DISCOVERY OF ASSOCIATION RULES OF THE RELATIONSHIP BETWEEN FOOD CONSUMPTION AND LIFE STYLE DISEASES FROM SWISS NUTRITION'S (MENUCH) DATASET & MULTIPLE SWISS HEALTH DATASETS FROM 1992 TO 2012

Timo Lustenberger[1], Helena Jenzer[2] and Farshideh Einsele[1]

[1]Section of Business Information,
Bern University of Applied Sciences, Switzerland
[2]Hospital of Psychiatry, University of Zurich, Switzerland

ABSTRACT

This article demonstrates that using data mining methods such as Weighted Association Rule Mining (WARM) on an integrated Swiss database derived from a Swiss national dietary survey (menuCH) and 25 years of Swiss demographical and health data is a powerful way to determine whether a specific population subgroup is at particular risk for developing a lifestyle disease based on its food consumption patterns. The objective of the study was to discover critical food consumption patterns linked with lifestyle diseases known to be strongly tied with food consumption. Food consumption databases from a Swiss national survey menuCH were gathered along with data of large surveys of demographics and health data collected over 25 years from Swiss population conducted by Swiss Federal Office of Public Health (FOPH). These databases were integrated and reported in a previous study as a single integrated database. A data mining method such as WARM was applied to this integrated database. A set of promising rules and their corresponding interpretation was generated. As an example, the found rules of the sample show that the consumption of alcohol in small quantities does not have a negative impact on health, whereas the consumption of vegetables is important for the supply of vitamins of the B group, which help the energy metabolism to pro-vide energy. These vitamins are particularly lacking in alcoholics and should then be taken with supplements. Another finding is that dietary supplements do little specially by diabetes. Applying WARM algorithm was beneficial for this study since no interesting rules were pruned out early and the significance of the rules could be highly increased as compared to a previous study using pure Apriori Algorithm.

KEYWORDS

Data Mining, WARM Association Analysis, Diet & Chronical Diseases, Health Informatics

1. INTRODUCTION

Lifestyle diseases are diseases that increase in frequency as countries become more industrialized and people get more aged. Lifestyle diseases include obesity, hypertension (blood pressure), heart disease, type 2 diabetes, cancer, mental disorders, and many others. They differ from the

David C. Wyld et al. (Eds): AISCA, NET, DNLP, BDHI - 2022
pp. 77-93, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.120207

infectious diseases originated from malnutrition, also called communicable diseases (CD) due to their contagious, dispersive nature. Lifestyle diseases are therefore among the so-called NC (non-communicable diseases) diseases. According to World Health Organization (WHO), the growing epidemic of chronic diseases afflicting both developed and developing countries are related to dietary and lifestyle changes [1].

Several researchers studied the relationship be-tween nutritional habits and lifestyle diseases aka chronic diseases. A. Fardet and Y. Boirie have agregated 304 pooled/meta-analyses and systematic reviews to obtain a qualitative overview of the associations between 17 food and beverage groups and the risk of diet-related chronic disease. The review of these authors confirmed that plant food groups were more protective than animal food groups against diet-related chronic diseases. Their results show that overweight, obesity, type 2 diabetes, cancer, and cardiovascular diseases accounted for 289 of the pooled/meta-analyses and systematic reviews [2]. Further, S. Fardet et al. conducted additional pooled analyses and meta-analyses of cohort studies and randomized controlled trials that linked fruit consumption with the risk of chronic disease and metabolic deregulation. Their results show that the degree of processing influences the health effects of fruit-based products. Fresh and dried fruits appeared to have a neutral or protective effect on health, 100% fruit juices had intermediary effects, and high con-sumption of canned fruit and sweetened fruit juice was positively associated with the risk of all-cause mortality and type 2 diabetes, respectively [3]. S. Schneider and al. conducted a mini–Nutritional Assessment as a promising score for evaluating malnutrition in the elderly, since nutrition intervention shortens the length of stay by diminishing the rate of complication and to identify malnourished patients and those who are at nutritional risk to treat and prevent malnutrition by chronic diseases by elderly [4].

Machine Learning and Data Mining methodologies for chronic diseases prediction and prevention in relationship with nutritional habits have been explored by different researchers Internationally. S. Lee et al conducted a study using stepwise logistic regression (SLR) analysis, decision tree, random forest, and support vector machine as an alternative and complement to the traditional statistical approaches to identify the factors that affect the health-related quality of life (HRQoL) of the elder-ly with chronic diseases and to subsequently develop from such factors a prediction model [5]. D. Qudsi and al. report in [6] from a study that aims to identify the potential benefits that data mining can bring to the health sector, using Indonesian Health Insurance company data as case study. Decision tree as a classification data mining method, was used to generate the prediction model by visualizing the tree to perform predictive analysis of chronic diseases. Z. Lei et al report in [7] of studying the relationship between nutritional ingredients and diseases such as diabetes, hypertension, and heart disease by using data mining methods. They have identified the first two or three nutritional ingredients in food that can benefit the rehabilitation of those diseases. R. McCabe et al. report in [8] of creating a simulation test environment using characteristic models of physician decision strategies and simulated populations of patients with type 2 diabetes, they state of employing a specific data mining technology that predicts encounter-specific errors of omission in representative databases of simulated physician-patient encounters and test the predictive technology in an administrative database of real physician-patient encounter data. D.W. Haslam and W.P. James re-port in [9] of an investigation in a population - based sample of 1140 children performed to derive dietary patterns related to children's obesity status. Their findings reveal that Rules derived through a data mining approach revealed the detrimental influence of the increased consumption of fried food, delicatessen meat, sweets, junk food and soft drinks. K. Lange et al. state in [10] that big data studies may ultimately lead to personalized genotype-based nutrition which could permit the prevention of diet-related diseases and improve medical therapy. A. Hearty and M. Gibney evaluate the usability of supervised data mining methods as ANNs and decision trees to predict an aspect of dietary quality an aspect of dietary quality based on dietary intake with a food-based coding system and a novel meal-based coding system [11]. A. von Reusten et al. used data from 23 531 participants of the EPIC-Potsdam study to analyze the associations between 45 single food groups and risk of major chronic diseases, namely, cardiovascular diseases (CVD), type 2 diabetes and cancer using multivariable-adjusted Cox regression. Their results show that higher intakes of low-fat dairy, butter, red meat, and sauce were associated with higher risks of chronic diseases [12]. E. Yu et al. demonstrate in [13] the usability of supervised data mining methods to extract the food groups related to bladder cancer. Their results show that beverages (non-milk); grains and grain products; vegetables and vegetable products; fats, oils and their products; meats and meat products were associated with bladder cancer risk.

