Personalised Coaching of Patients with Type 2 Diabetes Based on Their Individual Behavioural Characteristics

Tom M. Nyhoff (2514303)
Department of Psychology, University of Twente

202000379: Bachelor Thesis Research Methodology, Measurement, Data-analysis

1st Supervisor: Dr. Stephanie van den Berg

2nd Supervisor: Eclaire Hietbrink

July 5th, 2023

Summary

eHealth interventions have become a very important tool in the last years to support healthcare across many domains by helping patients to change specific lifestyles in a healthy way. One factor that is very important to make these eHealth interventions effective is personalisation, meaning that interventions are exactly fitted to the needs of every individual patient. This research examined how statistical methods like cluster analysis can be optimally used to support the development of eHealth interventions by allowing better personalisation. More specifically, this can be done by identifying groups of patients who share similar characteristics and would therefore profit from specific coaching strategies. The three unsupervised learning methods, K-Means clustering, Self-Organising Map algorithms and Swarm Intelligence Based clustering, were chosen and compared using a dataset of Type 2 Diabetes patients from the Diabetes and Lifestyle Cohort Twente. These three methods were applied to a dataset measuring the patients' physical activity using their step count at specific times as a measure, a dataset measuring the nutritional intake based on food diaries, and a dataset combining those two. The results of these methodologies were then evaluated based on interpretability and cluster quality. Interpretability was measured by analysing the number of variables showing significant differences between groups and how relevant these variables are for Type 2 Diabetes based on the literature. Mean silhouette scores were used as a measure of cluster quality. Overall, the results suggest that K-means clustering is a valid choice overall, as its results for the silhouette scores suggest the best cluster quality of the three methods. Interestingly SOM and Swarm Intelligence-based clustering performed worse than K-means contradicting expectations set by the literature. However, based on the low cluster quality for all measures and several limitations of the study, further exploration is suggested.

Personalised Coaching of Patients with Type 2 Diabetes Based on Their Individual Behavioural Characteristics

In today's rapidly evolving world, the importance of technology continues to grow in all areas. It is not surprising that these developments also show no bounds when going into the area of healthcare. eHealth or digital health describes the use of digital technology with the goal of delivering or further improving high-quality healthcare (Eysenbach, 2001). The application of eHealth ranges from telehealth which allows patients to get in contact with doctors from their homes to health wearables which constantly track patients' vital signs, like, for example, the glucose level of diabetes patients (HRSA, 2023; Freestyle Libre, 2023). The rapid growth in use and the broad range of applications seem really promising, especially when considering that eHealth has many strengths, as it is, for example, very cost-effective, reduces the work of health professionals and is easy to use for patients if designed well (Arief et al., 2013). This appears to be especially important as the shortage of healthcare workers that already leads to problems in several sectors appears to only grow further in the following years (Meershoek et al., 2022). Next to applying different technological health benefits individually, newer approaches to eHealth try to improve the health of patients systematically by combining different technological possibilities.

The Application of eHealth Interventions for Lifestyle Behavior Change

A specific area of attention within eHealth is interventions to stimulate health behaviour. These interventions use different forms of technology, like apps or programmes, that can help people change their own behaviour with the goal of improving their own health. This is important because healthy behaviour in many domains, like eating or physical activity, can often prevent the development of disease and help alleviate symptoms of already existing illnesses (Fisher et al., 2011). eHealth interventions are used in many different domains, which can be used very generally to, for example, support healthy ageing by remote monitoring of vital signs and suggesting training that fits through an app (Buyl et al., 2020). But they can also be very specific aimed at, for example, supporting male taxi drivers suffering from cardiovascular disease by trying to initiate more physical activity between drives through pop-up messages on the phone (McMahon et at., 2022). Studies about the effectiveness of digital interventions to change lifestyle behaviours and especially those related to physical activity and eating behaviour show that they can be very beneficial if used in a good way (Garcia-Ortiz et al., 2018; Schoeppe et al., 2016). Special about eHealth interventions in these cases is also that they are easy to distribute to very large numbers of patients and also to those patients who are hard to reach (Granja et al., 2018; McMahon et al., 2022).

To enable and improve the functionality of many of these eHealth interventions, data analysis plays an important role. Its main tasks lie in using the collected data to track the process of individual patients and give health professionals more objective information that they, in turn, can use to adapt the treatment of the patient by providing personalised advice. (Watson & Watson, 2014). Additionally, it is regularly used to evaluate how effective an intervention is by comparing the measurements of health parameters before and after an intervention, for example, shown in the study by McMahon et al. (2022). In the last years, objective lifestyle data started to support eHealth interventions in an additional way by allowing better development of more personalised interventions where lifestyle recommendations are based on individual data. These are very interesting developments as little attention was paid to this aspect before.

Personalisation

Personalisation or tailoring of eHealth interventions describes the process of adapting these technologies completely or in parts to the specific requirements of each individual user (Searby, 2003). The recent developments of including more personalisation into eHealth intervention appear to be very beneficial as personalised interventions seem to outperform more general ones across most health domains (Lau et al., 2020; Celis-Morales et al., 2015).

The area of eHealth interventions appears to be perfect for the use of personalisation, especially in application to large numbers of patients. This is because eHealth interventions can be programmed in a way so that they continuously collect the data of the patients and use this data to automatically personalise the intervention for the patient, without any time investment by humans. A regularly used way to do this is to compare all parameters collected from the patient to those collected from other patients and then create groups of patients based on their similarities. One example of this would be that some patients with binge eating disorders might express their binges at very specific and constant times of the day (Moghimi et al., 2021). By understanding that a person belongs to the group binging at this specific time, eHealth interventions can be personalised by offering more support during these more critical moments. The problem is that very often, these group similarities are not directly visible, which is why thorough data analysis is necessary to understand the data and identify the similarities.

The more data is collected and available, the better the foundation for data analysis to identify groups and make precise personalisations. However, especially in the early stages of applying these eHealth interventions, the available participant data is very limited, which leads to problems with the reliability of the results (Chevance et al.,2021). As collecting more data is a tedious process often related to high costs, this is often not a possibility at the beginning of development. Because of this, a considerate selection of the used methodologies to analyse the present data is often the best choice to allow good personalisation.

In general, unsupervised machine learning methods appear to be a good fit to support the personalisation of eHealth interventions as they can identify patterns and groups in the data which are hard to identify using other means (IBM, 2023). However, when taking the problem of a small sample size and the broad number of available unsupervised machine-learning techniques into account, it becomes hard to decide which method is best for the application in eHealth interventions.

Unsupervised Machine Learning Methods

Most of the statistical models that are commonly used in unsupervised machine learning are optimised towards higher sample sizes (Singh & Masuku, 2014). This means that even those methods that are usually considered very established and that show consistently good results in statistical research might encounter problems when applied to small samples. When planning to improve the selection of methods used, especially for small samples, it is therefore important to compare and evaluate carefully, as less established methods might have higher potential then more established ones perfectly optimised towards higher sample sizes.

K-Means Clustering

In order to enable good comparison and potentially find significant results, it is important to choose viable unsupervised learning methods. When considering which unsupervised learning method can be used, K-means clustering is commonly used and well-established in research with a broad background of publications (Jain et al., 1999; Kodinariya & Makwana, 2013). K-means clustering, therefore, appears to be a good benchmark to understand how well current methods perform in finding groups in eHealth intervention data. Besides, it is interesting to explore other more novel unsupervised learning techniques to see if there are differences in the quality of results in smaller sample sizes.

Self-Organising Map Algorithms

Another very interesting method for discussion in this paper is self-organising map algorithms (SOM). SOM seems like an interesting choice for the present case as it is more novel than K-means clustering and rarely used in the area of eHealth interventions, but at the same time already quite established and shows good evaluations across several other domains (Vesanto & Alhoniemi, 2000). Furthermore, it seems to outperform k-means clustering for smaller datasets, which makes it very relevant for the evaluation in the present case (Abbas, 2008). In contrast to k-means clustering, SOM needs less data preparation which also makes it easier to apply (Guthikonda, 2005). Overall, SOM seems very promising and is expected to perform better than K-means when applied to smaller datasets.

Swarm Intelligence Based Clustering

Swarm intelligence-based clustering with particle swarm optimisation is another unsupervised learning method relevant to this paper. This method is the newest of the explored techniques and is inspired by the communicative behaviour of animals. A meta-analysis by Alam et al. (2014) found that swarm intelligence-based clustering with particle optimisation outperforms many other unsupervised learning methods, including K-means. The fact that it is rarely used in the eHealth environment, in combination with its seemingly good performance, makes it a good choice for exploration in the present paper. Similarly to SOM, swarm-based clustering is also easier to implement than most other clustering techniques as it needs very little preparation and is easily scalable, which allows applicability to a variety of situations (Alam et al., 2014). Especially for the quality of the created clusters, Swarm Intelligence-based clustering is expected to strongly outperform K-means clustering (Mangat, 2012). Little research directly compares SOM and Swarm Intelligence-based clustering. However, based on comparisons of both techniques to third clustering techniques, it is expected that they perform quite equally with a slightly better performance of Swarm Intelligence-based clustering (Abbas, 2008; Mangat, 2012).

Even though the general performance of unsupervised learning techniques can be estimated using literature, the actual quality of results, when applied to specific datasets, can vary strongly, depending on the specifics of each dataset (Bishop & Nasrabadi, 2006). Nevertheless, conclusions drawn in individual datasets can still be used as an orientation for methodological decisions made on a broader scale. Therefore, it appears to be a good first step to explore the methodological possibilities of unsupervised machine learning techniques on a specific dataset when trying to understand which techniques perform best on smaller samples. The dataset that will be used for this case study is the DIALECT dataset of the Diameter Project.

Current Study

Diameter Project

In order to help Type Two Diabetes (T2D) patients get more insights into improving their health, the Diameter application was developed by Hospital Group Twente (Ziekenhuisgroep Twente; ZGT), Roessingh Research & Development (RRD) and the

University of Twente. In this application, the patients can self-monitor their physical activity, diet and glucose values and receive coaching messages in relation to several variables, such as stage of behaviour change or behavioural goal. The intention of this is to help them get a better insight into their own behaviour and make the right changes to help improve their quality of life with T2D (Hietbrink et al., 2023).

The project evaluations show that patients' general perceptions of the Diameter application are very positive (Hietbrink et al., 2021). However, in line with prior research, this study also shows that the coaching messages seemed too generic and could be more personalised. In order to be able to continue the optimisation of the Diameter application for T2D patients, it is therefore important to understand how to make the best use of the gathered data. This would, in turn, allow better analysis of the health behaviour of people and determine which groups of people can be optimally supported by which coaching strategy or treatment.

As a first approach to achieving this, De Gooijer (2022) developed individual coaching strategies by identifying and analysing patients' physical activity and eating behaviour using the Diameter application. More specifically, patterns in the patient's data were identified using a combination of principal component analysis (PCA) and K-means clustering, and meaningful categories of patients with different types of lifestyles were created based on these patterns. These were for example three groups of patients that were either inactive, moderately active or very active or three groups that had either high intake of carbohydrates, salt or fat. These findings were used to evaluate the current coaching messages, which are part of the Diameter app, and it was found that there is indeed change needed (De Gooijer, 2022). However, due to a variety of missing values in the dataset, the number of data points that could be used was very limited, indicating weaker statistical power of the identified groups. Because of this, it appears that applying different unsupervised learning methods, for example, SOM and Swarm Intelligence-based clustering, to the dataset might lead to better results for the present dataset, which makes it a good fit for the present case study. Improving the clustering method used can support the development of better coaching strategies, which could benefit the health of patients in the long run.

