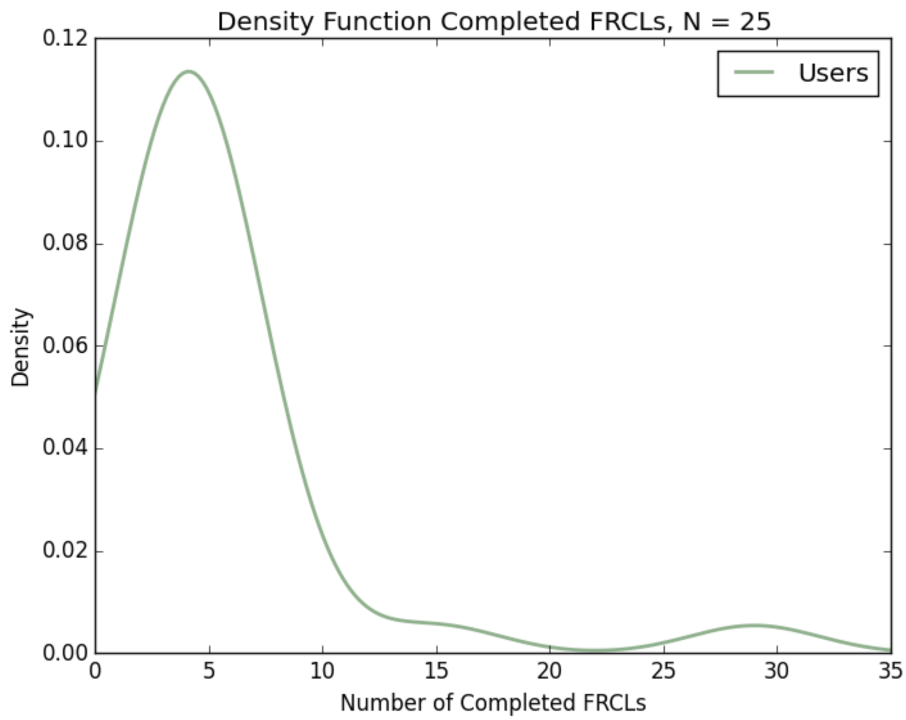


Figure 23 Completed FRCLs Density Function

Age Distribution

60% of the participating users were male and the average age was 38.9 years (standard deviation was 14.9). Figure 24 shows the distribution of the ages of the users. All users are either between 24 and 39 or 49 and 61, the population group with ages between 40 and 50 and above 60 are missing. The minimal age to participate in the study was 18 years. The age distribution indicates that the population is only partially represented in the sample.

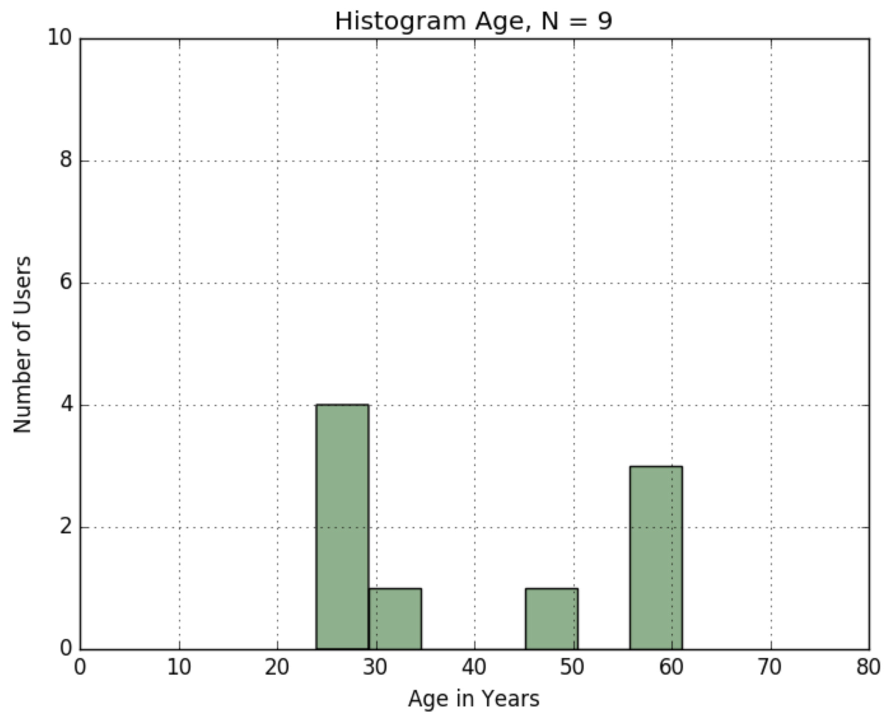


Figure 24 Age Distribution

Salt Intake Distribution

Analyzing the Swiss population's salt intake, it was found that the average daily salt intake is higher for men (10.6 g) than for women (8.1 g) (Beer-Borst et al. 2009). Figure 25 shows the salt intake distribution of the study participants using a boxplot. A box depicts the lower and upper quartiles of the sample, the lines, which are called whiskers, range from the lowest to the highest values (above the upper and below the lower quartile). As in the entire Swiss population, the salt intake of the male participants (average 10.3 g, median 10.1 g) is also significantly higher compared to the female (average 7.3 g, median 7.95 g) participants.

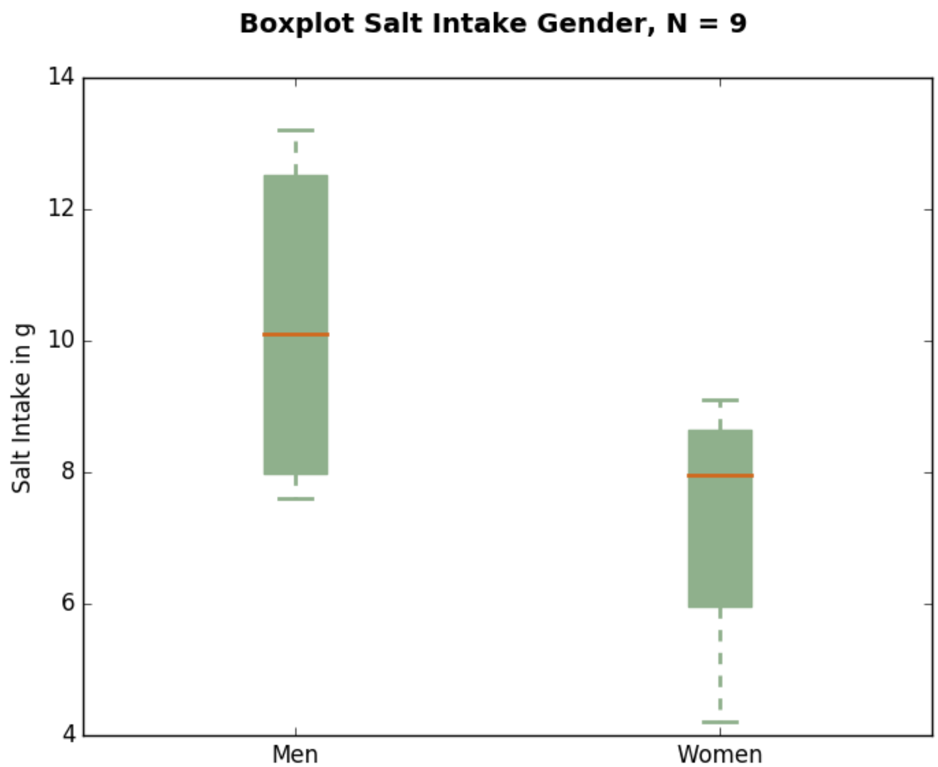


Figure 25 Salt Intake Distribution Gender
Medians are depicted in orange,
For the “Male” box the end of the lower whisker is at 7.6, the upper end at 13.2, the box ranges from 8.1 to 12.3
For the “Female” box the end of the lower whisker is at 4.2, the upper end at 9.1, the box ranges from 5.95 to 8.65

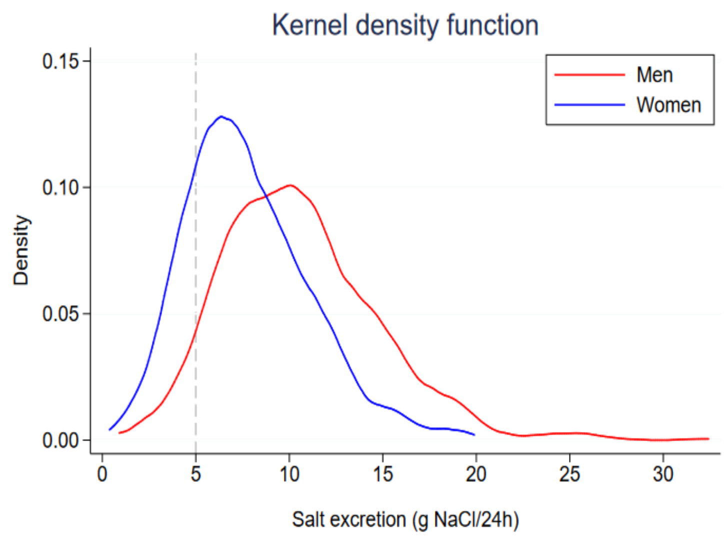


Figure 26 Kernel Density Function (Chappuis et al. 2011b)