To gain understanding about the impact of using data mining techniques for the analysis of lifestyle diseases that can be influenced by nutrition, we conducted a preliminary study on this matter [14]. In this preliminary previous study, we used a big database gained from a grocery store chain over a certain period along with associated health data of the same region. Association rule mining was successfully used to describe and predict rules linking food consumption patterns with lifestyle diseases. Additionally, we conducted a further study using real world health and nutritional data from Swiss population and gained interesting rules which showed the link between nutritional habits and chronical diseases [15]. In the current study, we use the same national Swiss dietary survey (menuCH) with a five times larger dataset (collected over 25 years) from the national Swiss health survey including demographical information. Based on the finding of the previous study [15], where it used the pure Apriori algorithm which resulted that some critical health-related dietary features were pruned out early in course of data mining, we have applied the Weighted Association Mining Rules (WARM) analysis to gain more accurate association rules that show the link between Swiss nutritional habits and chronical diseases.

2. DATABASE SELECTION

The data comes from the national surveys menuCH and the health survey that were carried out in Switzerland.

The national food survey menuCH [16] was carried out for the first time from January 2014 to February 2015. Over 2000 people living in Switzerland were asked about their eating habits and food consumption. The data resulting from the survey is the first representative, national nutritional survey data available in Switzerland from BLV.

The second data source comprises health data on the state of health and health-related behavior of the Swiss resident population over a period of 25 years. The Federal Statistical Office has been collecting health data from the population living in Switzerland every five years using a written and telephone questionnaire [17]. As part of this study, representative data from around 85,000 people from 1992, 1997, 2002, 2007 and 2012 are available. This data has already been pre-cleaned, attributes have been partially selected from the database and the data has been already transformed as reported in [18].

3. DATA PREPARATION FOR DATA MINING PURPOSES

In addition to the previous database reported in [18], a further reduction of the data was carried out due to the objective of the study. All data sets that could not be assigned to one of the four subject areas examined (alcohol, blood pressure, cholesterol, and diabetes) were deleted. This enabled the data volume to be massively reduced and the performance of operations with the MySQL database to be increased. A total of around 10 million responses from people to individual questions were available as a data set. A MySQL database was used to ensure the integrated collection of data from different sources on a permanent basis. The MySQL database server is very fast, reliable, and easy to use.

4. DATA TRANSFORMATION

Categories were taken over from our previous study [15] on blood pressure, cholesterol, diabetes, and alcohol consumption. Blood pressure was reduced into 6 categories. The cholesterol data was reduced to 4 categories. The diabetes data was reduced to 4 categories and finally alcohol consumption data was reduced to 4 categories. As an example, the alcohol consumption data was reduced as follows:

- Daily alcohol consumption up to 18 grams,
- Daily alcohol consumption > 18-23 grams,
- Daily alcohol consumption> 23-28 grams,
- Daily alcohol consumption> 28 grams

5. CREATION OF INTEGRATED, RELATIONAL DATABASES

The new reduced health database with 25 years data but 4 selected categories was then integrated with already existing menuCH database from our previous study [15] using five common demographical attributes available in both databases were used, such as gender, age group, household, marital status, and language to link the two databases into an integrated relational database. Figure 1 shows the revised scheme of the integrated database of health and nutrition data. The big advantage of the new structure is that the database can easily be expanded with additional topics for future investigations without having to adapt the database schema. This is very efficient and timesaving if the survey catalogue of the national health survey is expanded over the next few years and additional lifestyle topics are covered and analysed.