This research will aim to explore how well different methodologies perform in identifying behavioural characteristics of type 2 diabetes patients from the DIALECT cohort with the goal of allowing the development of better-personalised coaching messages. This goal will be supported by the following research questions.

Research Question

How can unsupervised machine learning methods such as cluster analysis be optimally used to support the development of tailored coaching strategies for eHealth interventions based on their individual behavioural characteristics?

Sub Questions:

1. Which unsupervised machine learning method can best be used when trying to cluster patients from the DIALECT data set into separate groups based on data describing their behavioural characteristics related to T2D?
2. Which conclusions can be drawn from the present case study of T2D patients to support the development of personalised eHealth Interventions?

Methods

Design

This study used a combination of different statistical methods with the goal of finding which clustering method can best be used to find patterns in the behavioural data of patients of eHealth interventions, especially when the available amount of data is very limited.

Data

The dataset used for this study is an anonymised collection of specific data of all patients of the Diabetes and Lifestyle Cohort Twente (DIALECT). This data was collected in an observational form in two stages in the timeframe between 2009 to 2019. Between the two stages, the types of data that were collected varied slightly, which is why different variables vary in the amount of collected data (Den Braber et al., 2021). This was mainly caused by the research underlying the data collection advancing and, therefore, needing more and different types of data to answer developing research questions. Next to the two variables of the data that were used in this study, there was also additional data collected in the DIALECT study, like, for example, the glucose levels of the patients at different points in time. This data was excluded from the study as it was expected to complicate the process while not providing much value to answering the research question.

Variables

Physical Activity. The first variable that was measured was physical activity. This was done in the form of a step count that was tracked using a Fitbit step tracker. The steps performed by each patient were measured per minute across several days. The number of days varied between patients from 7 up to 28 days, while most patients were measured for seven days. Physical activity was tracked in both stages of the research, which is why it includes the highest amount of data with 278 patients.

Nutritional Intake. The next variable that was measured was Nutritional intake. This was measured through food diaries. These diaries consisted of the patients giving information about their dietary intake during the course of a month. Based on the results of the food diaries, the nutrients can be calculated. This leads to a multitude of nutritional variables, including for example, Energy, Protein, Carbohydrates, Fat, Fiber and Sodium (Oosterwijk et al., 2019). The intake of these variables is divided per meal and hour of the day. Based on problems in the formatting of the original data, only 68 out of 112 participants could be used for the analysis.

Participants

The patients included in the DIALECT dataset are adults diagnosed with T2D patients. However, not all adult T2D patients could participate as certain exclusion criteria were defined. Patients that were excluded from the study included patients with end-stage kidney disease, those who were not able to speak an adequate level of Dutch, and those who were not able to and did not want to give informed consent. No demographic or other descriptive data of the patients will be used in this study.

Procedure

This study was conducted in several steps. Firstly, the given DIALECT dataset was reorganised and prepared to allow thorough analysis and clustering. In the next step, k-means clustering, self-organising map algorithm and Swarm intelligence-based clustering were applied to all relevant variables of the dataset, and the results were compared based on cluster quality and interpretability for each cluster. Based on the results of this comparison, the best clustering method was identified, and the difference in outcomes was analysed.

Data Analysis

Version 4.2.1 of RStudion was used for the analysis (R Core Team, 2022). Specifically, the data was analysed and managed using the packages tidyverse(1.3.2, Wickham et al., 2019), ggplot2 (3.4.2, Wickham, 2016), readxl (1.4.2, Wickham & Bryan, 2023), dplyr (1.0.10, Wickham et al., 2022), factoextra (1.0.7, Kassambra & Mundt, 2020), cluster (2.1.4, Maechler et al., 2022), FactoMiner (2.8, Le et al., 2008), plotly (4.10.1, Sievert, 2020), kohonen (3.0.11, Wehrens & Kruisselbrink, 2018), DatabionicSwarm (1.1.6, Thrun, 2018), GeneralizedUmatrix (1.2.5, Thrun & Ultsch, 2020), rgl (1.1.3, Murdoc & Adler, 2023), purr (1.0.1, Wickham & Henry, 2023), parallelDist (0.2.6, Eckert, 2022), broom (1.0.1, Robinson et al., 2022) and ProjectionBasedClustering (1.1.8, Thrun & Ultsch, 2020). The data analysis done in this paper can be split into three steps data organisation, clustering and comparison. To allow replicability, the randomisation of R Studio was adjusted by using the "set.seed" function and "2023" was chosen as the seed. A full overview of the script used in the form of an RMarkdown document was also published on RPubs (Nyhoff, 2023).

Data Organisation

Before any analyses could be performed, the data had to be organised and formatted. The data was made available in the form of different Excel sheets. Each patient and each of the three measures of these patients were stored in different sheets. Firstly, all Excel sheets of the same type of measure were manually organised to have exactly the same format. After this was done, they were imported and directly combined into one dataset for each of the three factors physical activity, nutritional intake and glucose.

These datasets were then further organised individually. As a first step in the physical activity dataset, the measures per minute were summarised into blocks of six hours. The blocks described for every day, the timeframes 12 pm to 6 am, 6 am to 12 am, 12 am to 18 pm and 18 pm to 12 am and were called night, morning, afternoon and evening, respectively. In the next step, the mean and standard deviation were calculated for each patient and time of day across all days of measurement. Lastly, the data frame was pivoted into the wide format using the different times of the days as column names and mean, standard deviation and variance as values.

The values in the dataset regarding nutritional intake were firstly grouped into different blocks of the day and then summarised by the mean and standard deviation in the same way as in the physical activity dataset. Afterwards, this dataset was also pivoted using the time of day as column names and the mean and standard deviation of the different nutritional units as values.

In the last step, the datasets that were transformed to the wide format were also combined into bigger datasets. This was meant to allow comparison in clusters between those datasets that, for example, only contain the data of physical activity and those that contain multiple, multiple datasets.

Clustering

K-Means Clustering. The first necessary step when performing k-means clustering was to scale the data. This allows the comparison of data from the different measurements. Next, the Euclidean distance between observations was calculated using the distance function "dist()" of the factoextra package. After this, it was calculated how many clusters were needed to perform the k-means clustering, which was done using an elbow plot and silhouette width (Kodinariya & Makwana, 2013). After the number of clusters was estimated, the k-means clustering was performed on both versions of the dataset using the estimated number of clusters. In the last step, the clustering results were visualised.

Self-Organising Map Algorithms (SOM). SOM is a relatively new clustering technique strongly based on neural networks (Guthikonda, 2005). Similarly to K-means clustering, scaling the data and calculating Euclidean distances is also the first step that needs to be performed when doing SOM. After this, a grid of nodes used for the SOM was created. This was done using varying grid sizes and a hexagonal grid shape, depending on what was fitting for a dataset of this size (Rojas et al., 2015). In the next step, the grid model was trained. This is done through an iterative process in which each data point is assigned to the grid slot that is most similar to it based on Euclidean distance. The weight of the chosen grid slot and its neighbours are then changed in relation to the newly added input, with decreasing changes the further the grid slot is away.

Swarm Intelligence-Based Clustering. The third method that was used was swarm intelligence-based clustering (Celebi & Aydin, 2016). After scaling the data, a distance matrix of a specific set of cases was defined, and a projection of the data was created based on these distances. Then a generalised u-matrix was calculated using the created projections. This u-matrix included information about how similar, different points of data are. Each unit shared this u-matrix information with its neighbouring units and then made movement decisions based on the gathered information. In this way, all units move across the data space until the most optimal clusters are identified. The optimal number of clusters that should be created in this way was prior evaluated using the silhouette width in the same way as in k-means clustering. Lastly, topographic visualisations of the cluster division were created, allowing a very intuitive understanding of identified clusters (Thrun, 2023). In this visualisation, the mountains and watersheds show borders between clusters, while valleys represent the clusters themselves (Thrun & Ultsch, 2020).

Comparison

To compare the results of the clustering techniques in a strategic way, the clusters were assessed using interpretability and cluster quality.

Interpretability. The first method used to identify the quality of the different methods was the interpretability of the results. This measure examines how usable and understandable the results of different methods are for experts. To approach this evaluation, anova or Kruskal Wallis tests were performed to identify to values of which clusters differ significantly. To determine the threshold for significance was used based on the number of variables compared (Weisstein, 2004). For those variables where a significant difference was found, the cluster with the significantly highest and lowest was identified. Interpretability was then evaluated by comparing the number of significant variables and relating the expression of variables to how important they seem based on the literature.

More precisely, the evaluation of the importance of specific nutrition suggested that especially carbohydrates, fibre, saturated fats, Sodium and alcohol are important variables to keep into account (Evert et al., 2014; Ley et al., 2014). For physical activity, no clear distinction can be drawn as it can only be said that high physical activity can be important to alleviate T2D, but not at which specific time of day this physical activity happens (Di Loreto et al., 2005).

Cluster Quality. The second method used to evaluate the quality of the results was cluster quality by means of silhouette scores. Silhouette scores can provide indications of how well different data points in one cluster are separated from each other. At the same time, it also measures how distinct different clusters are from each other. To do this in R Studio, the "silhouette" function was used. Lastly, the mean, standard deviation and variance of the silhouette scores were calculated to allow easy comparability between the clustering techniques. Silhouette scores can vary between -1 and 1. The closer these results are to 1, the better the clustering. Results that are close to -1 suggest the wrong classification.

Results

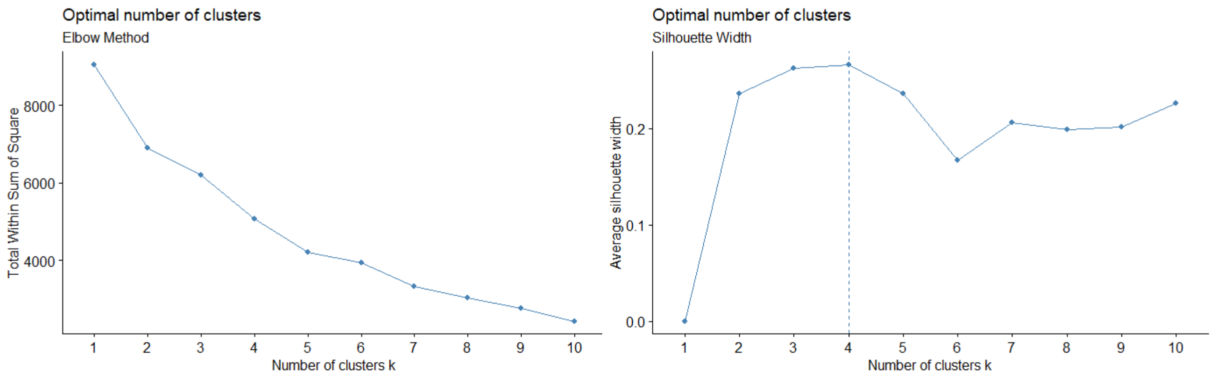
K-Means Clustering

Nutritional Intake

Based on the evaluation of elbow criterion and silhouette width (see Figure 1), it was decided that choosing four for the number of clusters would potentially lead to the best results.