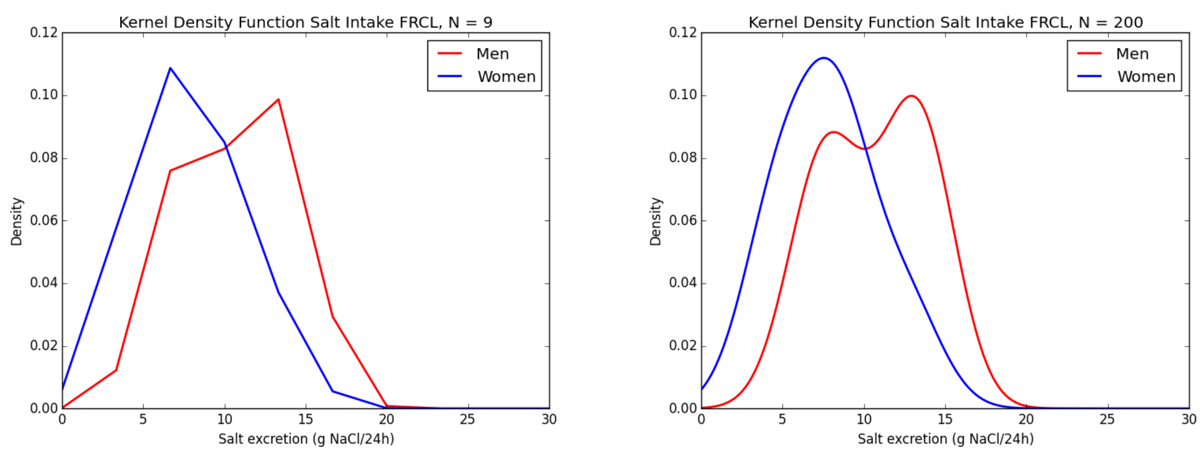


Figure 27 Kernel Density Function FRCL
On the left using only the original 9 users and on the right with 200 users gained using bootstrapping (Tan et al. 2006)

Another possibility to explore the salt intake distribution are density functions. Figure 26 shows the density function for men and women from a survey that took place between 2010 and 2011 using 24-hour urine collection from people throughout entire Switzerland with 1448 participants (Chappuis et al. 2011b).

Figure 27 depicts the density function using the users from the thesis study (graph on the left) and a smoothed graph (graph on the right) with 200 data points generated from the original sample using bootstrapping (Tan et al. 2006). Despite the very small sample size, the density function in both figures are relatively similar. Men have a higher salt intake than women and a relatively high density between 10 and 15 g. A difference however is that there are no extremely high salt intake values of more than 20 or even up to 30 g per day reported in the study, whereas this is the case in the 2010-2011 survey. The comparison indicates that, although very small, the sample in this thesis reflects the real Swiss population relatively well when it comes to the distribution of the average daily salt intake.

Salt Intake Categories

The collected data allows to analyze which categories had the strongest impact on the salt intake of the participating users. As already mentioned, there are two completely separate sets of categories used in this thesis, the categories defined in the FRCL and the categories defined by the “Schweizerische Gesellschaft für Ernährung” (SGE) and used in the Product-Item-Ingredient-Database (PIID).

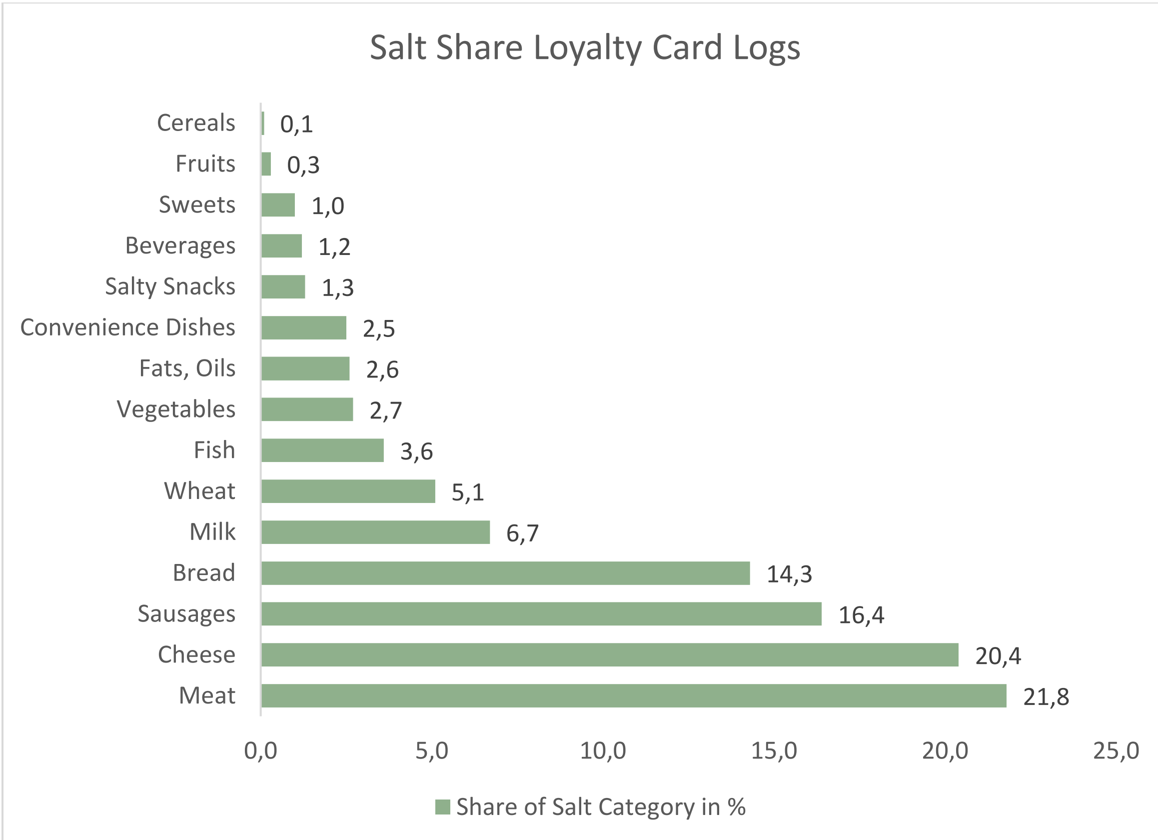


Figure 28 Salt Intake Share Loyalty Card Logs

Figure 28 shows the share of salt per category of the loyalty card logs as an average over all participating users. Compared to the top-five, although not identical, categories identified in previous studies (Beer-Borst et al. 2004; Chappuis et al. 2011b; Constanza et al. 2004) it stands out that meat or meat products, cheese and bread are main contributors to a person's salt intake, appearing both in the data collected and in the studies. Soups and ready meals, in the SGE categories both belonging to "convenience dishes", were also identified to be main contributions in earlier studies but have only a negligible impact in the loyalty card logs. This indicates that the salt intake of the category "convenience dishes" is underrepresented in the loyalty card logs.

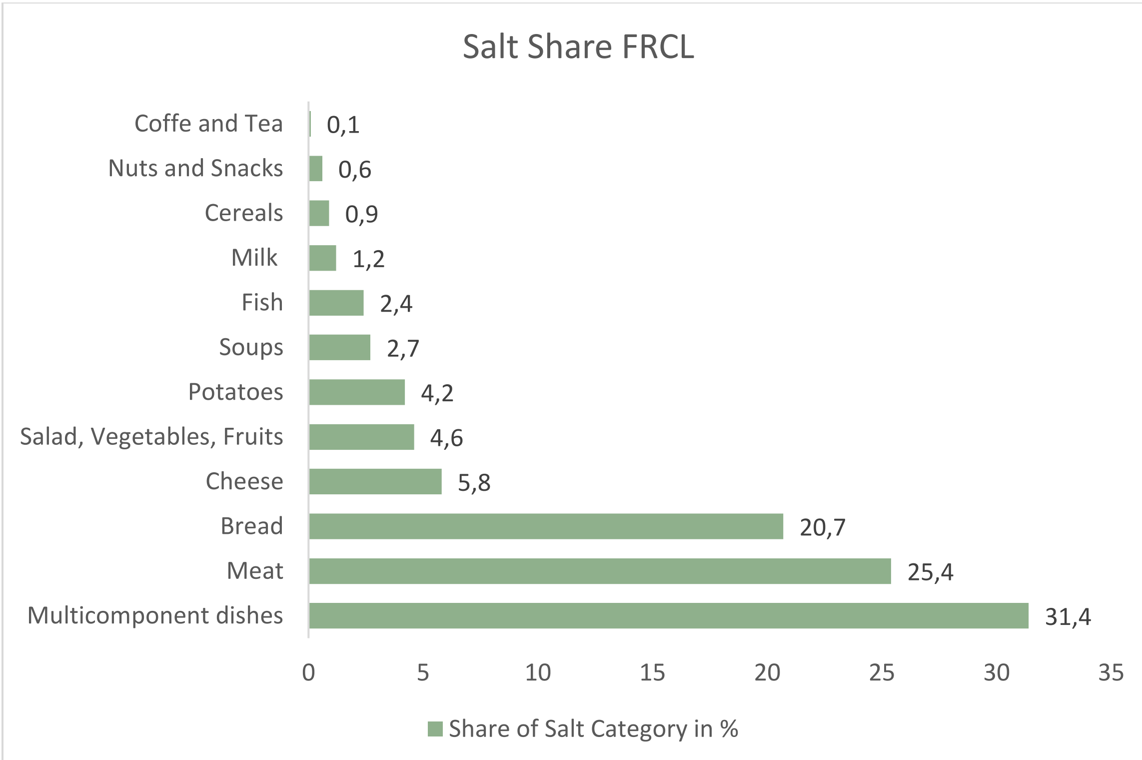


Figure 29 Salt Intake Share FRCL

In Figure 29 the salt share of the categories of the FRCL is illustrated, again as average from all users participating in the study. Cheese, bread and meat are again in the top-five categories, the absolute values however vary greatly compared to the loyalty card log data. The category with the highest share in the FRCL data is "multicomponent dishes". Food items belonging to this category are composed of several components, examples are pizza, pasta dishes, gratins and similar items. It is therefore no surprise that this category accounts for over 30% of the total salt intake, as many meals either cooked at home or consumed out-of-home consist of several, not clearly separable components. Soups however are again responsible for less salt intake than the previous studies showed. A possible explanation is the seasonality, as the study was conducted in late summer, whereas soups are usually consumed in the colder seasons. For better comparison, the five categories with the largest impact on the salt intake of the Swiss population (data from Beer-Borst et al. 2004), the loyalty card logs and the digital FRCL are illustrated in Table 6.