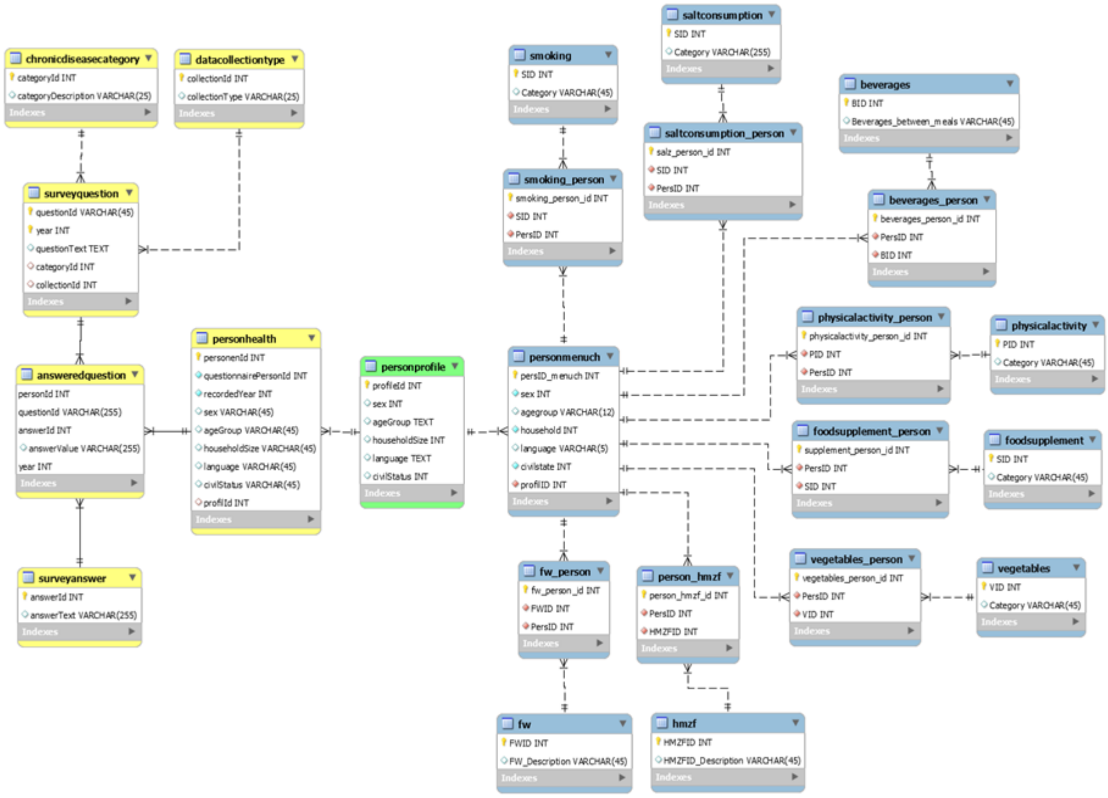

Figure 1. Scheme of integrated Database

6. Association Analysis using Apriori Algorithm

The most common method for association analysis is the Apriori algorithm [19], for which there exists several variants. In comparison, the Frequent Pattern (FP)-Growth algorithm was considered, which also examines the frequent occurrence of things together [20]. The two algorithms were contrasted and compared for the investigation using the criteria of complexity, performance, memory consumption, accuracy, iterations, and weighting. The algorithm was selected based on the relevant criteria for the task and considers improvement suggestions from the study for the survey year Occurred in 2012 [15]. Mewes et al. conclude in their study that *the weighting of features should be considered so that rules with higher technical relevance would not prun out too early in the data mining process*. Due to its simplicity and its sufficient efficiency and based on the findings of Mewes et al. in [15], we chose *Apriori algorithm in combination with the weighted association analysis (WARM)* over FR-Growth. In the weighted association analysis, the weight *w_i* assigned to each item i reflects the relative im-portance of one item to other items. The weight of an item set is derived from the individual weights for the items contained in a rule according to [21]. The aim of the association analysis is to find rules of the form "if feature A occurs, then feature B occurs with the probability of the confidence level" (A-> B). The calculation parameters support, confidence and lift were used to evaluate the rules. The algorithm continues until no item set fulfils the mini-mum support (Agrawal and Srikant, 1994). Item sets for rule formation were selected from theses 9 items. The item sets with the highest support and confidence value were selected for rule formation. In this study Apriori algorithm was applied to find rules for a set of 9 items as follows: 8 items from menuCH database and 1 item was the categorized chronical diseases from Swiss health database as described previously (see sec. categorization of health and menuCH data). Our multidisciplinary team consists of a dietician expert who carried out the assignment of weights of the 8 items in the menuCH table for application of weighted Apriori algorithm as previously described in this section.

6.1. Exemplary stepwise implementation of the Apriori algorithm with WARM for Blood Pressure

The association analysis with the Apriori algorithm in combination with WARM was carried out step by step. The pre-processed and transformed data of the integrated database of health and nutritional data form the basis for the analysis process. The total amount of data samples is calculated as total amount of interviewees multiplied by the number of asked questions for each investigated chronical disease in this study.

6.1.1. Step1: Weighting of the Characteristics

As a first step in the weighted association analysis, a specific weight was assigned to each characteristic as an item. Due to the different effects of nutritional characteristics on the chronic diseases examined, a different weighting was carried out for each subject area. The weighting was carried out in two stages, in which a weight was defined as the first stage for each characteristic within a category. The second stage is made up of all categories, which in turn have different meanings for the corresponding lifestyle disease. The following eight categories from the national nutrition survey were used for the association analysis and have positive or negative effects on chronic diseases.: Number of main meals, Activity, Consumption of vegetables, Beverages, Food supplements, Smoking, Salt consumption, Number of warm meals

The assignment of the weights per characteristic and category was carried out by a specialist in health and nutrition, as she has the necessary specialist expertise and can assess the effects of consumption on health. Figure 2 shows a section of the weighting carried out for the characteristics and categories in the clinical picture of blood pressure.

Categorie	ID	Features	Weighting Feature (% per category)	Weighting Category (% total of all categories)
Physical Activity	1	Seldom – Never (0 days per Week)	50	30
Physical Activity	2	Irregularly (1-4 days per week)	10	
Physical Activity	3	regularly (5-7 days per week)	30	
Physical Activity	4	No Answer/Don't Know/Empty	10	
Dietary Supplements	1	No Dietary Supplements	30	3
Dietary Supplements	2	takes Dietary Supplements ein	50	
Dietary Supplements	3	No Answer/Don't Know/Empty	20	
Salz Intake	1	Salz ohne Zusatz nie nachsalzen zu Hause	15	25
Salz Intake	2	Salz ohne Zusatz unregelmässig nachsalzen	10	
Salz Intake	3	Salz ohne Zusatz regelmässig nachsalzen zu	5	
Salz Intake	4	Salz mit Zusatz nie nachsalzen zu Hause	40	
Salz Intake	5	Salz mit Zusatz unregelmässig nachsalzen zu	20	
Salz Intake	6	Salz mit Zusatz regelmässig nachsalzen zu H	10	
Salz Intake	7	No Answer/Don't Know/Empty	5	
Smoking	1	Never	50	4
Smoking	2	Previously	40	
Smoking	3	Occasionally/Daily	5	
Smoking	4	No Answer/Don't Know/Empty	5	
Beverages	1	alcoholic	1	2
Beverages	a	Coffee/Tee	2	
Beverages	3	Coffee/Tea/Alcoholic	1	
Beverages	4	Coffee/Tee/Milk Drinks	3	

Figure 2. 1-Weighting of features and categories for Blood Pressure

6.1.2. Step 2: Calculation of the total weight per characteristic

Since the user-specific weighting was carried out by our dietary expert in two stages per category and characteristic, the total weight per characteristic was calculated in the second step. The total weight was required in the further course of the association analysis to derive the weight of an association rule from it. Figure 3 shows a section of the calculated total weights for the items in the categories exercise, dietary supplements, and salt consumption in the subject area of blood pressure. The two minimum values support and confidence were defined as user-specific parameters for each disease. Furthermore, the minimum and a maximum standard length for the number of characteristics of a disease was de-fined (Cengiz et al., 2019, p. 3). With the minimum rule length, the requirement was met that all highly weighted categories are included in the rules. The weights of all included categories add up to at least 95% of the total weight. As a result, only marginally relevant categories are usually not considered. For the maximum standard length, the clinical picture with all eight available categories from nutritional patterns was always used.