Figure 1

Elbow Plot and Silhouette Width for K-Means Clustering of Nutritional Intake



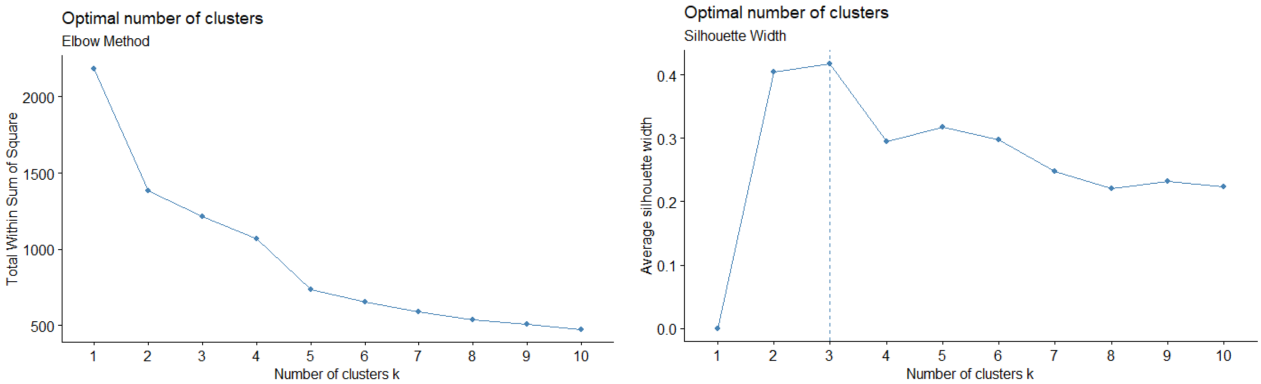
When performing k-means clustering, the data points patients were distributed to the different clusters, as seen in Appendix A.

Physical Activity

When judging the elbow criterion and silhouette width for the data frame of physical activity, three clusters seemed like the optimal option (see Figure 2).

Figure 2

Elbow Plot and Silhouette Width for K-Means Clustering of Physical Activity



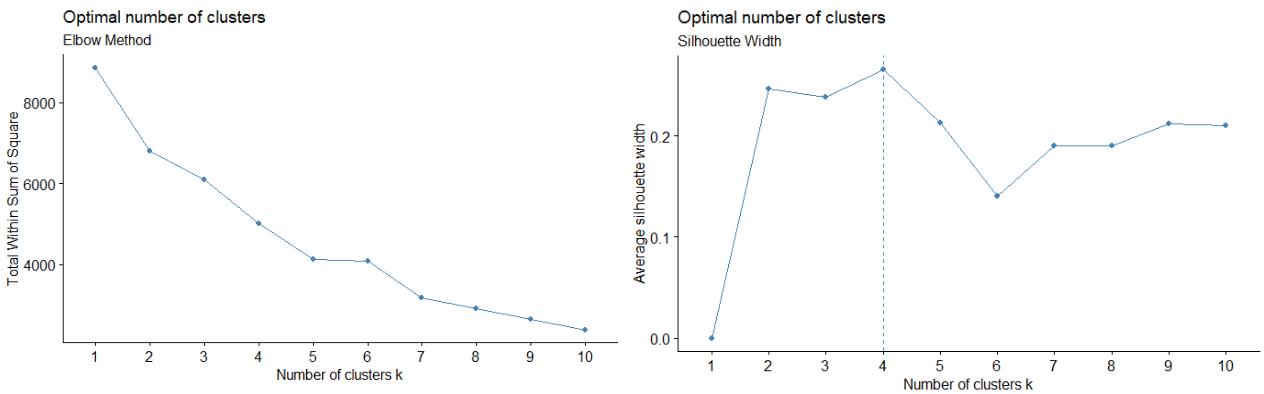
Performing k-means clustering using this number of clusters leads to the clustering distribution that can be seen in Appendix B.

Combination of Nutritional intake and Physical Activity

The elbow criterion and silhouette width for the combined dataset indicated that 4 is the optimal number of clusters (see Figure 3).

Figure 3

Elbow Plot and Silhouette Width for K-Means Clustering of Combined Dataset



When clustering the combined data frame using k-means and the given number of clusters, this led to the clustering division visible in Appendix C.

Self-Organising Map Algorithm

Nutritional Intake

When using the SOM on the data of nutritional intake, the observations were divided across the cells of the hex-map Appendix D. Based on the silhouette width estimated for the k-means clustering it was decided that four is a fitting number of clusters.

Physical Activity

The exact distribution of specific observations across the SOM-generated clusters can be seen in Appendix E. In line with prior analysis in the k-means clustering three was chosen like a fitting number of clusters.

Combination of Nutritional intake and Physical Activity

Lastly, the SOM model was also used to analyse the overall distribution of observations of the combined dataset across the clusters and the specific assignment to cluster in the hex map (see Appendix F). Four was chosen as a fitting number of clusters based on the analysis performed for k-means clustering.

Swarm Intelligence Based Clustering

Nutritional Intake

Based on the silhouette performed for the k-means clustering, it was decided that four should be chosen as the number of clusters and automatic clustering was performed. This clustering led to the division of observations across clusters, that is visible in Appendix G.

Physical Activity

Based on the silhouette width, it was decided to take three as an appropriate number of clusters. In the next step, the clustering of the data points into different groups was performed, which resulted in the division visible in Appendix H.

Combination of Nutritional intake and Physical Activity

Lastly, also the combination of nutritional intake and physical activity was analysed. Based on the silhouette width, it was decided that four is a good number of clusters, and automatic clustering was performed (see Appendix I).

Interpretability

Before the interpretability could be evaluated by identifying significant differences between the groups the significance thresholds had to be modified using the Bonferroni correction. Based on the number of variables the optimal thresholds were identified for the Nutritional Intake ($\alpha' = .000185$), Physical Activity ($\alpha' = .006$) and Combined ($\alpha' = .00018$)

dataset. Based on this, the number of variables with significant differences across clusters was evaluated across data frames (see Table 1).

Table 1

Number of Variables With Significant Differences Between Clusters

	Number of Significant Differences
Nutritional Intake K-means	86
Nutritional Intake SOM	78
Nutritional Intake Swarm Intelligence Based	43
Physical Activity K-means	8
Physical Activity SOM	8
Physical Activity Swarm Intelligence Based	6
Combined K-means	70
Combined SOM	77
Combined Swarm Intelligence Based	48

In the next step, it was analysed how the mean values of the significant variables are expressed across variables for all datasets related to Nutritional Intake (see Appendix J), Physical Activity (see Appendix K) and the combined dataset (see Appendix L). Specific patterns observable across clustering methods became directly visible here. One example for this is, that for Nutritional intake all clustering methods identified few to no variables with high expression, but a very high number of variables with low expression. Lastly, also the number of significant variables that are especially important based on the literature was evaluated (see Table 2).

Table 2

Number of Important Variables Across Methods and Datasets

	Carbohydrates	Fibre	Saturated Fats	Sodium	Alcohol
K-means Nutrition	2	3	4	3	2
SOM Nutrition	4	3	2	3	2
Swarm Nutrition	1	0	2	2	0
K-means Combined	3	3	2	2	2
SOM Combined	5	4	2	3	2
Swarm Combined	5	4	4	3	2

Cluster Quality

Finally, the cluster quality was evaluated using silhouette scores. The results in the form of mean, standard deviation and variance of silhouette scores can be seen in Table 3.

Table 3

Mean and Standard Deviation of Silhouette Scores Across Datasets and Methodologies

	Mean	Standard Deviation
Nutritional Intake K-means	-.004	.352
Nutritional Intake SOM	-.027	.411
Nutritional Intake Swarm Intelligence Based	.057	.332
Physical Activity K-means	.412	.232
Physical Activity SOM	.410	.236
Physical Activity Swarm Intelligence Based	.386	.275
Combined K-means	.091	.161
Combined SOM	.035	.090
Combined Swarm Intelligence Based	.034	.186

Discussion

Interpretation

This study found that none of the clustering techniques performed very well. Especially the Silhouette scores give a strong indication of the bad quality of the results. This is because all the scores are closer to zero, indicating no clustering than to one, which would indicate good clustering. Some of the results are even slightly negative, indicating wrong clustering. The only dataset for which the silhouette score results are somewhat better across all methods is the physical activity dataset. This appears to be in line with expectations based on the literature, as the main difference between this dataset and the other two datasets is that it has a significantly bigger sample size (Hahne et al., 2008).

When comparing the cluster quality results of the different unsupervised learning techniques to each other, making judgements becomes difficult. There are no clear trends visible in those results. It could potentially be argued that the overall results of k-means clustering are better. However, as this is not the case for the nutritional intake dataset and the results are very close to each other in general, this judgement should be made with care.

For interpretability, the results paint a somewhat similar picture. When comparing the number of variables with significant differences, it becomes apparent that the Swarm

Intelligence-based clustering leads to the smallest number of significant variables for all three datasets. At the same time, K-means clustering leads to the highest number of usable variables for nutritional intake, and SOM leads to the highest number for the combined dataset. For the physical activity dataset, the results of all techniques are quite similar, with Swarm Intelligence-based clustering having slightly less. Overall, it appears that Swarm Intelligence-based clustering performs worst at creating significantly different groups of the variables, while K-means and SOM perform quite equally. This is interesting to see as it is different to the results expected based on the literature, which suggests that Swarm Intelligence-based clustering outperforms K-means clustering and performs at least equal to SOM (Alam et al., 2014).

The evaluation of the number of significantly different important variables also shows comparable results. Here the results of the different methods are also very similar. These similarities make it hard to argue in favour of one of the methods. One point that is interesting to see is that the amount of significantly different groups identified by Swarm Intelligence-based clustering strongly changes between the Nutritional Intake dataset and the combined dataset. This is interesting as the combined dataset consists to a big extent of the Nutritional Intake dataset with only the Physical activity dataset added. However, Swarm Intelligence-based clustering also identifies fewer variables for Physical activity which makes it surprising to see that the interaction between those two datasets somehow allows Swarm Intelligence-based clustering to show better results. No logical explanations for this interaction appear obvious right away, and further research is necessary to get a better understanding of what is happening.

When trying to get a good understanding of the characteristics of individual patients in each group, it becomes quickly visible that making these judgments is not easy (see Appendix J, Appendix K, Appendix L). The high number of variables present for each cluster makes it hard to give clear descriptions of patients that are part of specific groups. To be able to make these statements, the most relevant variables need to be selected. This problem in understanding the specifics of the patients themselves also suggests that possibly the selection of specific important before conducting the analysis might lead to more interpretable results here. This, however, needs to be explored by further studies.

Implications

When trying to draw general implications from the results of this study, there are several points that need to be considered. Firstly, it has to be clearly stated that based on the results in the silhouette scores, none of the three chosen methods appears to be optimal for datasets of smaller sample sizes, which is in line with current research stating the sample size is potentially the most important predictor of clustering results (Von Luxburg & Ben-David, 2005). The main reason for this argument are the very low scores in cluster quality, combined with the fact that for the only dataset in this study that includes a bigger sample, these scores show significant improvements. This means in contrast to expectations, it cannot be said that SOM or Swarm Intelligence-based clustering perform generally better than K-means for smaller sample sizes (Mangat, 2012; Abbas, 2008). It also suggests that there might be additional factors next to sample size which influence the quality of the results of these methods.

As the evaluation criterion interpretability is not as commonly used by literature as cluster quality, it is harder to identify relations. Next to that, there are also no clear patterns in the distribution of significant important variables that allow good evaluation between methods. This is problematic and could suggest that the evaluation method of interpretability used in this paper might not be ideal. As interpretability is still believed to be an important factor for the evaluation of statistical methods, a different form of evaluation should be used for future studies.

In general, it can be said that all of the groups generated can be used to personalise coaching messages of patients, as they are of comparable quality as those currently used. However, there is still great room for improvement, and refining will definitely be needed to filter out variables that do not add much value, for example, Vitamin B6 consumption. It is also important to mention that the results of this paper need to be considered with care as there are several limitations present.

Limitations

Several limitations of the research of this paper must be named. Firstly, and most importantly, this paper can be seen as a case study using a very specific dataset and selected evaluation methodology. In order to be able to generate generalisable results, it is important to replicate the progress using multiple different datasets and compare the results. It could also be beneficial to add additional methods of evaluation in order to give more certainty to the results (Singh & Masuku, 2011).