Rank	Swiss Population	Loyalty Card Logs	FRCL
1	Bread	Meat	Multicomponent Dishes
2	Cheese	Cheese	Meat
3	Meat products	Sausages	Bread
4	Soups	Bread	Cheese
5	Ready meals	Milk	Salad, Vegetables, Fruits

Table 6 Comparison Categories Salt Impact

Loyalty Card Logs

For each of the participating users, the loyalty card logs were collected. In Table 7 some metrics calculated from the logs are shown. On average, a user purchased 1197 different products from the retailer, e.g. 1197 unique products. It is important to note that this number also includes the non-food items, as it is not possible to distinguish food and non-food items after only importing the data. Of course, only food items have been matched (from the receipt string to the real product name), in average 364 per user. As new food items are added frequently and not every food item has been matched, this number does not include all food items contained in the total purchased items. Also, the number of matched items per user varies greatly and depends heavily on which products were matched (this is reflected in the high standard deviation). During the matching process, some non-food items have already been labeled "irrelevant", 142 per user in average. Ideally (and also relevant for future studies), the number of matched items plus the number of irrelevant items should equal the total purchased items, meaning that every purchased product of a user is clearly classified as food or non-food product. Due to limited time and resources, not all items could be matched during the thesis project. However, items with a high salt amount or items which have been purchased a lot were matched first.

An interesting property of the data reveals the share of the matched items mapped to a product from the PIID and to the share mapped to the manually added items. Only 22% of the products could be found in the PIID and 78% had to be manually added. This underlines the importance and necessity of the manual adding of products. Fortunately, it can be expected that the number of products in the PIID used an in PIIDs in general will increase strongly over the next years due to regulations (as for example the EU 1169)

Loyalty Card Logs Metric	Mean	Standard Deviation
Unique Purchased Items (incl. non-food)	1197	388
Matched Items	364	235
Irrelevant Items	142	109
PIID Share	22%	2%
Manually Added Items Share	78%	4%

Table 7 Loyalty Card Logs Metrics 1

In Table 7 the metrics are calculated for the unique products purchased. Table 8 shows how many products have been matched when looking at all products purchased by a user, e.g. also taking into account in which quantity an item was purchased. On average, 3954 items have been purchased by a user and, as the standard deviation indicates, the range between the users is rather large (ranging from 852 to 6300). More important however is the share of matched or declared as irrelevant items. As frequently purchased items have been matched with priority, the share of matched items out of all purchased items is 64% on average. This means that by matching roughly 30% of the unique products (364 of 1197) 64% of all purchased items are covered.

Loyalty Card Logs Metric	Mean	Standard Deviation
Total Purchased Items	3954	1999
Matched or Irrelevant Items in %	64	20

Table 8 Loyalty Card Logs Metric 2

5.2 Linear Regression Models

The linear regression models described in section 3.4.5 were applied to the user sample. Unfortunately, the user sample did prove to be insufficient for the proposed models. Especially the small numbers of users in the sample were the key factor which prevented to find a correlation between the loyalty card log and the salt intake reported in the mobile app. The described models were also modified and tested with different subsets of the major and minor categories and even custom categories were created. The custom categories consisted mostly of the major categories with several mutations. The major category “meat and entrails” and “meat and sausages” were reorganized into “meat” and “sausages”, as sausages are known to be more salt-rich compared to other meat. Also, the category “milk and dairy products” was spitted into “milk” and “cheese”. The last mutation concerned the category “bread and cereals” which was divided into “bread” and “cereals”. Again, cheese and bread are known to contain more salt per 100 grams compared to other dairy products and cereals respectively. The separations should help to group salt-rich products in separate categories. Despite the changes, it was still not possible to create a regression model which could be considered reasonable (some models even indicated that the more food was consumed out of salt-rich categories the lower the salt intake would be). Instead of the linear regression it was decided to perform only the classification.

5.3 Classification

The small sample size does affect the classification and its expressiveness. Although meaningful results can be calculated using decision trees, these results must be treated with caution. Such a small sample size cannot be considered representative and neither the results based on this sample. Further, the results will most certainly suffer from overfitting, which means that the results of the model application become relatively complex or, in the case of decision trees, very deep (Tan et al. 2006). In other words, too many model variables are incorporated relative to the sample size and relationships are found in the data that are more likely random errors. A commonly used technique to prevent overfitting is cross-validation, where the data is split into a training set and a test set (Tan et al. 2006) and the model is first learned on the training set only. Given the already small sample size, a split further reducing the size is not expected to be helpful. Another option to mitigate the effects of overfitting is called pruning (Tan et al. 2006) which poses further restriction on the decision tree generation by for instance setting minimal information gain boundaries before splitting the tree. As pruning is not as dependent on the sample size compared to cross-validation, the method is applied in this thesis.

The classification can yield early stage findings on which further analysis, using a larger sample size, can build upon. The results presented in the following, therefore, focus on whether it is possible to find a correlation in the first place. Due to the small sample size, the subsamples of users described in 3.2.2 were not analyzed in detail but all models were tested with all participating users. Nevertheless, in later studies and given a larger sample size, the models should also be tested with the proposed subsamples as it can be expected that the subsamples will show stronger correlation between loyalty card logs and the self-reported salt intake. This also allows to find beneficial properties of a user (e.g. few meals consumed out-of-home or high share of grocery products from the retailer) regarding the correlation.

Based on the salt intake risk assessment in 2.3, three user categories are formed. Users with a salt intake of <=8 g are in the category "Low", users with a salt intake of >8 g and <12 g form the category "Medium" and users with a salt intake >=12g belong to the category "High".

Although not beneficial for the linear regression, the custom categories formed for the analysis are also used for the classification. Models 1, 3, 5 and 6, introduced in section 3.4.5, are also used with the newly created custom categories. These models were all based on the major SGE categories out of which the custom categories have been derived. The classification is therefore conducted with a total of ten models. Besides the custom categories, the gender is also used as a model variable as the analysis of the user data in 5.1 has shown that the gender has an effect on the salt intake (in the Swiss population as well as in the user sample). In the following, not all results of the models applied to the gathered data will be analyzed in detail and only the most promising ones are described. To assess the quality and the performance of the resulting decision trees the following metrics are used:

- Accuracy

  The accuracy simply measures how many objects have been categorized correctly, e.g. the number of correct categorizations divided by the total number of objects. (Tan et al. 2006)

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

- Precision
  The precision in the context of classification describes how many objects classified to a class also actually belong to this class. It is therefore a measure of how precisely objects are classified. For example, if the classifier predicted four users to belong to the category “Medium”, however in the data set there would only be three users in this category, the precision for this class would be 75%. The total precision is the average of each class’ precision. (Brownlee 2014)

$$Precision = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Positives}$$

- Recall
  Recall measures how many objects of a class have also been classified into this class, e.g. how many objects for a class have been “found”. The recall is therefore a measure of completeness. For instance, if there are 4 objects in a class but the classifier only classified 3 objects into this class the recall would be 75%. (Brownlee 2014)

$$Recall = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Negatives}$$

- F1-Score
  The F1-Measure or F1-Score incorporate both precision and recall into one metric.

$$F1 - Score = 2 * \frac{precision * recall}{precision + recall}$$

For each discussed result of a model application, the just defined performance metrics as well as the resulting decision tree will be listed.

Model 3 using custom categories

The first result discussed is the application of Model 3, which is based on the weight of the products purchased in a category. Figure 30 shows the resulting decision tree and Table 9 the corresponding performance metrics.

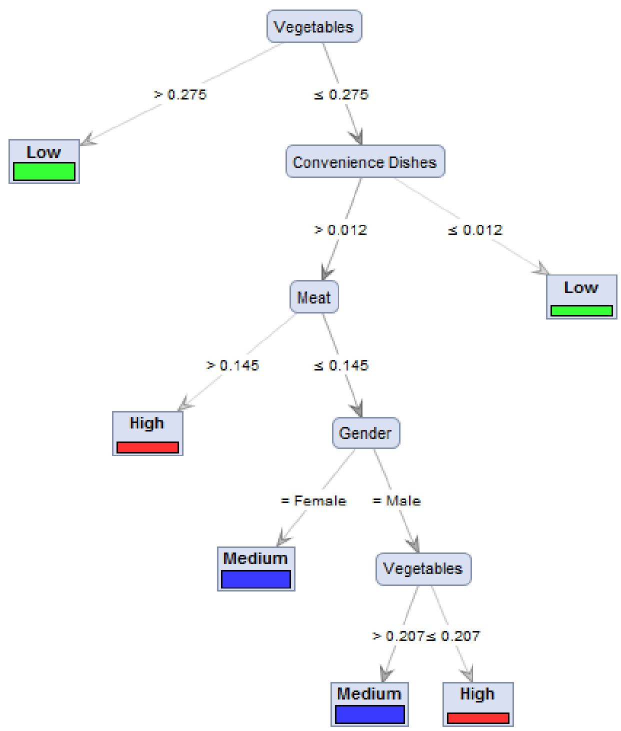


Figure 30 Resulting Decision Tree of Model 3

Model 3 Custom Categories	
Accuracy	100%
Precision	100% (1)
Recall	100% (1)
F1-Score	1

Table 9 Performance Metrics Model 3

The performance metrics of Model 3 all have the highest possible values as every single object was classified correctly. However, it is obvious that this result clearly suffers from overfitting (no pruning was applied) and the decision tree is rather complex (given the small sample). Nevertheless, the decision tree still provides some interesting information. It indicates that users buying more vegetables relative to others have a lower salt intake. As vegetables are known to have a rather small impact on a person's salt intake, this seems a valid statement. Users who buy more vegetables but also few convenience dishes are also correctly predicted to belong to category "Low". Users with many convenience dishes and meat in their loyalty card logs are classified as "High" which also makes sense. The last path in the tree which uses the variables "Gender" and again "Vegetables" is also coherent with real world findings but could also be a result of overfitting.