Physical Activity	Weight Feature	Weight Category	Total Weight
regularly (5-7 days per week)	0.3	0.3	9
irregularly (1-4 days per week)	0.1	0.3	3
No Answer/Don't Know/Empty	0.1	0.3	3
Seldom –Never (0 days per week)	0.5	0.3	15
Total	1	0.3	30

Dietray Supplements	Gewicht Merkmal	Gewicht Kategorie	Gesamtgewicht
No Dietray Supplements	0.3	0.03	0.9
Intake Nimmt Dietray Supplements	0.5	0.03	1.5
No Answer/Don't Know/Empty	0.2	0.03	0.6
Total	1	0.03	3

Salzkonsum	Gewicht Merkmal	Gewicht Kategorie	Gesamtgewicht
Salz mit Zusatz nie nachsalzen zu Hause	0.35	0.25	8.75
Salz mit Zusatz unregelmässig nachsalzen zu Hause (1-5	0.2	0.25	5
Salz ohne Zusatz nie nachsalzen zu Hause	0.15	0.25	3.75
Salz ohne Zusatz unregelmässig nachsalzen zu Hause (1	0.1	0.25	2.5
Salz mit Zusatz regelmässig nachsalzen zu Hause (6-10	0.1	0.25	2.5
No Answer/Don't Know/Empty	0.05	0.25	1.25
Salz ohne Zusatz regelmässig nachsalzen zu Hause (6-1	0.05	0.25	1.25
Total	1	0.25	25

Figure 3. Calculation of total weight per category by Blood Pressure

6.1.3. Step 3: Building frequent 2-Itemsets for Blood Pressure

The characteristics of the movement were combined with the characteristics of the blood pressure and from this the support, the weight and the weighted support were calculated. Movement was used for blood pressure for the 2-item set because it is the category with the highest user-specific weight (30%). The generation of all candidates from 2-item sets, 3-item sets, and all other item sets would have been too extensive for the scope of this study. For this reason, the procedure was descending according to the weight of the category and the characteristics were added to the item sets according to relevance. Sorting according to weight and the minimum length as parameters for the determination of association rules ensures that all characteristics of the clinical picture are included, which make up 95% of the total weight. See Figure 6. As can be seen from Figure 5, some item sets do not meet the predefined minimum support of 0.01. These subsets are marked in red in the figure and were no longer used to generate further candidates.

Blood Pressure	Transactions	Support	Weight	Weighted Support
not medically examined normal	582,624	0.541	0.25	0.1352
medically examined normal	220,958	0.205	0.3	0.0615
not medically examined low	99,435	0.092	0.15	0.0138
No Answer/Don't Know/Empty	83,941	0.078	0.05	0.0039
medically examined high	76,415	0.071	0.04	0.0028
medically examined low	5,196	0.005	0.2	0.0010
not medically examined high	8,613	0.008	0.01	0.0001

Figure 4. 1-itemset for Blood Pressure

Blood Pressure	Physical Activity	Transactions	Support	Weight	Weighted Support
not medically examined normal	regularly (5-7x per week)	492459	0.457	4.625	2.1144
medically examined normal	regularly (5-7x per week)	188534	0.175	4.65	0.8139
not medically examined low	regularly (5-7x per week)	85148	0.079	4.575	0.3616
No Answer/Don't Know/Empty	regularly (5-7x per week)	71078	0.066	4.525	0.2986
medically examined high	regularly (5-7x per week)	65231	0.061	4.52	0.2737
not medically examined normal	irregularly (1-4x per week)	57983	0.054	1.625	0.0875
not medically examined normal	seldom - never (0 days per week)	8357	0.008	7.625	0.0592
not medically examined normal	No Answer/Don't Know/Empty	23825	0.022	1.625	0.0359
not medically examined high	regularly (5-7x per week)	7335	0.007	4.505	0.0307
medically examined normal	irregularly (1-4x per week)	19436	0.018	1.65	0.0298
medically examined normal	seldom - never (0 days per week)	3666	0.003	7.65	0.0260
medically examined low	regularly (5-7x per week)	4463	0.004	4.6	0.0191
medically examined normal	No Answer/Don't Know/Empty	9322	0.009	1.65	0.0143
not medically examined low	irregularly (1-4x per week)	9139	0.008	1.575	0.0134
No Answer/Don't Know/Empty	irregularly (1-4x per week)	8239	0.008	1.525	0.0117
No Answer/Don't Know/Empty	seldom - never (0 days per week)	1404	0.001	7.525	0.0098
medically examined high	irregularly (1-4x per week)	6706	0.006	1.52	0.0095
not medically examined low	seldom - never (0 days per week)	947	0.001	7.575	0.0067
medically examined high	No Answer/Don't Know/Empty	3260	0.003	1.52	0.0046
not medically examined high	seldom - never (0 days per week)	180	0.000	7.505	0.0013
medically examined low	irregularly (1-4x per week)	446	0.000	1.6	0.0007
medically examined low	seldom - never (0 days per week)	86	0.000	7.6	0.0006

Figure 5. Frequent 2-itemsets for Blood Pressure

6.1.4. Step 4: Building frequent 3-Itemsets

To generate the 3-item set, salt consumption was added as an additional feature to the shortened 2-item set of blood pressure and exercise. Salt consumption is the nutritional trait with the second highest weighting (25%) after exercise. Figure 6 shows an excerpt from the frequent subsets as a 3-item set, consisting of the three attributes mentioned for the topic of blood pressure.

Physical Activity	Salt Intake	Transactions	Support	Weight	Weighted Support
regularlly (5-7 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	201904	0.1874	6.000	1.1246
regularlly (5-7 days per week)	Salz mit Zusatz unregelmässig nachsalzen zu	123698	0.1148	4.750	0.5455
regularlly (5-7 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	83998	0.0780	6.017	0.4692
regularlly (5-7 days per week)	Salz ohne Zusatz nie nachsalzen zu Hause	69349	0.0644	4.333	0.2790
regularlly (5-7 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	39310	0.0365	5.967	0.2177
regularlly (5-7 days per week)	Salz mit Zusatz unregelmässig nachsalzen zu	48782	0.0453	4.767	0.2159
regularlly (5-7 days per week)	Salz ohne Zusatz unregelmässig nachsalzen z	47027	0.0437	3.917	0.1710
regularlly (5-7 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	29439	0.0273	5.930	0.1621
regularlly (5-7 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	28105	0.0261	5.933	0.1548
irregularlly (1-4 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	28355	0.0263	4.000	0.1053
regularlly (5-7 days per week)	Salz ohne Zusatz nie nachsalzen zu Hause	25333	0.0235	4.350	0.1023
regularlly (5-7 days per week)	Salz mit Zusatz unregelmässig nachsalzen zu	16580	0.0154	4.680	0.0720
irregularlly (1-4 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	10038	0.0093	4.017	0.0374
regularlly (5-7 days per week)	Salz ohne Zusatz nie nachsalzen zu Hause	8729	0.0081	4.263	0.0345
Selten – nie (0 Tage pro Woche)	Salz mit Zusatz nie nachsalzen zu Hause	2058	0.0019	8.017	0.0153
irregularlly (1-4 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	3535	0.0033	3.930	0.0129
regularlly (5-7 days per week)	Salz mit Zusatz regelmässig nachsalzen zu Ha	2978	0.0028	3.847	0.0106
regularlly (5-7 days per week)	Salz mit Zusatz nie nachsalzen zu Hause	1891	0.0018	5.983	0.0105
irregularlly (1-4 days per week)	Salz mit Zusatz unregelmässig nachsalzen zu	3731	0.0035	2.767	0.0096
regularlly (5-7 days per week)	No Answer/Don't Know/Empty	2876	0.0027	3.517	0.0094
regularlly (5-7 days per week)	Salz ohne Zusatz regelmässig nachsalzen zu I	2777	0.0026	3.517	0.0091
No Answer/Don't Know/Empty	Salz mit Zusatz nie nachsalzen zu Hause	1949	0.0018	3.930	0.0071
regularlly (5-7 days per week)	Salz mit Zusatz unregelmässig nachsalzen zu	1183	0.0011	4.733	0.0052