In addition, it would also be important to compare more unsupervised learning methods. The amount of unsupervised learning methods existing is vast. The three methods used in this paper might be a good starting point but should not be seen as more than that. It is important to explore the results of other clustering techniques and compare those results to the present one to be able to find the best method for the DIALECT dataset. When using different methods, it would be suggested that especially density-based and hierarchical methods will be explored as those are not represented by the present study but show good results for smaller datasets (Abbas, 2008). Even though they were not considered for the present study as literature suggested slightly weaker performance than SOM and Swarm-Intelligence based clustering, it became evident that only direct evaluation on the specific dataset can really confirm these assumptions. Next to that, all methods in this study are based on Euclidean distances. As Euclidean distances appear to also have some weaknesses considering methods that are not based on them might also prove beneficial (Curriero, 2006).

Lastly, and also expected based on the research design of this paper, the generally small sample size, especially in the Nutritional Intake and Combined dataset, lead to problems with the quality of clustering results. Across unsupervised learning methods used, it was clearly visible that the bigger physical activity dataset showed better results for clustering quality. This is a strong suggestion that increasing the sample size improves the results to a certain extent. Taking these limitations in combination with the discussion of the results into account, a conclusion can be drawn.

Conclusion

Overall, K-means clustering appears to still be a better method that can be used to cluster patients as it shows better results regarding cluster quality than the other clustering techniques. Based on the literature, both other measures were expected to outperform K-means. When applied to a dataset, however, their performance was weaker than expected. These results make it quite prevalent that the specific characteristics of datasets have a very strong influence on the performance of clustering methods. This means to further aid the development of personalised eHealth interventions, more attention should be paid to identifying those exact characteristics of the used datasets that allow choosing the most appropriate method.

References

Abbas, O. A. (2008). Comparisons between data clustering algorithms. *International Arab Journal of Information Technology (IAJIT)*, *5*(3).

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Alam, S., Dobbie, G., Koh, Y. S., Riddle, P., & Ur Rehman, S. (2014). Research on particle swarm optimization based clustering: A systematic review of literature and techniques. *Swarm and Evolutionary Computation*, *17*, 1–13. https://doi.org/10.1016/j.swevo.2014.02.001

Arief, M., Hai, N., & Saranto, K. (2013). Barriers to and advantages of e-health from the perspective of elderly people: A literature review. *Finnish Journal of EHealth and EWelfare*, *5*, 50–56. http://www.finjehew.fi/index.php/stty/article/view/8172

Bilo, H. J. G., & Jong, M. (2003). Prevalence , incidence and mortality of type 2 diabetes mellitus revisited : A prospective populati ... *European Journal of Epidemiology*, *18*, 793–800.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, Issue 4). Springer.

Braber, N. Den, Laverman, G. D., Twente, Z., Science, C., & Science, C. (n.d.). *Physical activity profiles in patients with diabetes type 2*.

Buyl, R., Beogo, I., Fobelets, M., Deletroz, C., Van Landuyt, P., Dequanter, S., Gorus, E., Bourbonnais, A., Bourbonnais, A., Giguère, A., Giguère, A., Giguère, A., Lechasseur, K., Gagnon, M. P., Gagnon, M. P., & Gagnon, M. P. (2020). E-Health interventions for healthy aging: A systematic review. *Systematic Reviews*, *9*(1). https://doi.org/10.1186/s13643-020-01385-8

Celebi, M. E., & Aydin, K. (2016). Unsupervised learning algorithms. In *Unsupervised Learning Algorithms*. https://doi.org/10.1007/978-3-319-24211-8

Celis-Morales, C., Lara, J., & Mathers, J. C. (2015). Personalising nutritional guidance for more effective behaviour change. *Proceedings of the Nutrition Society*, *74*(2), 130–138.

Chevance, G., Perski, O., & Hekler, E. B. (2021). Innovative methods for observing and changing complex health behaviors: Four propositions. *Translational Behavioral Medicine*, *11*(2), 676–685.

Curriero, F. C. (2006). On the use of non-Euclidean distance measures in geostatistics. *Mathematical Geology*, *38*, 907–926.

Den Braber, N., Vollenbroek-Hutten, M. M. R., Westerik, K. M., Bakker, S. J. L., Navis, G., van Beijnum, B.-J. F., & Laverman, G. D. (2021). Glucose regulation beyond HbA1c in type 2 diabetes treated with insulin: real-world evidence from the DIALECT-2 cohort. *Diabetes Care*, *44*(10), 2238–2244.

Donath, M. Y., Ehses, J. A., Maedler, K., Schumann, D. M., Ellingsgaard, H., Eppler, E., & Reinecke, M. (2005). Mechanisms of β-cell death in type 2 diabetes. *Diabetes*, *54*(suppl_2), S108–S113.

Eckert, A. (2022). *parallelDist: Parallel Distance Matrix Computation using Multiple Threads*. https://cran.r-project.org/package=parallelDist

Eclaire Antoinette Gonnelien Hietbrink*, Niala den Braber, Anouk Middelweerd, Annemieke Konijnendijk, Miriam M.R. Vollenbroek-Hutten, G. D. L. (2021). A Pilot Study on the Diameter app: Lifestyle Support for Type 2 Diabetes Mellitus Patients. *A Pilot Study on the Diameter App: Lifestyle Support for Type 2 Diabetes Mellitus Patients*. https://research.utwente.nl/en/publications/a-pilot-study-on-the-diameter-app-lifestyle-support-for-type-2-di

Evert, A. B., Dennison, M., Gardner, C. D., Garvey, W. T., Hei, K., Lau, K., MacLeod, J., Mitri, J., Pereira, R., & Saslow, L. (2019). The American diabetes association. *Nutrition Therapy for Adults with Diabetes or Prediabetes: A Consensus Report*, 1–24.

Eysenbach, G. (2001). What is e-health? In *Journal of medical Internet research* (Vol. 3, Issue 2, p. E20). https://doi.org/10.2196/jmir.3.2.e20

Fisher, E. B., Fitzgibbon, M. L., Glasgow, R. E., Haire-Joshu, D., Hayman, L. L., Kaplan, R. M., Nanney, M. S., & Ockene, J. K. (2011). Behavior matters. *American Journal of Preventive Medicine*, *40*(5), e15–e30. https://doi.org/10.1016/j.amepre.2010.12.031

Fox, J. P., & Smink, W. A. C. (2021). Assessing an Alternative for "Negative Variance Components": A Gentle Introduction to Bayesian Covariance Structure Modeling for Negative Associations Among Patients With Personalized Treatments. *Psychological Methods*. https://doi.org/10.1037/met0000442

Freestyle Libre. (2023). *Je diabetes beter onder controle door uitstekende nauwkeurigheid*. https://www.freestyle.abbott/nl-nl/home.html

Garcia-Ortiz, L., Recio-Rodriguez, J. I., Agudo-Conde, C., Patino-Alonso, M. C., Maderuelo-Fernandez, J.-A., Repiso Gento, I., Puigdomenech Puig, E., Gonzalez-Viejo, N., Arietaleanizbeaskoa, M. S., Schmolling-Guinovart, Y., Gomez-Marcos, M. A., & Rodriguez-Sanchez, E. (2018). Long-Term Effectiveness of a Smartphone App for Improving Healthy Lifestyles in General Population in Primary Care: Randomized

Controlled Trial (Evident II Study). *JMIR Mhealth Uhealth*, *6*(4), e107. https://doi.org/10.2196/mhealth.9218

Gooijer, A. De. (2022). *Personalize coaching strategies within the Diameter using phenotypes based on physical activity , nutritional intake , and glucose levels* . 0–48.

Granja, C., Janssen, W., & Johansen, M. A. (2018). Factors determining the success and failure of ehealth interventions: Systematic review of the literature. *Journal of Medical Internet Research*, *20*(5), 1–21. https://doi.org/10.2196/10235

Guthikonda, S. M. (2005). Kohonen self-organizing maps. *Wittenberg University*, *98*.

Hahne, F., Huber, W., Gentleman, R., Falcon, S., Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning. *Bioconductor Case Studies*, 137–157.

Hietbrink, E. A. G., Middelweerd, A., van Empelen, P., Preuhs, K., Konijnendijk, A. A. J., Oude Nijeweme-d'Hollosy, W., Schrijver, L. K., Laverman, G. D., & Vollenbroek-Hutten, M. M. R. (2023). A Digital Lifestyle Coach (E-Supporter 1.0) to Support People With Type 2 Diabetes: Participatory Development Study. *JMIR Human Factors*, *10*, e40017. https://doi.org/10.2196/40017

HRSA. (2023). *What is telehealth?* https://telehealth.hhs.gov/patients/understanding-telehealth

IBM. (2023). *What is unsupervised learning?* https://www.ibm.com/topics/unsupervised-learning

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. https://cran.r-project.org/package=factoextra

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90–95.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining of cluster in K-means. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(6), 90–95. https://www.researchgate.net/publication/313554124

Lau, Y., Chee, D. G. H., Chow, X. P., Cheng, L. J., & Wong, S. N. (2020). Personalised eHealth interventions in adults with overweight and obesity: A systematic review and meta-analysis of randomised controlled trials. *Preventive Medicine*, *132*, 106001.

Lê, S., Josse, J., & Husson, F. (2008). {FactoMineR}: A Package for Multivariate Analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Lean, M. E. J., Leslie, W. S., Barnes, A. C., Brosnahan, N., Thom, G., McCombie, L., Peters, C., Zhyzhneuskaya, S., Al-Mrabeh, A., Hollingsworth, K. G., Rodrigues, A. M., Rehackova, L., Adamson, A. J., Sniehotta, F. F., Mathers, J. C., Ross, H. M., McIlvenna, Y., Stefanetti, R., Trenell, M., … Taylor, R. (2018). Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *The Lancet*, *391*(10120), 541–551. https://doi.org/https://doi.org/10.1016/S0140-6736(17)33102-1

Ley, S. H., Hamdy, O., Mohan, V., & Hu, F. B. (2014). Prevention and management of type 2 diabetes: dietary components and nutritional strategies. *The Lancet*, *383*(9933), 1999–2007.

Loke, A. (2022). *Diabetes*. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/diabetes

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). *cluster: Cluster Analysis Basics and Extensions*. https://cran.r-project.org/package=cluster

Mangat, V. (2012). Survey on particle swarm optimization based clustering analysis. *Swarm and Evolutionary Computation: International Symposia, SIDE 2012 and EC 2012, Held in Conjunction with ICAISC 2012, Zakopane, Poland, April 29-May 3, 2012. Proceedings*, 301–309.

McMahon, J., Thompson, D. R., Brazil, K., & Ski, C. F. (2022). An eHealth intervention (ManGuard) to reduce cardiovascular disease risk in male taxi drivers: protocol for a feasibility randomised controlled trial. *Pilot and Feasibility Studies*, *8*(1), 1–10. https://doi.org/10.1186/s40814-022-01163-4

Meershoek, A., Broek, L., & Crea-Arsenio, M. (2022). Perspectives from the Netherlands: Responses from, Strategies of and Challenges for Long-Term Care Health Personnel. *Healthcare Policy*, *17*.

Michael C. Thrun. (2023). *Short Intro to the Databionic Swarm (DBS)*. https://cran.r-project.org/web/packages/DatabionicSwarm/vignettes/DatabionicSwarm.html#3_Quality_Measures_of_Projection_and_Clustering

Moghimi, E., Davis, C., & Rotondi, M. (2021). The efficacy of eHealth interventions for the treatment of adults diagnosed with full or subthreshold binge eating disorder: systematic review and meta-analysis. *Journal of Medical Internet Research*, *23*(7), e17874.