Model 4

Model 4, like Model 3, makes use of the product and category weights but uses the more differentiated SGE minor categories. The decision tree and the performance metrics are depicted in Figure 31 and Table 10 respectively.

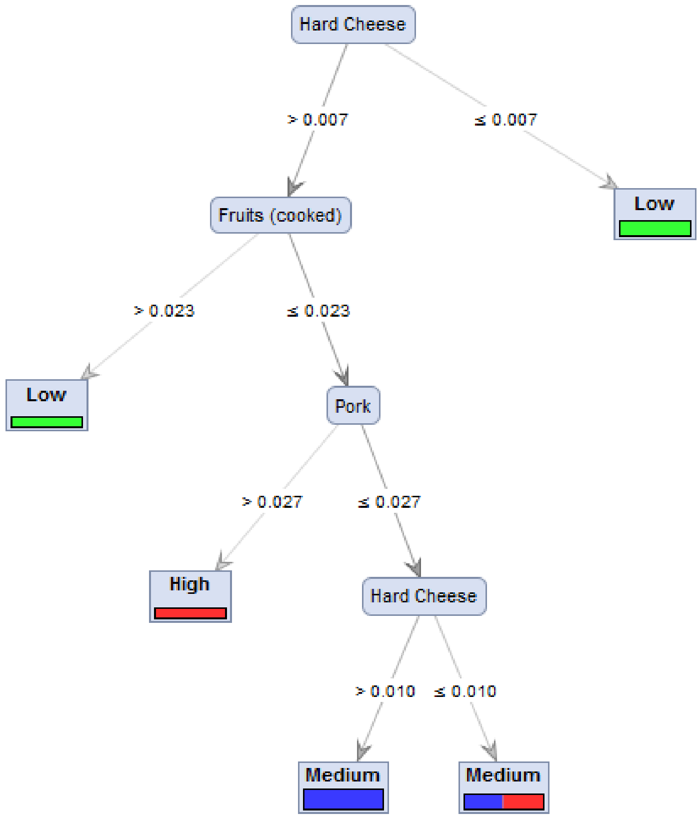


Figure 31 Resulting Decision Tree of Model 4

Model 4 Minor Categories	
Accuracy	88.9%
Precision	93.3% (0.933)
Recall	83.3% (0.833)
F1-Score	0.88

Table 10 Performance Metrics Model 4

In Model 4 the classifier has a much wider range of variables to choose from, potentially too much for the small sample size. The result however, although again suffering from overfitting and without using pruning, seems reasonable and the performance metrics are high. “Hard cheese”, belonging to the major category “Cheese” is a salt-rich category, fruits contain like vegetables relatively few salt per 100 g. The reasoning for classifying users to the category “Low”, namely a user buys not much hard cheese or if he does also buys a lot of fruits, seems valid. The same applied to the category “High” (purchases much hard cheese and pork but little fruits). The false predictions come from the category “Medium” with a class precision of only 80% while “Low” and “High” retain a precision of 100%. The recall however (83.3%) shows that not all users belonging to the category “High” (class recall of 50%) could have been found.

Model 5 using custom categories (without pruning)

In contrast to Model 3 and 4 not the total weight of the products but the amount of salt is relevant in Model 5 and the custom defined categories are used. Model 5 measures therefore which categories contribute the most to a user’s salt found in the loyalty card logs.

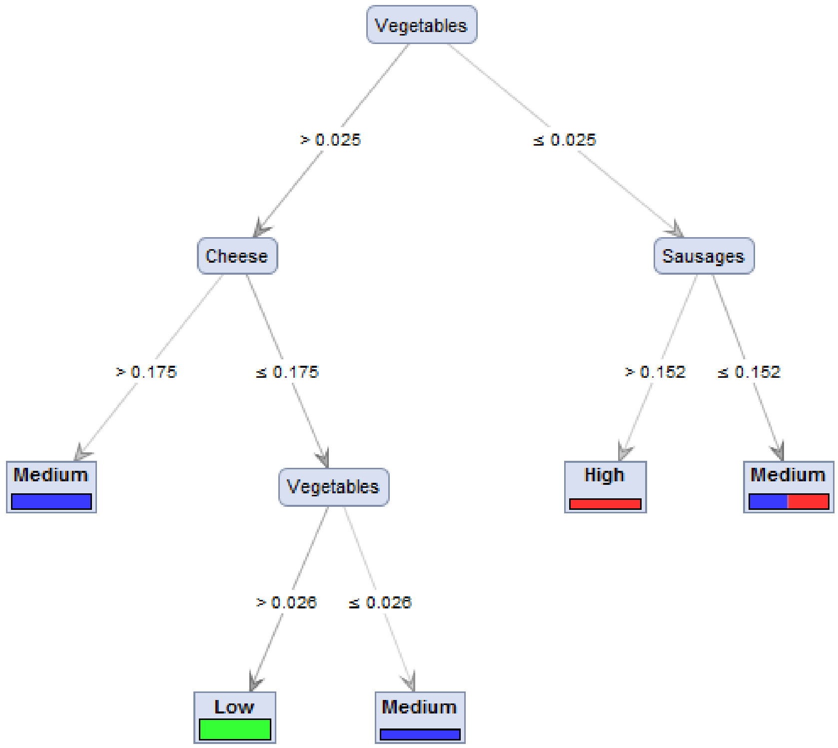


Figure 32 Resulting Decision Tree of Model 5 (no pruning)

Model 5 Custom Categories (no pruning)	
Accuracy	88.9%
Precision	93.3% (0.933)
Recall	83.3% (0.833)
F1-Score	0.88

Table 11 Performance Metrics Model 5 (no pruning)

The performance metrics (Table 11) are exactly the same as for Model 4, the decision tree (Figure 32 Resulting Decision Tree of Model 5 (no pruning)) however looks slightly less complex. Like in the results before the variables “Vegetable” (Model 1) and “Cheese” (“Hard Cheese” Model 4) are used for classifying the

“Low”-types. For the category “High” again the properties of purchasing few vegetables and relative more sausages (in previous models “pork” or “meat” was used) are utilized as split criteria for the decision tree. The prediction errors come again from the “Medium” (class precision of 80%) category for precision and accuracy and “High” for recall (class recall of 50%).

Model 5 using custom categories (without pruning)

The last result presented is again based on Model 5 with custom categories. The difference lies in the model application itself as pruning is used to reduce overfitting. The reason why the previously presented results did not made use of pruning are the relatively poor results such model applications yielded. The results for Model 5 with pruning are shown in Figure 33 and Table 12.

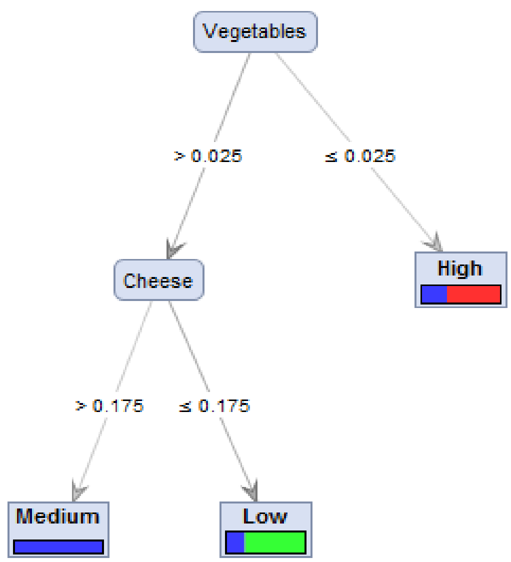


Figure 33 Resulting Decision Tree of Model 5 (with pruning)

Model 5 Custom Categories (with pruning)	
Accuracy	77.8%
Precision	80.6% (0.806)
Recall	83.3% (0.833)
F1-Measure	0.82

Table 12 Performance Metrics Model 5 (with pruning)

The metrics are the weakest presented regarding accuracy, precision and F1-Score. However, given that it is the only model application using pruning and the resulting decision tree is considerably simpler compared to the previous presented trees, it is not necessary the one with the lowest real world use. As only two variables are taken into account, the danger of overfitting is reduced and still almost 80% of the objects could be classified correctly. Again, the previously used variables "Vegetables" and "Cheese" were included in the computation, however the source of errors was not the category "Medium" but the categories "Low" (class prediction of 75%) and "High" (class prediction of 66.67 %) (regarding accuracy and precision). Looking at the recall the picture is different and the category "Medium" only has a 50% recall against the 100% class recall for "High" and "Low".