Figure 6. Frequent 3-itemsets for Blood Pressure

6.1.5. Further steps: Building frequent 9-Itemsets

This iterative procedure was repeated until all categories of nutritional behavior in combination with the corresponding clinical picture were includ-ed as an item set. For each frequent item set, the support, weight and weighted support were calcu-lated. All subsets that met the weighted minimum support were then used to generate the next candi-date item set. The last iteration is the generation of the 9-itemset with all available features. A section of this can be seen in Figure 7 for the clinical pic-ture of blood pressure.

After all candidate item sets had been generat-ed, the confidence value for each item set was cal-culated. For this purpose, the support of the corre-sponding item set was divided by the support of the disease as a premise. To identify the association rules in the next step, all item sets are used that meet both the minimum weighted support and the predefined minimum confidence value.

Blood Pressure	Physical Activity	Salt Intake	Main Meals	No. Hot Meals	Vegetable Intake	Smoking	Dietary supplements	Beverages	Transactions	Support	Weight	Weighted Support
not medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	irregullary warm meal(4	regullary vegetables	Previously	No Intake Dietary Supplement	Water/Coffee/Tea	6750	0.0063	3.027	0.01897
not medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	regullary warm meal(8	regullary vegetables	Never	Intake Dietary Supplements	Water/Coffee/Tea	5966	0.0055	3.249	0.01799
not medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	regullary warm meal(8	regullary vegetables	Never	No Intake Dietary Supplement	Water/Coffee/Tea	5699	0.0053	3.182	0.01684
not medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	regullary warm meal(8	regullary vegetables	Never	No Intake Dietary Supplement	Water/Coffee/Tea, SFGEI	4614	0.0043	3.182	0.01363
not medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	irregullary warm meal(4	regullary vegetables	Never	Intake Dietary Supplements	Water/Coffee/Tea	4150	0.0039	3.138	0.01209
medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	irregullary warm meal(4	regullary vegetables	Previously	No Intake Dietary Supplement	Water/Coffee/Tea	3743	0.0035	3.032	0.01054
not medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	irregullary warm meal(4	regullary vegetables	Never	No Intake Dietary Supplement	Water/Coffee/Tea	3658	0.0034	3.071	0.01043
medically exmined normal	regularly(5-7 d	Salz mit Zusatz Never nach	FS regel./ME regel. /AE re	regullary warm meal(8	regullary vegetables	Never	No Intake Dietary Supplement	Water/Coffee/Tea	3443	0.0032	3.188	0.01019

Figure 7. Excerpt of frequent 9-itemsets for Blood Pressure

6.1.6. Final step: Building Association Rules

The identification of the association rules is the second phase of the a priori procedure. The frequent patterns of the database from the previous step were used to derive interesting association rules using the calculated confidence value. From the frequent quantities of the four subject areas, all item sets were extracted that meet the specified minimum confidence value and the minimum standard length. Thereafter, all subsets were removed whose disease diagnosis was not assessed by a healthcare professional. Finally, the results were sorted in descend-ing order by confidence.

7. RESULTS USING WEIGHTED APRIORI ALGORITHM (WARM)

After completion of the algorithm rules were found that show the relationship between nutrition and chronic diseases. We report in the following gained rules for alcohol, blood pressure, cholesterol, and diabetes.

7.1. Rules for Alcohol

Rule 1: 4.1% of the sample, who consume 0-17 grams of alcohol daily, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee or tea between the meals, they eat irregularly hot meals, they prepare more than twice a week vegetable, they exercise regularly five to 10 times a week and take nutritional supplements. This rule occurs in 3.4% of the sample

Rule 2: 3.6% of the sample, who consume 0-17 grams of alcohol daily, show the following characteristics: the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between the meals, they eat regularly hot meals, they prepare more than twice a week vegetable, they exercise regularly five to 10 times.

Rule 3: 3.1% of the sample, who consume 18-22 grams of alcohol daily, show the following characteristics: the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between the meals, they eat irregularly hot meals, they prepare more than twice a week vegetable, they exercise regularly five to 10 times a week and take nutritional supplements. This rule occurs in 0.2% of the sample.

Rule 4: 2.9% of the sample, who consume 18-22 grams of alcohol daily, show the following characteristics: the following characteristics: they eat 3 times a day regularly, they drink water, coffee or tea between the meals, they eat regularly hot meals, they pre-pare more than twice a week vegetable, they exercise regularly five to 10 times a week and take no nutritional supplements. This rule occurs in 0.19% of the sample.

Rule 5: 2.3% of the sample, who consume more than 22 grams of alcohol daily, show the following characteristics: the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between the meals, they eat seldom or almost no hot meals, they don't prepare vegetables, they exercise regularly five to 10 times a week and take no nutritional supplements. This rule occurs in 0.17% of the sample.

Rule 6: 6.6% of the sample, who have a medic, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between the meals, they eat seldom or almost no hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 10 times a week and take no nutritional supplements. This rule occurs in 0.16% of the sample.

7.2. Rules for Blood Pressure

Rule 1: 6.6% of the sample, who have medically assessed high blood pressure, show the following characteristics: they eat 3 times a day regularly, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke or haven't ever smoked. This rule occurs in 0.47% of the sample.

Rule 2: 5.1% of the sample, who have medically assessed high blood pressure, show the following characteristics: they eat 3 times a day regularly, they eat irregularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke or haven't ever smoked. This rule occurs in 0.36% of the sample.

Rule 3: 4.1% of the sample, who have medically assessed high blood pressure, show the following characteristics: they eat 3 times a day regularly, they eat irregularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke but have previously smoked. This rule occurs in 0.29% of the sample.

Rule 4: 6.5% of the sample, who have medically assessed normal blood pressure, show the following characteristics: they eat 3 times a day regularly, they eat irregularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke or haven't ever smoked. This rule occurs in 1.33% of the sample.