Murdoch, D., & Adler, D. (2023). *rgl: 3D Visualization Using OpenGL*. https://cran.r-project.org/package=rgl

Nyhoff, T. (2023). *Personalised Coaching of Patients with Type 2 Diabetes Based on Their Individual Behavioural Characteristics*. RPubs. https://rpubs.com/TomNy/Thesis

Olafsdottir, A. F., Attvall, S., Sandgren, U., Dahlqvist, S., Pivodic, A., Skrtic, S., Theodorsson, E., & Lind, M. (2017). A clinical trial of the accuracy and treatment experience of the flash glucose monitor FreeStyle Libre in adults with type 1 diabetes. *Diabetes Technology & Therapeutics*, *19*(3), 164–172.

Oosterwijk, M. M., Soedamah-Muthu, S. S., Geleijnse, J. M., Bakker, S. J. L., Navis, G., Binnenmars, S. H., Gant, C. M., & Laverman, G. D. (2019). High dietary intake of vegetable protein is associated with lower prevalence of renal function impairment: results of the Dutch DIALECT-1 cohort. *Kidney International Reports*, *4*(5), 710–719.

Ried-Larsen, M., Johansen, M. Y., MacDonald, C. S., Hansen, K. B., Christensen, R., Wedell-Neergaard, A. S., Pilmark, N. S., Langberg, H., Vaag, A. A., Pedersen, B. K., & Karstoft, K. (2019). Type 2 diabetes remission 1 year after an intensive lifestyle intervention: A secondary analysis of a randomized clinical trial. *Diabetes, Obesity and Metabolism*, *21*(10), 2257–2266. https://doi.org/10.1111/dom.13802

Robinson, D., Hayes, A., & Couch, S. (2022). *broom: Convert Statistical Objects into Tidy Tibbles*. https://cran.r-project.org/package=broom

Rojas, I., Work-conference, I., Hutchison, D., Velázquez, M., Atencia, M., García, R., Pupo, D., Becerra, R., Veláquez, L., & Sandoval, F. (2015). *Advances in Computational Intelligence, Cluster Analysis of Finger-to-nose Test for Spinocerebellar Ataxia Assessment*.

Saran, R., Li, Y., Robinson, B., Ayanian, J., Balkrishnan, R., Bragg-Gresham, J., Chen, J. T. L., Cope, E., Gipson, D., He, K., Herman, W., Heung, M., Hirth, R. A., Jacobsen, S. S., Kalantar-Zadeh, K., Kovesdy, C. P., Leichtman, A. B., Lu, Y., Molnar, M. Z., … Abbott, K. C. (2015). US Renal Data System 2014 Annual Data Report: Epidemiology of Kidney Disease in the United States. In *American journal of kidney diseases : the official journal of the National Kidney Foundation* (Vol. 66, Issue 1 Suppl 1, p. Svii, S1-305). https://doi.org/10.1053/j.ajkd.2015.05.001

Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., & Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 127. https://doi.org/10.1186/s12966-016-0454-y

Searby, S. (2003). Personalisation - An overview of its use and potential. *BT Technology Journal*, *21*(1), 13–19. https://doi.org/10.1023/A:1022439824138

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC. https://plotly-r.com

Singh, A. S., Masuku, M. B., & Department. (2011). Sampling techniques & determination of sample size in applied statistics research. *Inwood Magazine*, *II*(96), 32–33.

Team, R. C. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Thrun, M. C., & Ultsch, A. (2020). Swarm intelligence for self-organized clustering. *IJCAI International Joint Conference on Artificial Intelligence*, *2021-Janua*(July), 5125–5129. https://doi.org/10.24963/ijcai.2020/706

Thrun, M. C. (2018). *Projection-Based Clustering through Self-Organization and Swarm Intelligence*. Springer. https://doi.org/10.1007/978-3-658-20540-9

Thrun, M. C., & Ultsch, A. (2020). Uncovering High-Dimensional Structures of Projections from Dimensionality Reduction Methods. *MethodsX*, *7*, 101093. https://doi.org/10.1016/j.mex.2020.101093

Thrun, M. C., & Ultsch, A. (2020). Using Projection based Clustering to Find Distance and Density based Clusters in High-Dimensional Data. *Journal of Classification*, *in press*. https://doi.org/10.1007/s00357-020-09373-2

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, *11*(3), 586–600. https://doi.org/10.1109/72.846731

Von Luxburg, U., & Ben-David, S. (2005). Towards a statistical theory of clustering. *Pascal Workshop on Statistics and Optimization of Clustering*, 20–26.

Watson, R., & Watson, R. (2014). Fernüberwachung. *50 Schlüsselideen Der Zukunft*, 92–95.

Wehrens, R., & Buydens, L. M. C. (2007). Self- and Super-Organizing Maps in {R}: The {kohonen} Package. *Journal of Statistical Software*, *21*(5), 1–19. https://doi.org/10.18637/jss.v021.i05

Weisstein, E. W. (2004). Bonferroni correction. *Https://Mathworld. Wolfram. Com/*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019).

Welcome to the {tidyverse}. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2023). *readxl: Read Excel Files*. https://cran.r-project.org/package=readxl

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation*. https://cran.r-project.org/package=dplyr

Wickham, H., & Henry, L. (2023). *purrr: Functional Programming Tools*. https://cran.r-project.org/package=purrr

Appendix A

Distribution of Patient Numbers Across clusters of Nutritional Intake Using K-Means

Cluster 1	Cluster 2	Cluster 3	Cluster 4
574	597	604	677
674	598	605	678
	601	607	697
	608	609	698
	615	611	711
	632	613	714
	633	619	717
	634	648	721
	636	649	727
	639	664	728
	642	696	741
	661	705	748
	669	710	760
	671	724	
	681	745	
	682	746	
	688	749	
	694	751	
	701	753	
	704	756	
	718	757	
	723		
	725		
	731		
	732		
	736		
	750		
	752		
	754		
	755		
	759		
	761		
	764		
	765		

Appendix B

Distribution of Patient Numbers Across clusters of Physical Activity Using K-Means

Cluster 1	Cluster 2	Cluster 3
588	361	1001
745	369	1002
751	371	353
761	380	364
	458	466
	470	473
	471	477
	476	479
	490	480
	497	482
	506	483
	509	485
	513	486
	552	487
	553	488
	558	489
	563	491
	567	504
	569	505
	574	507
	576	508
	577	510
	581	511
	582	512
	586	514
	591	515
	596	516
	600	517
	601	519
	604	520
	605	521
	607	522
	609	523
	610	524
	611	525
	612	527
	613	529
	615	530
	616	531
	618	532
	623	536
	625	537
	628	538
	629	540
	632	541
	634	542
	635	544

636	545
638	546
644	547
645	548
646	549
647	550
648	551
650	554
653	555
656	556
657	557
658	559
660	560
663	561
664	562
665	564
666	565
669	566
672	568
673	570
674	571
676	572
678	573
679	575
681	578
682	579
683	580
684	583
685	587
686	589
687	590
688	592
689	593
690	594
697	595
698	597
702	598
703	599
704	602
705	603
708	606
709	608
710	614
711	617
712	619
713	620
714	621
716	622
717	624
719	626
721	627

722	630
726	631
727	633
731	637
733	639
734	640
738	641
744	642
746	643
748	649
749	651
750	652
752	654
753	659
754	661
755	662
756	667
757	668
762	670
918	671
926	675
	677
	680
	691
	692
	693
	694
	695
	696
	699
	700
	701
	706
	707
	715
	718
	720
	723
	724
	725
	728
	729
	730
	732
	736
	737
	739
	741
	758
	759
	760

763
825
911
920
924
925

Appendix C

Distribution of Patient Numbers Across clusters of Physical Activity and Nutritional Intake Combined Using K-Means

Cluster 1	Cluster 2	Cluster 3	Cluster 4
598	751	597	574
601		615	674
604		632	677
605		633	678
607		634	711
608		636	717
609		639	721
611		642	727
613		661	741
619		669	760
648		671	
649		681	
664		688	
682		697	
694		701	
696		704	
698		714	
705		718	
710		723	
724		725	
745		728	
746		731	
748		732	
749		736	
753		750	
756		752	
757		754	
759		755	
		761	

Appendix D

Distribution of Patient Numbers Across clusters of Nutritional Intake

Cluster 1	Cluster 2	Cluster 3	Cluster 4
597	674	678	574
598		711	604
601		721	605
608			607
615			609
632			611
633			613
634			619
636			648
639			649
642			664
661			677
669			696
671			698
681			705
682			710
688			724
694			727
697			745
701			746
704			748
714			749
717			751
718			753
723			756
725			757
728			760
731			
732			
736			
741			
750			
752			
754			
755			
759			
761			
764			
765			

Appendix E

Distribution of Patient Numbers Across clusters of Physical Activity

Cluster 1	Cluster 2	Cluster 3
1001	588	361
1002	745	369
353	751	371
364	761	380
466		458
473		470
477		471
479		476
480		490
482		497
483		506
485		509
486		513
487		552
488		553
489		558
491		563
504		567
505		569
507		574
508		576
510		577
511		581
512		582
514		586
515		591
516		596
517		600
519		601
520		604
521		605
522		606
523		607
524		609
525		610
527		611
529		612
530		613
531		615
532		616
536		618
537		621
538		623
540		625
541		628
542		629
544		632

545	634
546	635
547	636
548	638
549	644
550	645
551	646
554	647
555	648
556	650
557	653
559	656
560	657
561	658
562	660
564	663
565	664
566	665
568	666
570	669
571	672
572	673
573	674
575	676
578	678
579	679
580	681
583	682
587	683
589	684
590	685
592	686
593	687
594	688
595	689
597	690
598	697
599	698
602	702
603	703
608	704
614	705
617	708
619	709
620	710
622	711
624	712
626	713
627	714
630	716
631	717

633	719
637	721
639	722
640	726
641	727
642	731
643	733
649	734
651	738
652	744
654	746
659	748
661	749
662	750
667	752
668	753
670	754
671	755
675	756
677	757
680	762
691	918
692	926
693	
694	
695	
696	
699	
700	
701	
706	
707	
715	
718	
720	
723	
724	
725	
728	
729	
730	
732	
736	
737	
739	
741	
758	
759	
760	
763	
825	

911
920
924
925

Appendix F

Distribution of Patient Numbers Across clusters of Nutritional Intake and Physical Activity Combined

Cluster 1	Cluster 2	Cluster 3	Cluster 4
678	597	633	574
711	598	718	601
721	608	736	604
	613		605
	615		607
	632		609
	634		611
	636		619
	639		648
	642		649
	661		664
	669		677
	671		696
	674		698
	681		705
	682		710
	688		724
	694		727
	697		745
	701		746
	704		748
	714		749
	717		751
	723		753
	725		756
	728		757
	731		760
	732		
	741		
	750		
	752		
	754		
	755		
	759		
	761		

Appendix G

Distribution of Patient Numbers Across clusters of Nutritional Intake

1	2	3	4
574	597	601	604
605	598	607	609
674	608	677	619
711	611	721	664
749	613	727	696
	615	741	698
	632	746	705
	633	760	745
	634		751
	636		
	639		
	642		
	648		
	649		
	661		
	669		
	671		
	678		
	681		
	682		
	688		
	694		
	697		
	701		
	704		
	710		
	714		
	717		
	718		
	723		
	724		
	725		
	728		
	731		
	732		
	736		
	748		
	750		
	752		
	753		
	754		
	755		
	756		
	757		
	759		
	761		
	764		