5.4 Research Questions

The description of the classification results implies that especially vegetables, cheese and meat related food items including sausages and probably also fruits and convenience dishes are good predictors for the salt intake of a person given his loyalty card logs. Vegetables and fruits could be used to identify people with a low dietary salt intake. To be precise, a person whose predominant salt intake category is "Vegetable" does not necessarily need to be a person with a low salt intake in general. It is possible that such a person simply consumes large amounts of vegetables and thus still has a high salt intake. However, the results found indicate that someone who either buys a lot of fruits or vegetables (Model 3 and 4) or has vegetables as a major contributor to his salt found in the loyalty card logs, is also expected to have low dietary salt intake. In contrast, meat, cheese and sausages which are known to be salt-rich, can be used to predict high salt intake. The category “Medium”, people with an average salt intake between 8 and 12 g, was found to be the most difficult to predict regarding the average class precision (of all presented models) which was 90% compared to 92% for “High” and 94% for “Low”.

The variables best suited for predicting the salt intake level can be categorized into two groups. Variables belonging to the first group follow the logic that the more items of this category are purchased the higher the predicated salt intake will be. Examples are the categories “Meat”, “Cheese” or “Sausages”. The categories “Fruits” and especially “Vegetables” indicate the opposite and form the second group. The more items are purchased belonging to these categories, the lower the predicted salt intake. The two groups indicate that it is not only important to focus on categories that are known to have a strong impact on a person’s salt intake but also on the categories that do not. Categories as “Vegetables” can likely be used to predict who has a sufficiently low salt intake and does not need further interventions, leaving more resources for people who do need consultation or coaching. The variable “Gender” does not belong to any of the groups but can also be helpful for prediction as women have in general a lower salt intake compared to men.

Table 13 summarizes which variables have been used by the decision trees while applying the models the data. Further, it is illustrated how many times a variable was used indicating the importance of a variable regarding future analysis and prediction models. Minor categories are counted to the respective major category (e.g. “Hard Cheese” to “Cheese”). Variables belonging to the same group are colored equally.

Variable (Food Category)	Occurrences in Results	Model(s)
Vegetables	3	3, 5 (no pruning), 5 (with pruning)
Cheese	3	4, 5 (no pruning), 5 (with pruning)
Meat	2	3, 4
Sausages	1	5 (no pruning)
Fruits	1	4
Gender	1	3
Convenience Dishes	1	3

Table 13 Occurrences of Variables in Tree Results

Although Research Question 1 cannot be answered with certainty, the analysis of the results indicates that a person's salt intake risk level can be derived automatically from his loyalty card logs using the models developed during this thesis under the requirement of a larger sample size.

Research Question 2 could not be answered as the sample was not sufficient to conduct the linear regressions and must be left to future work.

CHAPTER 6. DISCUSSION

The last chapter of this thesis consists of the conclusion as a summary of the findings. It is also explained what the limitations of the thesis are and how future work could extend the thesis and even make use of the technical solutions developed during the thesis project.

6.1 Conclusion

This thesis performed the steps necessary to automatically link dietary salt intake to loyalty card logs. A technical framework for collecting loyalty card logs and automatic linkage to the self-reported dietary salt intake has been developed and the paper-based Food-Record-Checklist (FRCL) for salt intake measurement has been translated into a digital context. Further, first proposals were made which statistical models and food categories are suited to find a correlation between the reported salt intake and the loyalty card logs. Due to the limited sample size, a correlation cannot yet be completely confirmed, although the first analysis using the data collected and the models developed are promising. Especially the variables found to be helpful to predict a person's salt intake risk level can be reused in further studies and future work.

Special care was taken to create all resources in this thesis in such a way that they can be reused in future work and research. The background services run autonomously and further import new products from the Product-Item-Ingredient-Database (PIID), the functionality for exporting the latest data for every model presented in this thesis has been included into a web service and the data can therefore be downloaded by authorized researches with access to the internet.

To emphasize that this thesis aims to act as a starting point and supporter of future research in the field of analyzing and linking loyalty card logs or digital receipts to self-reported dietary intake, the main barriers found in this thesis together with potential solutions are specified to conclude this thesis.

Integration of PIID

To have a solid data foundation is a crucial part in order to map items in the loyalty card logs to real products. Therefore, a PIID or a combination of several PIIDs should be carefully selected and the data quality must be assessed. Depending on the format and quality of the data further actions to correct or harmonize the data must be taken.

User acquisition and loyalty card credentials

User acquisition should be started as soon as possible, the effort for a user to obtain its loyalty card credentials must not be underestimated and if there is no public API to obtain the loyalty card logs from a retailer, additional time must be conceded to develop a solution for importing the logs.

Matching and manual adding of items

The matching is an integral part of the automatic linkage. Further processing is only possible if the product in the loyalty card logs can be identified. The matching would be rendered useless if the product names in the loyalty card logs are the same as in the PIID but for many retailers this will not be the case. Besides the matching, the manual addition of products missing in the PIID is important, as the mapping analysis in section 5.1 has shown. In the future, this could become unnecessary as more and more food items will be included in PIIDs due to regulations like the EU 1169.

The generated models and the results of the application of the models on the data collected are promising. As already stated, the small sample size does not allow to completely accept the first research question stated in 3.4.3. Nevertheless, the results indicate that using the respective loyalty card logs, it is possible to predict the risk level of a person regarding his salt intake automatically, rendering the self-reporting of meals and food items unnecessary.

6.2 Limitation

The small and not representative sample size does pose a limitation to the presented results. In order to be representative, a larger number of users would be needed. Nevertheless, the statements made in this thesis still holds, for a further and more detailed analysis a larger sample size is a key component. The presented results are therefore also not representative and can only show which factors are potentially helpful to verify a correlation between loyalty card logs and self-reported salt intake.

The self-reported salt intake is used as a person's "true" salt intake in this thesis. The self-reported salt intake is however only an approximation of the real salt intake or even an approximation of an approximation. The paper-based FRCL is an approximation of the 24-hours-urine-samples, which can accurately measure the real salt intake value. The digital FRCL in the mobile application used for the thesis experiment is an approximation of the paper-based FRCL. As the paper-based FRCL does correlate with the urine-samples and the digital FRCL is a translation of the paper-based FRCL, the self-reported salt intake in the thesis study is also expected to correlate with the urine-samples and subsequently with the true salt intake value. Nevertheless, future work must include further studies which examine the correlation between the paper-based and app version of the FRCL. A clinical study directly comparing the FRCL in the mobile app with 24-hours-urine-samples should also be conducted and, besides the analysis of the correlation between these two, the results should be compared to the existing studies examining paper-based FRCL and urine-samples.

A further restriction is the focus on loyalty card logs from only one retailer. Although a necessary restriction, (no other loyalty card logs can be obtained digitally in Switzerland at the time of writing this thesis) it limits possible users to customers of this specific retailer. Food purchased at other retailers and food that was consumed in restaurants and as take-away food or that was home grown is not reflected in the analysis. This problem cannot be solved by a further study itself but is dependent on how quickly other retailers and companies adopt digital receipts and provide APIs or other ways to access the loyalty card logs of a user.

The used loyalty card logs are not necessarily reflecting the purchases of an individual but of often the purchases of an entire household. Therefore, it is possible that products on the loyalty card logs are not evenly consumed by the users of the loyalty card. This bears the danger of over- or underestimating the impact of certain products towards a person's salt intake. It can be expected that the assignment of purchased products to an individual will become easier in the future due to the broader adoption of digital receipts from retailers and restaurants. In the short term, the correlation of the loyalty card logs and the self-reported salt intake can be analyzed using only small households, e.g. using a subsample as described in 3.2.2.

The described matching process makes clear that not all items purchased by the participating users were also matched to the real product. This means that the salt values of the loyalty card logs are not completely accurate as not all purchases were incorporated. By extending the matching table (and also the missing items table if necessary) this limitation can be resolved.

The data quality in the PIID is a further limitation. Although the product information is directly inserted by the respective manufacturer faulty salt or sodium entries cannot be completely ruled out.

As the mobile app was only launched in Switzerland, all statements are only valid for people living in Switzerland. The eating and shopping behavior in other countries may differ.

6.3 Future Work

Based on the experiment setup, the technical foundation and the presented results, further studies have to be conducted using a larger sample size. Provided that a larger sample size can be found, it is possible to extract more information from the data collected. Reapplying the models on a larger dataset and analyzing the results should also help to make the models more robust or even find new models being better suited. Research for identifying a person's salt intake level and help to reduce the salt intake if necessary is highly relevant and can be a contributor to help prevent deaths and diseases caused by increased or excessive salt intake.[19]

The application of the linear regression models could not be conducted. The method itself is however still expected to have the potential to find a correlation between the loyalty card logs and the self-reported salt level. In contrast to the classification, the linear regression could even be able to predict the salt intake value instead of only the risk level. Given a significantly larger sample size, which will be necessary to apply the linear regression successfully, the application of the regression models presented in 3.4.5 should be conducted in a future study.

If the correlation between the loyalty card logs and the salt intake measured by the FRCL can be proven to be strong enough, it is possible to completely remove the process of self-reporting. This leads to applications which allow for analyzing a user's salt intake in a very convenient way, reducing the user's effort to a minimum. A possible point of contact to build upon could be the "shopping tips" described in section 4.4.3. This a promising part of this thesis and should be further

[19] http://www.who.int/mediacentre/factsheets/fs393/en/

refined and developed in future work. An extensive analysis of the purchase history of a user could allow to give very personal advice to use low-salt products, even incorporating the user's personal taste. As most parts of the thesis, this can not only be applied to salt but to all nutrition ingredients as fat or carbohydrate. A crucial part would probably be immediate or more contemporary feedback. In the setup developed for this thesis it can take up to 48 hours until new purchases are reflected in the assessment, for more coaching orientated applications it is important to drastically shorten this response time. A possible solution would be to execute the periodically executed tasks much more frequently and no longer separate the import of new purchases and the calculation of the assessment into different tasks.