Rule 5: 5.3% of the sample, who have medically assessed normal blood pressure, show the following characteristics: they eat 3 times a day regularly, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke or haven't ever smoked. This rule occurs in 1% of the sample.

Rule 6: 5% of the sample, who have medically assessed normal blood pressure, show the following characteristics: they eat 3 times a day regularly, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke but have previously smoked. This rule occurs in 0.78% of the sample.

7.3. Rules for Cholesterol

Rule 1: 6.3% of the sample, who have medically assessed high cholesterol, show the following characteristics: they eat 3 times a day regularly, they eat irregularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, and don't smoke or haven't ever smoked. This rule occurs in 0.21% of the sample.

Rule 2: 4.8% of the sample, who have medically assessed high cholesterol, show the following characteristics: they eat 3 times a day regularly, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke or haven't ever smoked. This rule occurs in 1.33% of the sample.

Rule 3: 3.8% of the sample, who have medically assessed high cholesterol, show the following characteristics: they eat 3 times a day regularly, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don't add salt during eating and don't smoke but have previously smoked. This rule occurs in 1.33% of the sample.

Rule 4: 6.3% of the sample, who have medically assessed normal cholesterol, show the following characteristics: they eat 3 times a day regularly, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don’t add salt during eating and don’t smoke or haven’t ever smoked. This rule occurs in 0.76% of the sample.

Rule 5: 4.4% of the sample, who have medically assessed normal cholesterol, show the following characteristics: they eat 3 times a day regularly, they eat irregularly hot meals, they prepare vegetables more than twice a week, they exercise regularly five to 7 times a week, they cook with salt but don’t add salt during eating and don’t smoke or haven’t ever smoked. This rule occurs in 0.54% of the sample

7.4. Rules for Diabetes

Rule 1: 6.3% of the sample, who have medically assessed high diabetes, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between meals, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise 2 to 5 times a week, they don’t take any meal supplements. This rule occurs in 0.04% of the sample.

Rule 2: 3.3% of the sample, who have medically assessed high diabetes, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between meals, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise 2 to 5 times a week, they take any meal supplements. This rule occurs in 0.04% of the sample.

Rule 3: 3.3% of the sample, who have medically assessed high diabetes, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between meals, they eat irregularly hot meals, they prepare vegetables more than twice a week, they exercise 2 to 5 times a week, they don’t take any meal supplements. This rule occurs in 0.04% of the sample.

Rule 4: 3.4% of the sample, who have medically assessed normal diabetes, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between meals, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise 2 to 5 times a week, they don’t take any meal supplements. This rule occurs in 0.14% of the sample.

Rule 5: 3.3% of the sample, who have medically assessed normal diabetes, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between meals, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise 2 to 5 times a week, they take meal supplements. This rule occurs in 0.13% of the sample.

Rule 6: 2.9% of the sample, who have medically assessed normal diabetes, show the following characteristics: they eat 3 times a day regularly, they drink water, coffee, or tea between meals, they eat regularly hot meals, they prepare vegetables more than twice a week, they exercise 2 to 5 times a week, they take any meal supplements. This rule occurs in 0.11% of the sample.

8. KNOWLEDGE INTERPRETATION

8.1. Alcohol

The consumption of alcohol is culturally strongly anchored in Western society and has been an integral part of social life for centuries. Excessive alcohol consumption is nevertheless a major cause of premature mortality and damage to physical (especially liver and digestive disorders), mental and social health. It is also a cause of violent behavior, accidents, early disability, lost work, or social exclusion.

Rules 1 and 2 are probably people with a high level of health and nutrition awareness. Warm meals are an expression of an even greater nutritional awareness.

Rules 3 and 4 are about people who have found a good balance between nutrition and health awareness on the one hand, and enjoyment of life and culture on the other. If alcohol and cuisine are sensibly combined, this indicates a balanced person.

Rules 5 and 6 with an alcohol consumption of more than 28 grams of alcohol per day result in significant health disadvantages, especially on the liver. Without vegetables, important B vitamins are missing, which can lead to a manifest deficiency for alcoholics. In this case, these micronutrients should be taken with supplements.

According to the rules, alcohol consumption in quantities of 18-22 grams per day is ideal for health. A complete renunciation of alcohol or slightly above-average consumption is not absolutely necessary as long as the alcohol dehydrogenase can oxidize the amount absorbed into acetaldehyde and acetic acid. If this metabolic reaction is permanently overwhelmed, it ultimately leads to a hardened liver (cirrhosis), which increasingly hinders the blood flow. This blood flow is necessary insofar as the liver is the most important detoxification organ of the organism. This gives people an advantage who can find a good balance between indulgence in modest amounts and avoiding chronic consumption of alcohol. Exercise can help improve circulation in the liver as well. It is ideally combined with plenty of fluids (drinking water, coffee, tea) between meals to eliminate metabolic waste products from the liver. This behavior is followed for all rules.

The consumption of regular but not too frequent meals provides building blocks and energy sources. The building blocks that are not used immediately are deposited until they are hungry, in which they are released from the depot. The distinction between warm and not regularly warm meals can be an expression of nutritional awareness, but also the result of time management (e.g., sandwich lunch). It does little to alleviate or increase the effects of alcohol consumption. The frequency of meals is important if a cantable (consuming) metabolic process predominates (e.g., a serious illness such as cancer, obstructive pulmonary disease, HIV, or others). In this case, food has to be fed frequently or continuously. In healthy people, eating twice a day can be much healthier than eating three times a day (interval diet). The consumption of vegetables is important for the supply of vitamins of the B group, which help the energy metabolism (citrate cycle and respiratory chain) to provide energy. This is important for endurance athletes. For sprinters and strength athletes, the short-term energy is mainly taken from the carbohydrate metabolism. Vegetable intake is an expression of a high level of health awareness. For athletes, vegetable intake can usefully be supplemented with supplements, especially for endurance sports.

8.2. Blood Pressure

High blood pressure is a chronic disease that is usually caused by a decline in cardiac output, declining kidney function (lack of filtration capacity) and the hardening of the blood vessels. Often one finds genetic predispositions combined with improper nutrition. Excess fats and cholesterol in particular lead to vascular problems in the fat metabolism. This can cause excessive foam cells to form and lead to diabetes. In this respect, high blood pressure, hyperlipidemia, hypercholesterolemia and diabetes (type 2) are often associated with one another. A healthy heart can fight against increased counter pressure for years and overcome it until a stroke suddenly occurs in the heart or brain. As a comparison, imagine a garden pump for irrigation with a hose and sprinkler. A clogged sprinkler and calcified hoses will sooner or later bring any engine to a standstill due to overheating if the power is no longer sufficient to transport water.