765	

Appendix H

Distribution of Patient Numbers Across clusters of Physical Activity

1	2	3
1001	361	588
1002	369	
353	371	
364	380	
476	458	
477	466	
479	470	
480	471	
482	473	
483	490	
485	497	
486	506	
487	509	
488	513	
489	552	
491	553	
504	558	
505	563	
507	567	
508	569	
510	572	
511	574	
512	576	
514	577	
515	581	
516	582	
517	586	
519	587	
520	591	
521	595	
522	596	
523	600	
524	601	
525	604	
527	605	
529	606	
530	607	
531	609	
532	610	
536	611	
537	612	
538	613	
540	614	
541	615	
542	616	
544	618	
545	621	

546	623
547	625
548	628
549	629
550	632
551	634
554	635
555	636
556	638
557	644
559	645
560	646
561	647
562	648
564	650
565	653
566	656
568	657
570	658
571	660
573	663
575	664
578	665
579	666
580	667
583	669
589	672
590	673
592	674
593	676
594	678
597	679
598	681
599	682
602	683
603	684
608	685
617	686
619	687
620	688
622	689
624	690
626	
627	
630	
631	
633	
637	
639	
640	
641	

642
643
649
651
652
654
659
661
662
668
670
671
675
677
680
691
692
694
693

Appendix I

Distribution of Patient Numbers Across clusters of Nutritional Intake and Physical Activity Combined

1	2	3	4
574	597	604	632
598	605	674	633
601	608	677	636
607	634	678	639
609	642	682	661
611	648	705	669
613	710	711	671
615	727	714	681
619	745	721	688
649	749	746	694
664	751		697
696	753		701
698	755		704
717	756		718
724	759		723
748	760		725
757			728
			731
			732
			736
			741
			750
			752
			754
			761

Appendix J

Expression of Significant Variables of Nutritional Intake Across Methodologies

K-Means	High Expression	Low Expression
Cluster 1	Energie(kcal)_mean_Nacht Energie(kcal)_sd_Abend Energie(kcal)_sd_Nacht Magnesium(mg)_sd_Abend Magnesium(mg)_sd_Nacht Nicotinezuur(mg)_sd_Abend	
Cluster 2		Calcium(mg)_mean_Morgen Calcium(mg)_sd_Morgen Eiwit(g)_mean_Morgen Eiwit(g)_sd_Abend Eiwit(g)_sd_Morgen Energie(kcal)_mean_Abend Energie(kcal)_mean_Morgen Energie(kcal)_sd_Abend Energie(kcal)_sd_Morgen Energie(kcal)_sd_Nacht Energie(kcal)_sd_Vormittag Foliumzuur(µg)_mean_Morgen Foliumzuur(µg)_sd_Vormittag IJzer(mg)_mean_Abend IJzer(mg)_mean_Morgen IJzer(mg)_sd_Abend IJzer(mg)_sd_Morgen IJzer(mg)_sd_Nacht IJzer(mg)_sd_Vormittag Jodium(µg)_mean_Morgen Jodium(µg)_sd_Morgen Kalium(mg)_mean_Abend Kalium(mg)_mean_Morgen Kalium(mg)_sd_Morgen Kalium(mg)_sd_Vormittag Koolhydr(g)_mean_Morgen Koolhydr(g)_sd_Morgen Magnesium(mg)_mean_Abend Magnesium(mg)_mean_Morgen Magnesium(mg)_sd_Abend Magnesium(mg)_sd_Morgen Magnesium(mg)_sd_Nacht Magnesium(mg)_sd_Vormittag Natrium(mg)_mean_Abend Natrium(mg)_mean_Morgen Natrium(mg)_sd_Morgen Nicotinezuur(mg)_mean_Abend Nicotinezuur(mg)_mean_Morgen Nicotinezuur(mg)_sd_Abend Selenium(µg)_mean_Abend Selenium(µg)_mean_Morgen Selenium(µg)_mean_Nacht Selenium(µg)_sd_Morgen Selenium(µg)_sd_Nacht Verz.vet(g)_mean_Abend Verz.vet(g)_mean_Morgen

		Verz.vet(g)_sd_Abend Verz.vet(g)_sd_Morgen Verz.vet(g)_sd_Nacht Vet(g)_mean_Abend Vet(g)_mean_Morgen Vet(g)_sd_Abend Vet(g)_sd_Morgen Vet(g)_sd_Nacht Vezels(g)_mean_Morgen Vezels(g)_sd_Morgen Vezels(g)_sd_Vormittag Vit.A(µg)_mean_Morgen Vit.A(µg)_sd_Morgen Vit.B1(mg)_mean_Morgen Vit.B12(µg)_mean_Morgen Vit.B12(µg)_mean_Nacht Vit.B12(µg)_sd_Morgen Vit.B12(µg)_sd_Nacht Vit.B2(mg)_mean_Morgen Vit.B2(mg)_sd_Morgen Vit.B6(mg)_mean_Abend Vit.B6(mg)_mean_Morgen Vit.B6(mg)_sd_Abend Vit.D(µg)_mean_Morgen Vit.D(µg)_sd_Morgen Vit.E(mg)_mean_Abend Vit.E(mg)_mean_Morgen Vit.E(mg)_sd_Morgen Water(g)_mean_Morgen Water(g)_sd_Morgen Zink(mg)_mean_Abend Zink(mg)_mean_Morgen Zink(mg)_sd_Morgen Zink(mg)_sd_Nacht Zink(mg)_sd_Vormittag Zout(g)_mean_Abend Zout(g)_mean_Morgen Zout(g)_sd_Morgen
Cluster 3	Energie(kcal)_mean_Abend Energie(kcal)_mean_Morgen Energie(kcal)_sd_Morgen Energie(kcal)_sd_Vormittag Vet(g)_sd_Morgen Verz.vet(g)_mean_Abend Verz.vet(g)_sd_Morgen Koolhydr(g)_mean_Morgen Koolhydr(g)_sd_Morgen Eiwit(g)_mean_Morgen Eiwit(g)_sd_Morgen Vezels(g)_mean_Morgen Vezels(g)_sd_Morgen Zout(g)_mean_Abend Zout(g)_sd_Morgen Water(g)_mean_Morgen Natrium(mg)_mean_Abend Natrium(mg)_sd_Morgen	

	Kalium(mg)_mean_Abend	
	Kalium(mg)_mean_Morgen	
	Kalium(mg)_sd_Morgen	
	Kalium(mg)_sd_Vormittag	
	Calcium(mg)_mean_Morgen	
	Calcium(mg)_sd_Morgen	
	Magnesium(mg)_mean_Abend	
	Magnesium(mg)_mean_Morgen	
	Magnesium(mg)_sd_Morgen	
	IJzer(mg)_mean_Morgen	
	IJzer(mg)_sd_Abend	
	IJzer(mg)_sd_Morgen	
	IJzer(mg)_sd_Vormittag	
	Selenium(µg)_mean_Morgen	
	Selenium(µg)_mean_Nacht	
	Selenium(µg)_sd_Morgen	
	Selenium(µg)_sd_Nacht	
	Zink(mg)_mean_Morgen	
	Zink(mg)_sd_Morgen	
	Zink(mg)_sd_Vormittag	
	Vit.A(µg)_mean_Morgen	
	Vit.A(µg)_sd_Morgen	
	Vit.D(µg)_mean_Morgen	
	Vit.D(µg)_sd_Morgen	
	Vit.E(mg)_mean_Abend	
	Vit.E(mg)_mean_Morgen	
	Vit.E(mg)_sd_Morgen	
	Vit.B1(mg)_mean_Morgen	
	Vit.B2(mg)_mean_Morgen	
	Vit.B2(mg)_sd_Morgen	
	Vit.B6(mg)_mean_Abend	
	Vit.B6(mg)_mean_Morgen	
	Vit.B6(mg)_sd_Abend	
	Foliumzuur(µg)_mean_Morgen	
	Foliumzuur(µg)_sd_Vormittag	
	Vit.B12(µg)_mean_Morgen	
	Vit.B12(µg)_sd_Morgen	
	Nicotinezuur(mg)_mean_Abend	
	Nicotinezuur(mg)_mean_Morgen	
	Jodium(µg)_mean_Morgen	
	Jodium(µg)_sd_Morgen	
Cluster 4	Vet(g)_mean_Abend	Energie(kcal)_mean_Nacht
	Vet(g)_mean_Morgen	Zink(mg)_mean_Nacht
	Vet(g)_sd_Abend	
	Vet(g)_sd_Nacht	
	Verz.vet(g)_mean_Morgen	
	Verz.vet(g)_sd_Abend	
	Verz.vet(g)_sd_Nacht	
	Eiwit(g)_sd_Abend	
	Vezels(g)_sd_Vormittag	
	Zout(g)_mean_Morgen	
	Water(g)_sd_Morgen	
	Natrium(mg)_mean_Morgen	
	Magnesium(mg)_sd_Vormittag	
	IJzer(mg)_mean_Abend	
	IJzer(mg)_sd_Nacht	

	Selenium(µg)_mean_Abend	
	Zink(mg)_mean_Abend	
	Zink(mg)_mean_Nacht	
	Zink(mg)_sd_Nacht	
	Vit.B12(µg)_mean_Nacht	
	Vit.B12(µg)_sd_Nacht	
SOM		
Cluster 1	Energie(kcal)_mean_Abend	Energie(kcal)_mean_Abend
	Energie(kcal)_mean_Morgen	Energie(kcal)_mean_Morgen
	Energie(kcal)_sd_Abend	Energie(kcal)_sd_Abend
		Energie(kcal)_sd_Morgen
		Energie(kcal)_sd_Vormittag
		Vet(g)_mean_Abend
		Vet(g)_mean_Morgen
		Vet(g)_sd_Morgen
		Verz.vet(g)_mean_Morgen
		Verz.vet(g)_sd_Morgen
		Koolhydr(g)_mean_Morgen
		Koolhydr(g)_sd_Abend
		Koolhydr(g)_sd_Morgen
		Koolhydr(g)_sd_Vormittag
		Eiwit(g)_mean_Morgen
		Eiwit(g)_sd_Morgen
		Vezels(g)_mean_Morgen
		Vezels(g)_sd_Morgen
		Vezels(g)_sd_Vormittag
		Zout(g)_mean_Morgen
		Zout(g)_sd_Morgen
		Zout(g)_sd_Vormittag
		Alcohol(g)_mean_Morgen
		Alcohol(g)_sd_Morgen
		Water(g)_mean_Morgen
		Water(g)_sd_Morgen
		Water(g)_sd_Vormittag
		Natrium(mg)_mean_Morgen
		Natrium(mg)_sd_Morgen
		Natrium(mg)_sd_Vormittag
		Kalium(mg)_mean_Morgen
		Kalium(mg)_sd_Morgen
		Kalium(mg)_sd_Vormittag
		Calcium(mg)_mean_Morgen
		Calcium(mg)_sd_Morgen
		Magnesium(mg)_mean_Abend
		Magnesium(mg)_mean_Morgen
		Magnesium(mg)_sd_Abend
		Magnesium(mg)_sd_Morgen
		Magnesium(mg)_sd_Vormittag
		IJzer(mg)_mean_Abend
		IJzer(mg)_mean_Morgen
		IJzer(mg)_sd_Abend
		IJzer(mg)_sd_Morgen
		IJzer(mg)_sd_Vormittag
		Selenium(µg)_mean_Morgen
		Selenium(µg)_sd_Morgen
		Selenium(µg)_sd_Vormittag
		Zink(mg)_mean_Abend