The collected data from the loyalty logs can also be used to perform impact analysis. The goal of such an analysis is to identify products purchased by the user whose substitution with a product containing lower salt values would have the biggest impact, e.g. lowering the salt intake.

When diving further into interventions aiming to change the eating behavior of people, the field of Persuasive Technology is addressed. Persuasive Technology however also has problems and limitations, for example the focus on measurable effects or that they often assume that people act and choose rationally (Huber and Hilty 2015). A potential extension or alternative presents gamification, which means that elements usually used in games and game design are applied to "non-game contexts" (Huber and Hilty 2015). Although Huber and Hilty (2015) were researching Persuasive Technology and gamification in the context of sustainable consumption, these concepts can also be applied to health and nutrition related topics. Both, sustainable behavior and reduction of salt intake, aim to change a person's conduct towards a "better" (e.g. more sustainable or healthier) state. Gamification can lead to higher user engagement as it can help to raise and keep up motivation through the use of game design concepts as cooperation and competition elements or letting the user feel the notion of progress and improvement of his skills.

Further future work could include the improvement of the matching process, which is manually performed in this thesis. It would be necessary to compare several string similarity measures and find the one which fits best the use case of mapping strings from receipts to real product names. Machine learning could be used to identify terms which support or hinder the matching, e.g. the brand name can be helpful if one wants to find the exact branded product but also dangerous as it can have a too big weight when using string similarity measures. It could also be beneficial to use additional data sources where the same products are stored with other or more information that might support the matching.

Most importantly, the introduced concepts and the technical foundation built in this thesis can be leveraged to not only support further analysis and recommendations in relation to salt or sodium but potentially for any nutrition ingredient. The PIID already provides this data and also for the manually added missing items the values for energy, fat, saturated fat, carbohydrate, sugar, fibers and protein are stored. Naturally, for every nutrition ingredient different research must be taken into account and conflict of aims must be resolved, for example when reducing one ingredient would lead to an increase of another. However, this thesis provides a foundation on which it is possible to build such systems and come closer towards a solution to fully analyze a person's nutrition and give behavioral recommendations by using loyalty card logs. Even more so as it can be expected that usage and general availability of digital receipts are going to grow drastically in the coming years, eliminating some limitations of the current research performed in this thesis, namely the focus on loyalty card logs from only one retailer.

CHAPTER 7. APPENDIX

7.1 User Survey Questionnaire

Upon registration, each user was asked the following questions within the mandatory user survey (translated from German):

1. *How often do you do sports per week?*

This question is asked to possibly find out whether the frequency of physical activity has an effect on salt intake.

2. *Which nutrition profile does describe you best?*

Asked to find out if the nutrition profile (e.g. vegan, vegetarian etc.) has any influence on salt intake.

3. *How many adults are living with you?*
4. *How many children are living with you?*

As many households and families are sharing a loyalty card it is important to know how many adults and children live in the same household and food is bought for. The distinction between adults and children is based on the assumption that children eat a smaller share of the purchased food than adults.

5. *How often do you (your household) buy groceries per week?*

6. *What is the share of groceries bought at Migros?*

As it is only possible to obtain the data of products bought at Migros, it is important to know how often a person buys food at Migros. People with a low share will most likely have to be excluded from further analyses.

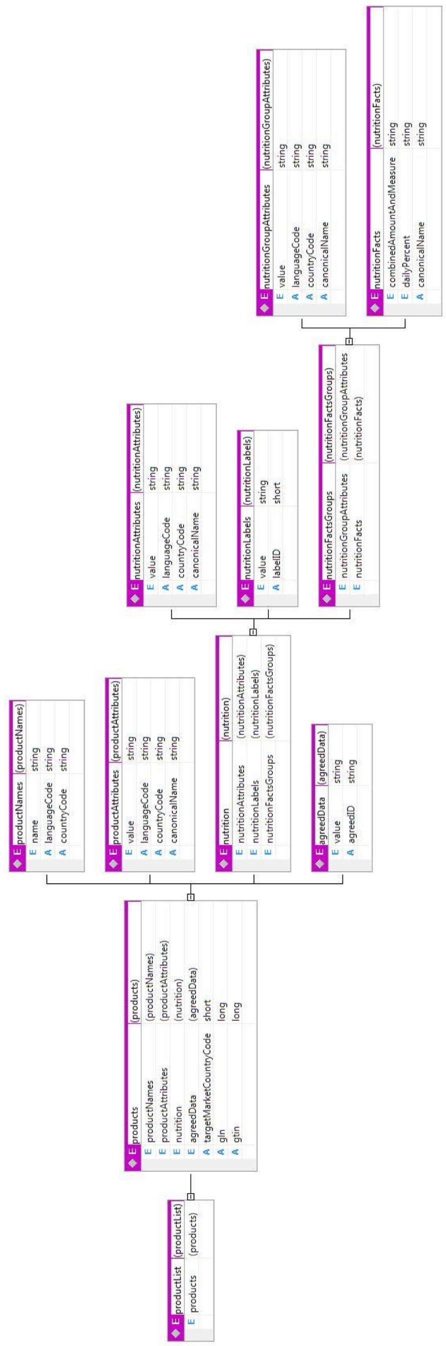
7. *How often do you use your Cumulus-Card when you shop at Migros?*

This question is a follow-up to question (6) as only purchases completed by scanning the Cumulus card are stored by Migros and can later be retrieved.

8. *How many of the main meals do you eat out-of-home per week?*

This question tries to collect information about where a person eats. For people who seldom eat at home the fraction of the salt intake that can be found in the loyalty card logs is only a small share of their total salt intake.