Rules 1, 2 and 3 contain some references to culinary lovers. This is not a problem as long as the weight is sufficiently controlled. In the case of high blood pressure, any weight reduction is helpful in reducing the symptoms.

Rules 4, 5 and 6 concern people who have normal blood pressure (normotonic). They eat irregularly warm meals more often, but they are nutritionally conscious in terms of vegetables and regular food intake.

The most important subset in this subject area are people with medically diagnosed high blood pressure. Weight reduction, exercise, smaller amounts of food and less frequent food intake help here.

In the case of high blood pressure, the warm food is probably an indication of health awareness. Salt should be reduced when the kidney is already inadequate. In terms of blood pressure rules, salt consumption is the same everywhere. Vegetables are also fed in all cases.

8.3. Cholesterol

Hypercholesterolemia is a disease of the fat metabolism. Cholesterol is produced by the body's own biosynthesis. If too much carbohydrate is ingested, it is broken down into glucose. The breakdown product acetyl-CoA can be linked to fatty acids or via steroid biosynthesis to hormones such as sex hormones, corticoids or mineralocorticoids. The internal excess cholesterol biosynthesis is genetically determined and more difficult to control than the dietary cholesterol. Hypercholesterolemia usually leads to vascular problems, atherosclerosis, diabetes or a stroke.

Rules 1 to 5 have identical physical activity. It is important for people to promote the breakdown of cholesterol.

Rules 1 to 3 show that, as with high blood pressure, health awareness is also well developed in diagnosed hypercholesterolemics. This is shown by the fact that the people do not smoke and in some cases have not smoked before. Smoking would seal the vessels even more. The behavioral patterns are also the same when it comes to the intake of salt and vegetables.

The difference between **Rules 1 to 3** and **Rules 4 to 5** is probably only quantitative in relation to fat intake. However, this was not collected and could not be considered in this research.

8.4. Diabetes

Diabetes is a carbohydrate metabolism disease. Diabetes is divided into type 1 diabetes and type 2 diabetes, which occurs with increasing age. As with cholesterol, genetics are also essential in glucose metabolism. The disease is found genetically more frequently in certain ethnic groups. As an example, people in the Indian subcontinent from countries like India, Pakistan, Tibet or Nepal can be named. In this area, most people die from cardiovascular problems at a young age. Gestational diabetes can often be diagnosed in women due to their genetic predisposition. In this case, the high progesterone level during pregnancy leads to gluconeogenesis and thus to increased blood glucose levels. It is for this reason that the term gestational diabetes is also used.

A permanently high blood sugar level leads to the exhaustion of the insulin secretion and the islet cells of Langerhans. While previously the insulin injection and irritation of the pancreas were used to release more insulin, there are now new antidiabetic drugs that slow down the digestion of carbohydrates in the gastrointestinal tract and thus lead to a massively higher efficiency of the pancreatic work. Researchers have discovered the mechanism in lizards, which ingest prey at very long intervals and digest it so slowly that the next food intake is not necessary for a long time.

Rules 1 to 3 show that a confirmed diagnosis of diabetes leads to significantly better health and nutritional behavior and understanding of nutrition. This is probably reflected in the great care taken in the preparation of meals (warm, with lots of vegetables and regularly in small quantities).

From **Rules 1 to 3** it can also be inferred that the intake of food supplements has little effect on the metabolic situation, unless movement is increased drastically and energy is generated with the micronutrients (B vitamins, iron, copper and the like) improved.

With **Rules 4 to 6**, people without a medical diagnosis of diabetes generally eat less regularly, in a lavish manner, with or without supplements.

9. CONCLUSION AND FUTURE WORK

In this paper, we apply a data mining method such as WARM Apriori algorithm to a big integrated database comprising of Swiss nutrition and health data to gain rules that show the effects of nutritional habits on some chronical diseases such as high alcohol consumption, high blood pressure, Diabetes, and high Cholesterol. For this purpose, we use an integrated database comprised of collected data from various Swiss national surveys as reported in a previous study (Lustenberger, 2021). Our database includes health data on the state of health and health-related habits of around 85,000 people over a period of 25 years (1992-2012) as well as data on food consumption, cooking, eating and physical activity habits of around 2,000 people in 2015 and 2016.

A deficiency in the data used for data mining is the deviating collection period for health and consumption data. While the health data was collected from the Swiss resident population every five years over a period of 25 years, data for food consumption comes from a national survey from 2014 to 2015. This affects the result insofar as consumer habits as a cause is not directly reflected in health as an effect. The habits of the population, such as physical activities or diet, changes over a longer period and in turn has an impact on health. Ideally, both surveys should be carried out repeatedly to gain even more informative value of the results and to ensure the causal
The algorithmic improvements as suggested in the study by Mewes et al. could be successfully implemented in the execution of the association analysis. The problem with removing sets of attribute values when using the a priori algorithm was solved with the WARM method. Using an expert weighting of characteristics enables the related dietary relevance to be better considered in the results than is the case with pure Apriori and ensures this relevance to not be pruned out in the mining process.

The interpretation of the derived rules reveals interesting aspects about the selected Swiss population subgroup. The study shows that the consumption of alcohol in small quantities does not have a negative impact on health. In addition to this, regular exercise in combination with an adequate increase in fluids in the form of water, coffee or tea between meals improves the circulation in the liver and helps to eliminate the metabolic waste products. Additionally, the consumption of vegetables is also important for the supply of vitamins of the B group, which help the energy metabolism to provide energy. These vitamins are particularly lacking in alcoholics and should then be taken with supplements.

Furthermore, for people with medically diagnosed high blood pressure, weight loss, regular exercise, smaller amounts of food and less frequent food intake help as measures. The analysis shows that the people in the sample with high blood pressure probably gained better health awareness because of the diagnosis. Smoking habit with diagnosed high blood pressure is either non-smokers or a status after smoking cessation. In addition, these people eat hot meals more often, which is also an indication of awareness. Notwithstanding, for an even more comprehensive assessment within the scope of this study, the body mass index (BMI) is missing as an important risk factor.