		Zink(mg)_mean_Morgen
		Zink(mg)_sd_Morgen
		Zink(mg)_sd_Vormittag
		Vit.A(µg)_mean_Morgen
		Vit.A(µg)_sd_Morgen
		Vit.D(µg)_mean_Morgen
		Vit.D(µg)_sd_Morgen
		Vit.E(mg)_mean_Abend
		Vit.E(mg)_mean_Morgen
		Vit.E(mg)_sd_Morgen
		Vit.B1(mg)_mean_Morgen
		Vit.B1(mg)_sd_Morgen
		Vit.B2(mg)_mean_Morgen
		Vit.B2(mg)_sd_Morgen
		Vit.B6(mg)_mean_Abend
		Vit.B6(mg)_mean_Morgen
		Vit.B6(mg)_sd_Abend
		Vit.B6(mg)_sd_Morgen
		Vit.B6(mg)_sd_Vormittag
		Foliumzuur(µg)_mean_Morgen
		Foliumzuur(µg)_sd_Morgen
		Foliumzuur(µg)_sd_Vormittag
		Vit.B12(µg)_mean_Morgen
		Vit.B12(µg)_sd_Morgen
		Nicotinezuur(mg)_mean_Abend
		Nicotinezuur(mg)_mean_Morgen
		Nicotinezuur(mg)_sd_Morgen
		Nicotinezuur(mg)_sd_Vormittag
		Jodium(µg)_mean_Morgen
Cluster 2	Energie(kcal)_sd_Morgen	
	Energie(kcal)_sd_Vormittag	
	Vet(g)_mean_Abend	
Cluster 3	Vet(g)_mean_Morgen	
	Vet(g)_sd_Morgen	
	Verz.vet(g)_mean_Morgen	
	Verz.vet(g)_sd_Morgen	
	Koolhydr(g)_mean_Morgen	
	Koolhydr(g)_sd_Abend	
Cluster 4	Koolhydr(g)_sd_Morgen	
	Koolhydr(g)_sd_Vormittag	
	Eiwit(g)_mean_Morgen	
	Eiwit(g)_sd_Morgen	
	Vezels(g)_mean_Morgen	
	Vezels(g)_sd_Morgen	
	Vezels(g)_sd_Vormittag	
	Zout(g)_mean_Morgen	
	Zout(g)_sd_Morgen	
	Zout(g)_sd_Vormittag	
	Alcohol(g)_mean_Morgen	
	Alcohol(g)_sd_Morgen	
	Water(g)_mean_Morgen	
	Water(g)_sd_Morgen	
	Water(g)_sd_Vormittag	
	Natrium(mg)_mean_Morgen	
	Natrium(mg)_sd_Morgen	
	Natrium(mg)_sd_Vormittag	

	Kalium(mg)_mean_Morgen
	Kalium(mg)_sd_Morgen
	Kalium(mg)_sd_Vormittag
	Calcium(mg)_mean_Morgen
	Calcium(mg)_sd_Morgen
	Magnesium(mg)_mean_Abend
	Magnesium(mg)_mean_Morgen
	Magnesium(mg)_sd_Abend
	Magnesium(mg)_sd_Morgen
	Magnesium(mg)_sd_Vormittag
	IJzer(mg)_mean_Abend
	IJzer(mg)_mean_Morgen
	IJzer(mg)_sd_Abend
	IJzer(mg)_sd_Morgen
	IJzer(mg)_sd_Vormittag
	Selenium(µg)_mean_Morgen
	Selenium(µg)_sd_Morgen
	Selenium(µg)_sd_Vormittag
	Zink(mg)_mean_Abend
	Zink(mg)_mean_Morgen
	Zink(mg)_sd_Morgen
	Zink(mg)_sd_Vormittag
	Vit.A(µg)_mean_Morgen
	Vit.A(µg)_sd_Morgen
	Vit.D(µg)_mean_Morgen
	Vit.D(µg)_sd_Morgen
	Vit.E(mg)_mean_Abend
	Vit.E(mg)_mean_Morgen
	Vit.E(mg)_sd_Morgen
	Vit.B1(mg)_mean_Morgen
	Vit.B1(mg)_sd_Morgen
	Vit.B2(mg)_mean_Morgen
	Vit.B2(mg)_sd_Morgen
	Vit.B6(mg)_mean_Abend
	Vit.B6(mg)_mean_Morgen
	Vit.B6(mg)_sd_Abend
	Vit.B6(mg)_sd_Morgen
	Vit.B6(mg)_sd_Vormittag
	Foliumzuur(µg)_mean_Morgen
	Foliumzuur(µg)_sd_Morgen
	Foliumzuur(µg)_sd_Vormittag
	Vit.B12(µg)_mean_Morgen
	Vit.B12(µg)_sd_Morgen
	Nicotinezuur(mg)_mean_Abend
	Nicotinezuur(mg)_mean_Morgen
	Nicotinezuur(mg)_sd_Morgen
	Nicotinezuur(mg)_sd_Vormittag
	Jodium(µg)_mean_Morgen
Swarm Based Clustering	
Cluster 1	Energie(kcal)_mean_Morgen
	Vet(g)_mean_Morgen
	Vet(g)_sd_Morgen
	Verz.vet(g)_mean_Morgen
	Verz.vet(g)_sd_Morgen
	Koolhydr(g)_mean_Morgen

	Eiwit(g)_sd_Morgen	
	Zout(g)_mean_Morgen	
	Zout(g)_sd_Morgen	
	Alcohol(g)_mean_Morgen	
	Alcohol(g)_sd_Morgen	
	Water(g)_mean_Morgen	
	Water(g)_sd_Morgen	
	Natrium(mg)_mean_Morgen	
	Natrium(mg)_sd_Morgen	
	Kalium(mg)_mean_Morgen	
	Kalium(mg)_sd_Morgen	
	Calcium(mg)_mean_Morgen	
	Calcium(mg)_sd_Morgen	
Cluster 2		Energie(kcal)_mean_Morgen
		Vet(g)_mean_Morgen
		Vet(g)_sd_Morgen
		Verz.vet(g)_mean_Morgen
		Verz.vet(g)_sd_Morgen
		Koolhydr(g)_mean_Morgen
		Eiwit(g)_sd_Morgen
		Zout(g)_mean_Morgen
		Zout(g)_sd_Morgen
		Alcohol(g)_mean_Morgen
		Alcohol(g)_sd_Morgen
		Water(g)_mean_Morgen
		Water(g)_sd_Morgen
		Natrium(mg)_mean_Morgen
		Natrium(mg)_sd_Morgen
		Kalium(mg)_mean_Morgen
		Kalium(mg)_sd_Morgen
		Calcium(mg)_mean_Morgen
		Calcium(mg)_sd_Morgen
		Magnesium(mg)_mean_Abend
		Magnesium(mg)_mean_Morgen
		Magnesium(mg)_sd_Morgen
		IJzer(mg)_mean_Morgen
		Selenium(µg)_mean_Morgen
		Zink(mg)_mean_Morgen
		Vit.A(µg)_mean_Morgen
		Vit.D(µg)_mean_Morgen
		Vit.D(µg)_sd_Morgen
		Vit.E(mg)_mean_Morgen
		Vit.B1(mg)_mean_Morgen
		Vit.B1(mg)_sd_Morgen
		Vit.B2(mg)_mean_Morgen
		Vit.B2(mg)_sd_Morgen
		Vit.B6(mg)_mean_Morgen
		Vit.B6(mg)_sd_Morgen
		Foliumzuur(µg)_mean_Morgen
		Foliumzuur(µg)_sd_Morgen
		Vit.B12(µg)_mean_Morgen
		Vit.B12(µg)_sd_Morgen
		Nicotinezuur(mg)_mean_Morgen
		Nicotinezuur(mg)_sd_Morgen
		Vit.C(mg)_mean_Morgen
		Vit.C(mg)_sd_Morgen

Cluster 3	Magnesium(mg)_mean_Abend
	Magnesium(mg)_mean_Morgen
	Magnesium(mg)_sd_Morgen
	IJzer(mg)_mean_Morgen
	Selenium(µg)_mean_Morgen
	Zink(mg)_mean_Morgen
	Vit.A(µg)_mean_Morgen
	Vit.D(µg)_mean_Morgen
Cluster 4	Vit.D(µg)_sd_Morgen
	Vit.E(mg)_mean_Morgen
	Vit.B1(mg)_mean_Morgen
	Vit.B1(mg)_sd_Morgen
	Vit.B2(mg)_mean_Morgen
	Vit.B2(mg)_sd_Morgen
	Vit.B6(mg)_mean_Morgen
	Vit.B6(mg)_sd_Morgen
	Foliumzuur(µg)_mean_Morgen
	Foliumzuur(µg)_sd_Morgen
	Vit.B12(µg)_mean_Morgen
	Vit.B12(µg)_sd_Morgen
	Nicotinezuur(mg)_mean_Morgen
	Nicotinezuur(mg)_sd_Morgen
	Vit.C(mg)_mean_Morgen
	Vit.C(mg)_sd_Morgen

Appendix K

Expression of Significant Variables of Physical Activity Across Methodologies

K-Means	High Expression	Low Expression
Cluster 1	meansteps_night meansteps_morning meansteps_afternoon meansteps_evening sdsteps_night	
Cluster 2	sdsteps_morning sdsteps_afternoon sdsteps_evening	
Cluster 3		meansteps_night meansteps_morning meansteps_afternoon meansteps_evening sdsteps_night sdsteps_morning sdsteps_afternoon sdsteps_evening
SOM		
Cluster 1		meansteps_night meansteps_morning meansteps_afternoon meansteps_evening sdsteps_night sdsteps_morning sdsteps_afternoon sdsteps_evening
Cluster 2	meansteps_night meansteps_morning meansteps_afternoon meansteps_evening sdsteps_night	
Cluster 3	sdsteps_morning sdsteps_afternoon sdsteps_evening	
Swarm Intelligence Based		
Cluster 1		sdsteps_afternoon meansteps_morning meansteps_afternoon meansteps_evening sdsteps_morning sdsteps_evening
Cluster 2	sdsteps_afternoon	
Cluster 3	meansteps_morning meansteps_afternoon meansteps_evening sdsteps_morning	

sdsteps_evening

Appendix L

Expression of Significant Variables of Combined Datasets Across Methodologies

K-Means	High Expression	Low Expression
Cluster 1	Energie(kcal)_sd_Morgen	Energie(kcal)_mean_Nacht
	Vet(g)_sd_Morgen	Energie(kcal)_sd_Nacht
	Verz.vet(g)_sd_Morgen	Kalium(mg)_mean_Morgen
	Koolhydr(g)_mean_Morgen	Calcium(mg)_sd_Morgen
	Koolhydr(g)_sd_Abend	Vit.B2(mg)_sd_Morgen
	Koolhydr(g)_sd_Morgen	
	Vezels(g)_sd_Morgen	
	Zout(g)_sd_Morgen	
	Alcohol(g)_mean_Morgen	
	Water(g)_sd_Vormittag	
	Natrium(mg)_mean_Morgen	
	Natrium(mg)_sd_Morgen	
	Kalium(mg)_sd_Morgen	
	Kalium(mg)_sd_Nacht	
	Kalium(mg)_sd_Vormittag	
	Calcium(mg)_mean_Morgen	
	Magnesium(mg)_mean_Nacht	
	IJzer(mg)_sd_Morgen	
	Selenium(µg)_mean_Abend	
	Selenium(µg)_mean_Morgen	
	Selenium(µg)_sd_Morgen	
	Zink(mg)_mean_Morgen	
	Zink(mg)_sd_Morgen	
	Vit.A(µg)_sd_Morgen	
	Vit.D(µg)_sd_Morgen	
	Vit.E(mg)_mean_Morgen	
	Vit.E(mg)_sd_Morgen	
	Vit.B1(mg)_mean_Morgen	
	Vit.B1(mg)_sd_Morgen	
	Vit.B1(mg)_sd_Vormittag	
	Vit.B2(mg)_mean_Morgen	
	Vit.B6(mg)_mean_Nacht	
	Vit.B6(mg)_sd_Abend	
	Vit.B6(mg)_sd_Morgen	
	Vit.B6(mg)_sd_Vormittag	
	Foliumzuur(µg)_sd_Morgen	
	Vit.B12(µg)_mean_Morgen	
	Vit.B12(µg)_sd_Morgen	
	Nicotinezuur(mg)_mean_Morgen	
	Nicotinezuur(mg)_sd_Morgen	
Cluster 2	Energie(kcal)_mean_Morgen	
	Eiwit(g)_mean_Morgen	
	Eiwit(g)_sd_Morgen	
	Vezels(g)_mean_Morgen	
	Magnesium(mg)_sd_Morgen	
	Magnesium(mg)_sd_Vormittag	
	Vit.A(µg)_mean_Morgen	