7.2 Trustbox Database Schema



7.3 EatFitService Database Schema

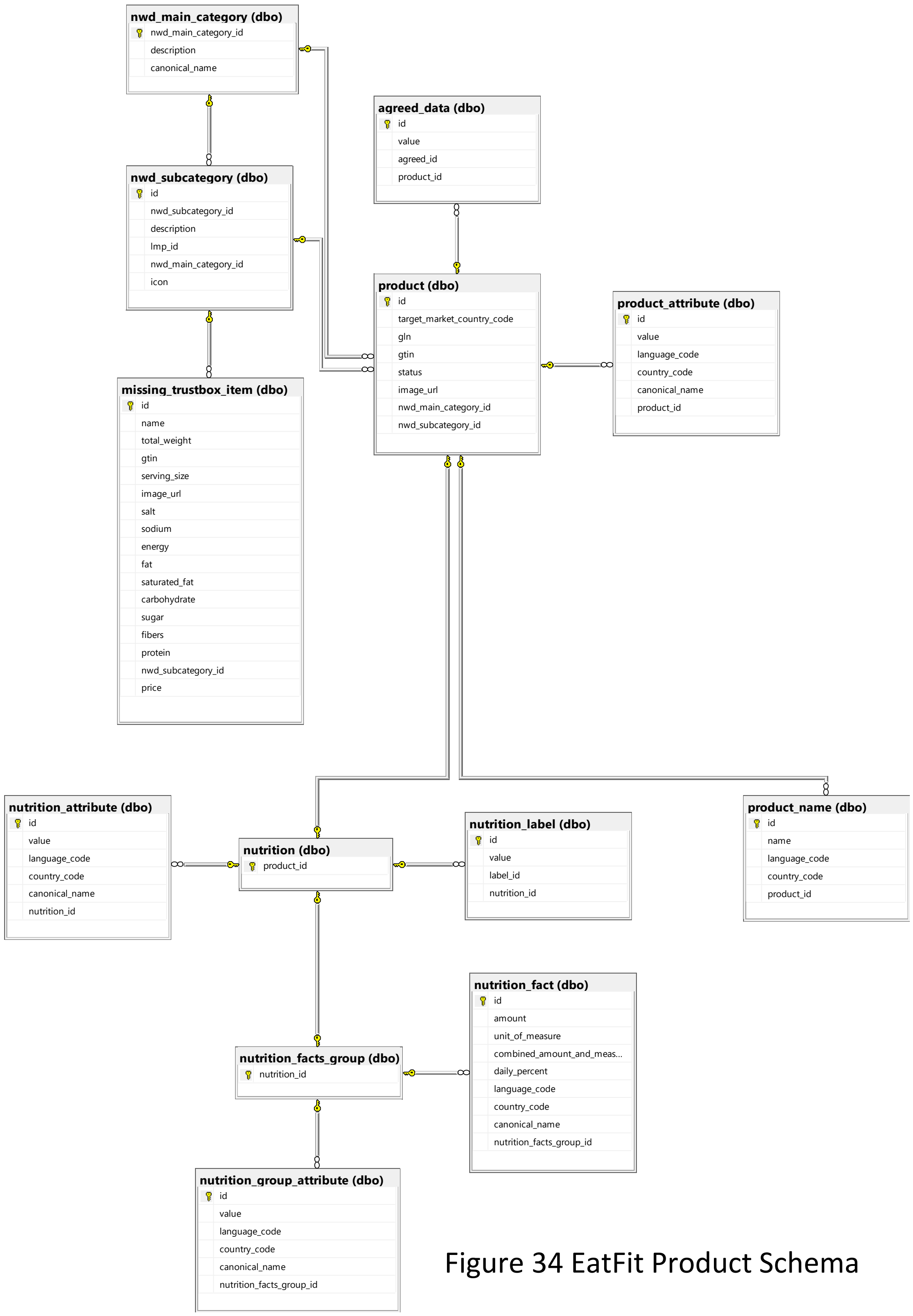


Figure 34 EatFit Product Schema

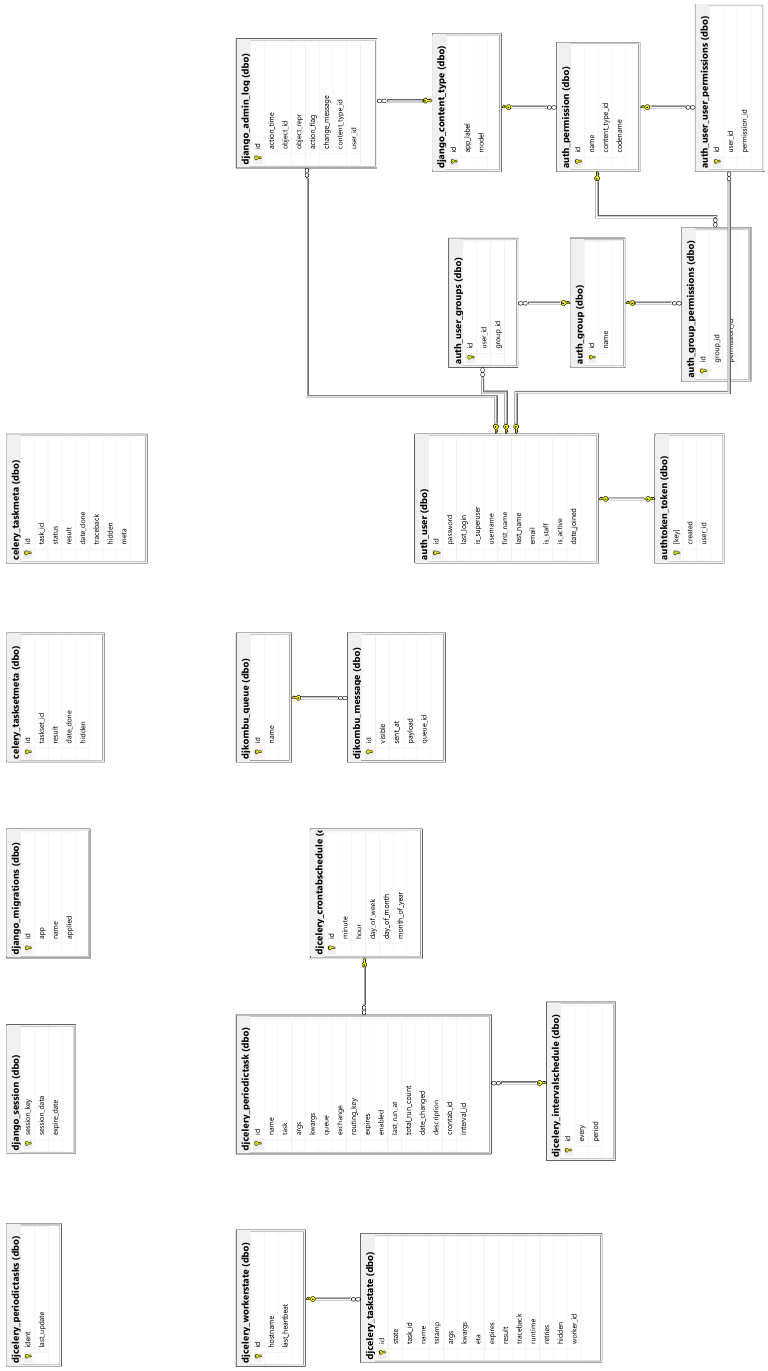


Figure 35 EatFit Application Schema

7.4 Content of CD

The attached CD contains the following:

- Thesis (PDF-format)
- Abstract as text file in English and German
- Source Code structured into three separate projects
    - SaltTrackerApp
    - SaltTrackerService
    - EatFitService

7.5 Bibliography

Aburto, Nancy J. et al. 2013. “Effect of Lower Sodium Intake on Health: Systematic Review and Meta-Analyses.” *BMJ (Clinical Research Ed.)* 346(April):f1326.

Bachmann, M. et al. 2004. “Salzkonsum Und Bluthochdruck: Expertenbericht Der Eidgenössischen Ernährungskommission.” *Bundesamt Für Gesundheit BAG* (0):1–62.

Beer-Borst, S., M. C. Costanza, A. Pechère-Bertschi, and A. Morabia. 2009. “Twelve-Year Trends and Correlates of Dietary Salt Intakes for the General Adult Population of Geneva, Switzerland.” *European Journal of Clinical Nutrition* 63(2):155–64.

Beer-Borst, Sigrid, Michael Costanza, and Alfredo Morabia. 2004. “10-Year Trends and Correlates of Dietary Salt in the Geneva General Adult Population. Report - Part 1 on Behalf of the Federal Office of Public Health.” *Geneva: University Hospitals Geneva* (April):1–30.

Beer-Borst, Sigrid, Corinna Krause, Stefan Siegenthaler, and Stefanie Hayoz. 2015. “Measuring the Effectiveness of Salt Intake Intervention . Can a Sodium and Potassium Specific Food Record Checklist Complement or Replace 24-Hour and Spot-Urine Collections?” P. 3012 in *ICDAM9 2015 (poster)*, vol. 41.

Beer, Michael. 2009. *Salz Strategie 2008-2012 Direktionsbereich Verbraucherschutz Actionsanté : Zentrale Aktionsbereiche*.

Block, Gladys et al. 2015. "A Fully Automated Diabetes Prevention Program, Alive-PD: Program Design and Randomized Controlled Trial Protocol." *JMIR Research Protocols* 4(1):e3–e3.

Brinkerhoff, Kristina M. et al. 2011. "Linking Supermarket Sales Data to Nutritional Information: An Informatics Feasibility Study." *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium* 2011:598–606.

Brown, Ian J. et al. 2013. "Estimating 24-Hour Urinary Sodium Excretion from Casual Urinary Sodium Concentrations in Western Populations." *American Journal of Epidemiology* 177(11):1180–92.

Brownlee, Jason Brownlee. 2014. "Classification Accuracy Is Not Enough: More Performance Measures You Can Use - Machine Learning Mastery." *Website*. Retrieved (http://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/).

Casperson, Shanon L. et al. 2015. "A Mobile Phone Food Record App to Digitally Capture Dietary Intake for Adolescents in a Free-Living Environment: Usability Study." *JMIR mHealth and uHealth* 3(1):e30.

Chappuis, A. et al. 2011a. "Studie Zum Salzkonsum Der Schweizer Bevölkerung." *Service de Néphrologie et Institut Universitaire de Médecine Sociale et Préventive Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland*.

Chappuis, A. et al. 2011b. "Swiss Survey on Salt Intake: Main Results." *Service de N´ephrologie et Institut Universitaire de M´edecine Sociale et Pr´eventive Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Suisse.* 1–32.

Chidambaram, Valliammai, Philip J. Brewster, Kristine C. Jordan, and John F. Hurdle. 2013. "qDIET: Toward an Automated, Self-Sustaining Knowledge Base to Facilitate Linking Point-of-Sale Grocery Items to Nutritional Content." *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium* 2013:224–33.

Cobiac, Linda J., Theo Vos, and J.Lennert Veerman. 2010. "Cost-Effectiveness of Interventions to Reduce Dietary Salt Intake." *Heart (British Cardiac Society)* 96(23):1920–25.

Coll, Sami. 2013. "Consumption as Biopower: Governing Bodies with Loyalty Cards." *Journal of Consumer Culture* 13(3):201–20.

Constanza, Michael C., Sigrid Beer-Borst, Hans Wolff, and Alfredo Morabia. 2004. "A Population-Based Assessment of Salty Foods and Blood Pressure. Report - Part 2 on Behalf of the Federal Office of Public Health." *Geneva: University Hospital Geneva*.

Drewnowski, A. and B. M. Popkin. 1997. "The Nutrition Transition: New Trends in the Global Diet." *Nutrition Reviews* 55(2):31–43.

European Commission. 2014. "E-Invoiving Directive 2014/55/EU." *Directive 2014/55/EU*.

European Commission. 2016a. "Directive 95/46/EC (General Data Protection Regulation)." 1–88.

European Commission. 2016b. "Protection of Personal Data."

European Commission. 2016c. "Recommendations on the Use of a Semantic Data Model to Support Interoperability for Electronic Invoicing."

Eyles, H., A. Rodgers, and C. Ni Mhurchu. 2010. "Use of Electronic Sales Data to Tailor Nutrition Education Resources for an Ethnically Diverse Population." *Journal of Human Nutrition and Dietetics* 23(1):38–47.

Fielding, Roy T. 2000. "Representational State Transfer (REST)." *Website*. Retrieved (http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm).

Forouzanfar, Mohammad H. et al. 2015. "Global, Regional, and National Comparative Risk Assessment of 79 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks in 188 Countries, 1990-2013: A Systematic Analysis for the Global Burden of Disease Study 2013." *The Lancet* 386(10010):2287–2323.

Fowler, Martin. 2006. "GUI Architectures." *Website*. Retrieved (http://martinfowler.com/eaaDev/uiArchs.html).

Franco, Veronica and Suzanne Oparil. 2006. "Salt Sensitivity, a Determinant of Blood Pressure, Cardiovascular Disease and Survival." *Journal of the American College of Nutrition* 25(3 Suppl):247S–255S.