Regarding a high cholesterol value (hypercholesterolemia) as a disease of the fat metabolism, a low-cholesterol diet can be effective, but only if the genetic predisposition is not given and a fat intake that is much too high is reduced. If the person concerned already consumes little fat, an improvement in the diet can hardly be achieved. In addition, people diagnosed with high cholesterol levels show a pronounced health awareness. This is shown in the presented study by the fact that the people do not smoke. Smoking would, in addition to causing illness, block the blood vessels even more.

Furthermore, the rules and patterns in diabetes (type 2) show that the medical diagnosis of diabetes leads to significantly better health and nutritional habits and understanding of nutritional values of food intake. This was shown by the balanced preparation of meals (warm, vegetable-rich, and regular). Another finding is that dietary supplements do little when it comes to diabetes.

REFERENCES

[1] WHO, 2003. Diet, Nutrition, and the Prevention of Chronic Diseases. Report of a Joint WHO/FAO Ex-pert Consultation. In World Health Organization aper templates.

[2] Fardet, A., Boirie, Y. 2008. Associations between food and beverage groups and major diet-related chronic diseases: an exhaustive review of pooled/meta-analyses and systematic reviews, In Nutr Rev. 2014 Dec; 72(12):741-62. doi: 10.1111/nure.12153

[3] Fardet, A. Richonnet, C., Mazur, A., 2019, Association between consumption of fruit or processed fruit and chronic diseases and their risk factors: a systematic review of meta-analyses, Nutrition Reviews. In Nu-trition Reviews, Volume 77, Issue 6, Pages 376-387.

[4] Schneider, S., Heuterne, X., 2000, Moore, R., Lopes, J., 1999. Prediction Model for Health-Related Quality of Life of Elderly with Chronic Diseases using Ma-chine Learning Techniques. In Healthc Inform Res. 2014 Apr;20(2):125-134.

[5] Kee, S. K., Son, Y. J, Kim H.G., Lee J. Il., Cho, H.S., Lee, S., 2014, Associations between food and bever-age groups and major diet-related chronic diseases: an exhaustive review of pooled/meta-analyses and systematic reviews, In Nutr Rev. 2014 Dec; 72(12):741-62. doi: 10.1111/nure.12153

[6] Qudsi, D., Kartiwi, M., Saleh, N.B., 2017, Predictive data mining of chronic diseases using decision tree: A case study of health insurance company in Indonesia. In International Journal of Applied Engineer-ing Research 12(7):1334-1339

[7] Lei Z., Yang, S., Liu, H., Aslam, S., Liu, J., Bugingo, E., Zhang, D., 2018, Mining of Nutritional Ingredi-ents in Food for Disease Analysis, In IEEE Access 6(1):52766-52778

[8] McCabe, R.M, Adomavicius, G., Johnson P.E., Rund, E., Rush, A., Sperl-Hillen, A., 2008, Using Data Mining to Predict Errors in Chronic Disease Care, Advances in Patient Safety: In New Directions and Alternative Approaches in Vol. 3: Performance and Tools.

[9] Haslam, D.W., James, W.P.T., Obesity, In the Lancet, Volume 366, Issue 9492, Pages 1197-1209

[10] Lange, K.W., James W.P.T., Makulska-Gertruda E., Nakamura Y., Reissmann, A., 2008, A. Sperl-Hillen, Using Data Mining to Predict Errors in Chronic Disease Care, Advances in Patient Safety. In New Directions and Alternative Approaches (Vol. 3: Performance and Tools)

[11] Hearty, A.P., Gibney, M.J., 2008, A. Richonnet, C., Mazur, A., Analysis of meal patterns with the use of supervised data mining techniques—artificial neural networks and decision trees, In The American Journal of Clinical Nutrition, Volume 88, Issue 6, Pages 1632–1642.

[12] Von Ruesten, A., Feller, S., Bergmann, N.M., Boeing, H., 2013, S., Diet and risk of chronic diseases: results from the first 8 years of follow-up in the EPIC-Potsdam study, In European Journal of Clinical Nutrition volume 67, pages412–419.

[13] Yu E. Y. W., Wesselius A., Sinhart C., Wolk A., 2020, A data mining approach to investigate food groups related to incidence of bladder cancer, In the Bladder cancer Epidemiology and Nutritional Determinants International Study, Cambridge University Press

[14] Einsele, F., Sadeghi, L., Ingold, R., Jenzer, H., 2015, A Study about Discovery of Critical Food Consumption Patterns Linked with Lifestyle Diseases using Data Mining Methods, In HealthInf, BIOSTEC - International Joint Conference on Biomedical Engineering Systems and Technologies, Lisbon.

[15] Mewes I., Jenzer H., Einsele, F., 2020, building an integrated relational database from Swiss Nutrition's (menuCH) and Swiss Health datasets for Data Mining Purposes, submitted and accepted In ICAFNH 2021: International Conference on Agrilife, Food, Nutrition and Health

[16] BLV (2021). https://www.blv.admin.ch/blv/de/home/ lebensmittel-und-ernaehrung /ernaehrung/ menuch.html, date: 2/12/2021.

[17] BAG (2021). https://www.bag.admin.ch/bag/de/home/ zahlen-und-statistiken.html, date: 2/12/2021.

[18] Lustenberger, T., Jenzer, H., Einsele, F., 2021, Building an Integrated Relational Database from Swiss Nutrition's (menuCH) and multiple Swiss Health Datasets acquired from 1992 to 2012 for Data Mining Purposes, In Proceedings of the 10th International Conference on Data Science, Technology and Applications (DATA 2021), pages 150-15

[19] Agrawal R., Srikant, R., 1994, Fast algorithms for mining association rules. In IBM Research Report RJ9839, IBM Almaden Research Center, San Jose, California

[20] Cleve, J. & Lämmel, U. (Hg.). (2016). De Gruyter Studium. Data Mining (2. Aufl.). De Gruyter. https://doi.org/10.1515/9783110456776

[21] Cengiz, A. B., Birant, K. U. & Birant, D. (2019). Analysis of Pre-Weighted and Post-Weighted Association Rule Mining. In 2019 Innovations in Intelligent Systems and Applications Conference (ASYU).

AUTHORS

Farshideh Einsele, Prof. Dr., Lecturer and researcher in the Business section of Bern University of Applied Sciences.
She teaches business informatic subjects and her research work is dedicated to epidemiology, big data, and data mining.

Helena Jenzer, Dr., Chief Pharmacist FPH - PhD, university, Hospital PUK Zurich, formerly; Prof Dr Head of applied R&D Nutrition & Dietetics in Bern University of Applied Sciences.

Timo Lustenberger, Student, and member of research team Epidemiology Business Department of Bern University of Applied Sciences.