	Vit.D(µg)_mean_Morgen	
	Vit.B1(mg)_mean_Nacht	
	Vit.B6(mg)_mean_Morgen	
	Foliumzuur(µg)_mean_Morgen	
	Vit.B12(µg)_mean_Nacht	
Cluster 3		Energie(kcal)_sd_Morgen
		Vet(g)_sd_Morgen
		Verz.vet(g)_sd_Morgen
		Koolhydr(g)_mean_Morgen
		Koolhydr(g)_sd_Abend
		Koolhydr(g)_sd_Morgen
		Vezels(g)_sd_Morgen
		Zout(g)_sd_Morgen
		Alcohol(g)_mean_Morgen
		Water(g)_sd_Vormittag
		Natrium(mg)_mean_Morgen
		Natrium(mg)_sd_Morgen
		Kalium(mg)_sd_Morgen
		Kalium(mg)_sd_Nacht
		Kalium(mg)_sd_Vormittag
		Calcium(mg)_mean_Morgen
		Magnesium(mg)_mean_Nacht
		IJzer(mg)_sd_Morgen
		Selenium(µg)_mean_Abend
		Selenium(µg)_mean_Morgen
		Selenium(µg)_sd_Morgen
		Zink(mg)_mean_Morgen
		Zink(mg)_sd_Morgen
		Vit.A(µg)_sd_Morgen
		Vit.D(µg)_sd_Morgen
		Vit.E(mg)_mean_Morgen
		Vit.E(mg)_sd_Morgen
		Vit.B1(mg)_mean_Morgen
		Vit.B1(mg)_sd_Morgen
		Vit.B1(mg)_sd_Vormittag
		Vit.B2(mg)_mean_Morgen
		Vit.B6(mg)_mean_Nacht
		Vit.B6(mg)_sd_Abend
		Vit.B6(mg)_sd_Morgen
		Vit.B6(mg)_sd_Vormittag
		Foliumzuur(µg)_sd_Morgen
		Vit.B12(µg)_mean_Morgen
		Vit.B12(µg)_sd_Morgen
		Nicotinezuur(mg)_mean_Morgen
		Nicotinezuur(mg)_sd_Morgen
		Vit.C(mg)_mean_Morgen
		Energie(kcal)_mean_Morgen
		Eiwit(g)_mean_Morgen
		Eiwit(g)_sd_Morgen
		Vezels(g)_mean_Morgen
		Magnesium(mg)_sd_Morgen

		Magnesium(mg)_sd_Vormittag Vit.A(µg)_mean_Morgen Vit.D(µg)_mean_Morgen Vit.B1(mg)_mean_Nacht Vit.B6(mg)_mean_Morgen Foliumzuur(µg)_mean_Morgen Vit.B12(µg)_mean_Nacht Energie(kcal)_sd_Abend Vet(g)_mean_Morgen Verz.vet(g)_mean_Morgen Vezels(g)_sd_Vormittag Zout(g)_mean_Morgen Alcohol(g)_sd_Morgen Water(g)_mean_Morgen Water(g)_sd_Morgen Magnesium(mg)_mean_Morgen Magnesium(mg)_sd_Abend Foliumzuur(µg)_sd_Vormittag Nicotinezuur(mg)_mean_Abend
Cluster 4	Energie(kcal)_mean_Nacht Energie(kcal)_sd_Abend Energie(kcal)_sd_Nacht Vet(g)_mean_Morgen Verz.vet(g)_mean_Morgen Vezels(g)_sd_Vormittag Zout(g)_mean_Morgen Alcohol(g)_sd_Morgen Water(g)_mean_Morgen Water(g)_sd_Morgen Kalium(mg)_mean_Morgen Calcium(mg)_sd_Morgen Magnesium(mg)_mean_Morgen Magnesium(mg)_sd_Abend Vit.B2(mg)_sd_Morgen Foliumzuur(µg)_sd_Vormittag Nicotinezuur(mg)_mean_Abend	
SOM		
Cluster 1	Energie(kcal)_mean_Morgen Energie(kcal)_mean_Nachmittag Energie(kcal)_sd_Abend Energie(kcal)_sd_Morgen Vet(g)_mean_Morgen	Energie(kcal)_mean_Morgen Energie(kcal)_mean_Nachmittag Energie(kcal)_sd_Abend
Cluster 2	Vet(g)_sd_Morgen Verz.vet(g)_mean_Morgen Verz.vet(g)_sd_Morgen Koolhydr(g)_mean_Abend	Energie(kcal)_sd_Morgen Vet(g)_mean_Morgen Vet(g)_sd_Morgen Verz.vet(g)_mean_Morgen Verz.vet(g)_sd_Morgen Koolhydr(g)_mean_Abend Koolhydr(g)_mean_Morgen Koolhydr(g)_sd_Abend Koolhydr(g)_sd_Morgen

	Koolhydr(g)_sd_Vormittag
	Eiwit(g)_mean_Morgen
	Eiwit(g)_sd_Morgen
	Vezels(g)_mean_Morgen
	Vezels(g)_mean_Nachmittag
	Vezels(g)_sd_Morgen
	Vezels(g)_sd_Nachmittag
	Vezels(g)_sd_Vormittag
	Zout(g)_mean_Morgen
	Zout(g)_sd_Morgen
	Zout(g)_sd_Vormittag
	Alcohol(g)_mean_Morgen
	Alcohol(g)_sd_Morgen
	Water(g)_mean_Morgen
	Water(g)_sd_Morgen
	Water(g)_sd_Vormittag
	Natrium(mg)_mean_Morgen
	Natrium(mg)_sd_Morgen
	Natrium(mg)_sd_Vormittag
	Kalium(mg)_mean_Morgen
	Kalium(mg)_sd_Morgen
	Kalium(mg)_sd_Vormittag
	Calcium(mg)_mean_Morgen
	Calcium(mg)_sd_Morgen
	Magnesium(mg)_mean_Morgen
	Magnesium(mg)_mean_Nachmittag
	Magnesium(mg)_sd_Abend
	Magnesium(mg)_sd_Morgen
	Magnesium(mg)_sd_Vormittag
	IJzer(mg)_mean_Morgen
	IJzer(mg)_mean_Nachmittag
	IJzer(mg)_sd_Abend
	IJzer(mg)_sd_Morgen
	IJzer(mg)_sd_Vormittag
	Selenium(µg)_mean_Morgen
	Selenium(µg)_sd_Morgen
	Zink(mg)_mean_Abend
	Zink(mg)_mean_Morgen
	Zink(mg)_sd_Abend
	Zink(mg)_sd_Morgen
	Vit.A(µg)_mean_Morgen
	Vit.A(µg)_sd_Morgen
	Vit.D(µg)_mean_Morgen
	Vit.D(µg)_sd_Morgen
	Vit.E(mg)_mean_Morgen
	Vit.E(mg)_sd_Morgen
	Vit.B1(mg)_mean_Morgen
	Vit.B1(mg)_sd_Morgen
	Vit.B2(mg)_mean_Morgen
	Vit.B2(mg)_sd_Morgen
	Vit.B6(mg)_mean_Abend

		Vit.B6(mg)_mean_Morgen
		Vit.B6(mg)_sd_Abend
		Vit.B6(mg)_sd_Morgen
		Foliumzuur(µg)_mean_Morgen
Cluster 3		Foliumzuur(µg)_sd_Morgen
		Foliumzuur(µg)_sd_Vormittag
		Vit.B12(µg)_mean_Morgen
		Vit.B12(µg)_sd_Morgen
		Nicotinezuur(mg)_mean_Abend
		Nicotinezuur(mg)_mean_Morgen
		Nicotinezuur(mg)_mean_Nachmittag
		Nicotinezuur(mg)_sd_Abend
		Nicotinezuur(mg)_sd_Morgen
		Jodium(µg)_mean_Morgen
Cluster 4	Koolhydr(g)_mean_Morgen	
	Koolhydr(g)_sd_Abend	
	Koolhydr(g)_sd_Morgen	
	Koolhydr(g)_sd_Vormittag	
	Eiwit(g)_mean_Morgen	
	Eiwit(g)_sd_Morgen	
	Vezels(g)_mean_Morgen	
	Vezels(g)_mean_Nachmittag	
	Vezels(g)_sd_Morgen	
	Vezels(g)_sd_Nachmittag	
	Vezels(g)_sd_Vormittag	
	Zout(g)_mean_Morgen	
	Zout(g)_sd_Morgen	
	Zout(g)_sd_Vormittag	
	Alcohol(g)_mean_Morgen	
	Alcohol(g)_sd_Morgen	
	Water(g)_mean_Morgen	
	Water(g)_sd_Morgen	
	Water(g)_sd_Vormittag	
	Natrium(mg)_mean_Morgen	
	Natrium(mg)_sd_Morgen	
	Natrium(mg)_sd_Vormittag	
	Kalium(mg)_mean_Morgen	
	Kalium(mg)_sd_Morgen	
	Kalium(mg)_sd_Vormittag	
	Calcium(mg)_mean_Morgen	
	Calcium(mg)_sd_Morgen	
	Magnesium(mg)_mean_Morgen	
	Magnesium(mg)_mean_Nachmittag	
	Magnesium(mg)_sd_Abend	
	Magnesium(mg)_sd_Morgen	
	Magnesium(mg)_sd_Vormittag	
	IJzer(mg)_mean_Morgen	
	IJzer(mg)_mean_Nachmittag	
	IJzer(mg)_sd_Abend	
	IJzer(mg)_sd_Morgen	
	IJzer(mg)_sd_Vormittag	

	Selenium(µg)_mean_Morgen
	Selenium(µg)_sd_Morgen
	Zink(mg)_mean_Abend
	Zink(mg)_mean_Morgen
	Zink(mg)_sd_Abend
	Zink(mg)_sd_Morgen
	Vit.A(µg)_mean_Morgen
	Vit.A(µg)_sd_Morgen
	Vit.D(µg)_mean_Morgen
	Vit.D(µg)_sd_Morgen
	Vit.E(mg)_mean_Morgen
	Vit.E(mg)_sd_Morgen
	Vit.B1(mg)_mean_Morgen
	Vit.B1(mg)_sd_Morgen
	Vit.B2(mg)_mean_Morgen
	Vit.B2(mg)_sd_Morgen
	Vit.B6(mg)_mean_Abend
	Vit.B6(mg)_mean_Morgen
	Vit.B6(mg)_sd_Abend
	Vit.B6(mg)_sd_Morgen
	Foliumzuur(µg)_mean_Morgen
	Foliumzuur(µg)_sd_Morgen
	Foliumzuur(µg)_sd_Vormittag
	Vit.B12(µg)_mean_Morgen
	Vit.B12(µg)_sd_Morgen
	Nicotinezuur(mg)_mean_Abend
	Nicotinezuur(mg)_mean_Morgen
	Nicotinezuur(mg)_mean_Nachmittag
	Nicotinezuur(mg)_sd_Abend
	Nicotinezuur(mg)_sd_Morgen
	Jodium(µg)_mean_Morgen
Swarm Intelligence Based	
Cluster 1	
Cluster 2	
Cluster 3	
Cluster 4	