Freeman, Erick and Elisabeth Freeman. 2013. *Head First Design Patterns*. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/23386138).

FSVO. 2013. “Salt Strategy for 2013 - 2016. Paper on a Strategy for Reducing Salt Consumption.” *FSVO* 1–20.

FSVO. 2016. “Federal Food Safety and Veterinary Office (FSVO). Swiss Food Composition Database (SFCD).”

Fuchs, Klaus Ludwig, Valerius Huonder, Denis Vuckovac, and Alexander Dr. Ilic. 2016. “Swiss FoodQuiz: Inducing Nutritional Knowledge via a Visual Learning Based Serious Game.” in *ECIS 2016 Proceedings (accepted)*.

Gossman, John. 2005. “Introduction to Model/View/ViewModel Pattern for Building WPF Apps.” *Website*. Retrieved (https://blogs.msdn.microsoft.com/johngossman/2005/10/08/introduction-to-modelviewviewmodel-pattern-for-building-wpf-apps/).

He, Feng J., Jiafu Li, and Graham A. Macgregor. 2013. “Effect of Longer Term Modest Salt Reduction on Blood Pressure: Cochrane Systematic Review and Meta-Analysis of Randomised Trials.” *BMJ (Clinical Research Ed.)* 346:f1325.

He, Feng J. and Graham A. MacGregor. 2003. “How Far Should Salt Intake Be Reduced?” *Hypertension* 42(6):1093–99.

He, Feng J. and Graham A. MacGregor. 2010. “Reducing Population Salt Intake Worldwide: From Evidence to Implementation.” *Progress in Cardiovascular Diseases* 52(5):363–82.

Hebden, L., A. Cook, H. P. Van Der Ploeg, and M. Allman-Farinelli. 2012. “Development of Smartphone Applications for Nutrition and Physical Activity Behavior Change.” *Journal of Medical Internet Research* 14.

Hicham, Gueddah, Yousif Abdallah, and Belkasmi Mustapha. 2012. “Introduction of the Weight Edition Errors in the Levenshtein Distance.” *IJARAI - International Journal of Advanced Research in Artificial Intelligence* 1(5):30–32.

Huber, Martina Z. and Lorenz M. Hilty. 2015. “ICT Innovations for Sustainability.” *Advances in Intelligent Systems and Computing* 310:367–85. Retrieved (http://www.scopus.com/inward/record.url?eid=2-s2.0-84927651454&partnerID=tZOtx3y1).

Illner, A. K. et al. 2012. “Review and Evaluation of Innovative Technologies for Measuring Diet in Nutritional Epidemiology.” *International Journal of Epidemiology* 41(4):1187–1203.

Ipjian, Michelle Lauren, Carol Johnston, Christina Shepard, and Melinda Johnson. 2016. “The Popular Calorie Counter App, MyFitnessPal, Used to Improve Dietary Sodium Intake: A Four-Week Randomized Parallel Trial - ProQuest.” *Arizona State University Digital Repository*.

Karpischek, Stephan, Florian Michahelles, and Elgar Fleisch. 2014. “Detecting Incorrect Product Names in Online Sources for Product Master Data.” *Electronic Markets* 24(2):151–60.

Kumanyika, S. 1991. “Behavioral Aspects of Intervention Strategies to Reduce Dietary Sodium.” *Hypertension (Dallas, Tex. : 1979)* 17(1 Suppl):I190-5.

LifeSum. 2016. “LifeSum.” *Website*. Retrieved (https://www.lifesum.com/).

Lose-it! 2016. “Lose-It!” *Website*. Retrieved (https://www.loseit.com/).

Moore, David S., George P. McCabe, Layth C. Alwan, Bruce A. Craig, and William M. Duckworth. 2011. *The Practice of Statistics for Business and Economics*. New York: W.H. Freeman and Company.

Murielle Bochud, MD PhD, MD PhD Pedro Marques-Vidal, MD Michel Burnier, and MD Msc Fred Paccaud. 2012. “Dietary Salt Intake and Cardiovascular Disease : Summarizing the Evidence.” *Public Health Reviews* 33(2):530–52.

MyFitnessPal. 2016. “MyFitnessPal Website.” *Website*. Retrieved (https://www.myfitnesspal.com/).

Närhinen, M., A. Nissinen, and P. Puska. 1998. “Sales Data of a Supermarket--a Tool for Monitoring Nutrition Interventions.” *Public Health Nutrition* 1:101–7.

O’Flaherty, M. et al. 2016. “Exploring Potential Mortality Reductions in 9 European Countries by Improving Diet and Lifestyle: A Modelling Approach.” *International Journal of Cardiology* 207:286–91.

Popkin, Barry M. 2002. “An Overview on the Nutrition Transition and Its Health

Implications: The Bellagio Meeting.” *Public Health Nutrition* 5(1a):93–103.

Pretzer, Matthias. 2003. “Clustering Und Klassifikation.” 1–40. Retrieved (http://www.diko-project.de/dokumente/ausarbeitungen/pretzer.pdf).

Rankin, Janet Walberg et al. 1998. “Food Purchase Patterns at the Supermarket and Their Relationship to Family Characteristics.” *Journal of Nutrition Education* 30(2):81–88.

Rokach, Lior and Oded Maimon. 2005. “Top-down Induction of Decision Trees Classifiers - A Survey.” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 35(4):476–87.

Smith, Josh. 2009. “Patterns - WPF Apps With The Model-View-ViewModel Design Pattern.” *Website*. Retrieved (https://msdn.microsoft.com/en-us/magazine/dd419663.aspx).

Sommer, Andreas and Stephan Krusche. 2013. “Evaluation of Cross-Platform Frameworks for Mobile Applications.” *Proceedings of the 1st European Workshop on Mobile Engineering* 363–76. Retrieved (http://subs.emis.de/LNI/Proceedings/Proceedings215/363.pdf).

Srinath Reddy, K. and Martijn B. Katan. 2004. “Diet, Nutrition and the Prevention of Hypertension and Cardiovascular Diseases.” *Public Health Nutrition* 7(1A):167–86.

Strazzullo, Pasquale, Lanfranco D’Elia, Ngianga-Bakwin Kandala, and Francesco P. Cappuccio. 2009. “Salt Intake, Stroke, and Cardiovascular Disease: Meta-Analysis of Prospective Studies.” *BMJ (Clinical Research Ed.)* 339:b4567.

Stumbo, Phyllis J. et al. 2010. “Web-Enabled and Improved Software Tools and Data Are Needed to Measure Nutrient Intakes and Physical Activity for Personalized Health Research.” *The Journal of Nutrition* 140(12):2104–15.

Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. 2006. “Classification : Basic Concepts , Decision Trees , and.” *Introduction to Data Mining* 67(17):145–205. Retrieved (http://www-users.cs.umn.edu/~kumar/dmbook/index.php).

Tate, D. F., R. R. Wing, and R. a Winett. 2001. “Using Internet Technology to Deliver a Behavioral Weight Loss Program.” *JAMA, Journal of the American Medical Association* 285:1172–77.

Thomas, J.Graham, Tricia M. Leahey, and Rena R. Wing. 2015. “An Automated Internet Behavioral Weight-Loss Program by Physician Referral: A Randomized Controlled Trial.” *Diabetes Care* 38(1):9–15.

Tin, Sandar Tin, Cliona Ni Mhurchu, and Chris Bullen. 2007. “Supermarket Sales Data: Feasibility and Applicability in Population Food and Nutrition Monitoring.” *Nutrition Reviews* 65(1):20–30.

Turner-McGrievy, Gabrielle and Deborah Tate. 2011. “Tweets, Apps, and Pods: Results of the 6-Month Mobile Pounds Off Digitally (Mobile POD) Randomized Weight-Loss Intervention among Adults.” *Journal of Medical Internet Research* 13(4):e120.

Weinberger, Myron H. 2002. “Salt Sensitivity Is Associated with an Increased Mortality in Both Normal and Hypertensive Humans.” *Journal of Clinical Hypertension (Greenwich, Conn.)* 4(4):274–76.

Wharton, Christopher M., Carol S. Johnston, Barbara K. Cunningham, and Danielle Sterner. 2014. “Dietary Self-Monitoring, but Not Dietary Quality, Improves with Use of Smartphone App Technology in an 8-Week Weight Loss Trial.” *Journal of Nutrition Education and Behavior* 46(5):440–44.

WHO. 2003. “Diet, Nutrition and the Prevention of Chronic Diseases.” *World Health Organization Technical Report Series* 916.

WHO. 2014. “Reducing Salt Intake in Populations : Report of a WHO Forum and Technical Meeting, 5-7 October 2006, Paris, France.” *WHO Library* (1):1–65.

Williamson, Donald A. et al. 2006. “Two-Year Internet-Based Randomized Controlled Trial for Weight Loss in African-American Girls.” *Obesity (Silver Spring, Md.)* 14(7):1231–43.

Winett, R. A. et al. 1991. “Altering Shoppers’ Supermarket Purchases to Fit Nutritional Guidelines: An Interactive Information System.” *Journal of Applied Behavior Analysis* 24(1):95–105.

Yang, Quanhe et al. 2011. “Sodium and Potassium Intake and Mortality Among US Adults.” *Arch Intern Med* 171(13):1183–91.

Zülli, Susan, Claudine Allemann, and Wissenschaftliche Mitarbeiterin. 2011. “Reduktion Des Salzkonsums: Reduktion Des Salzgehalts in Verarbeiteten Lebensmitteln.” *Berner Fachhochschule, Schweizerische Hochschule Für Landwirtschaft*.