Predicting the Hours of Home Health Care Needed, based on Personal and Treatment Characteristics

Britt Zweers
STUDENT NUMBER: 2018817
Wordcount: 7961

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

Marijn van Wingerden
Mojtabi Rostami Kandroodi

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
June 2022

Preface

For this thesis, data was provided by EscuLine, a Business Intelligence and Healthcare Analytics company working for healthcare organizations in the Netherlands. Data was collected by Zorgbalans, a home healthcare organization in the Netherlands. I would like to thank EscuLine and Zorgbalans for their time and help. Furthermore, I would especially like to thank Dr. Daan Ooms and Dr. Marijn van Wingerden for their help and insights during the writing of this thesis.

Data Source/Code/Ethics Statement

Work on this thesis was based on a dataset provided by EscuLine, a Business Intelligence and Health Analytics company in the Netherlands: they received the dataset from a Dutch home health care organization. The health care organization collected the data in compliance with General Data Protection Regulation (GDPR). EscuLine signed a processing agreement with the home health care organization that allows them to process this data in order to provide insights. The data involves data from human participants and is anonymous. The original owner of the data and code used in this thesis retains ownership of the data and code during and after the completion of this thesis. The author of this thesis acknowledges that they do not have any legal claim to this data or code. The author of this thesis has evaluated her project according to the "Ethics checklist Student research with human participants". The code used for this thesis is only available to EscuLine, not to the general public. Images, tables and figures in this thesis were produced by the author. The code used in this thesis is partly based on the thesis written by Novakovic (2022) at Tilburg University on the same topic, as this thesis tried to improve the model she created.

Predicting the Hours of Home Health Care Needed, based on Personal and Treatment Characteristics

Britt Zweers

In this work the prediction of hours of home healthcare (HHC) needed in a week was investigated, based on personal and treatment characteristics known when a new client enters the HHC-system. The role of patient clusters in predicting hours of HHC was also evaluated. Previous research on predicting HHC hours is scarce: it is related to hours of care in a hospital setting, is not studied on an individual level or data on personal and treatment characteristics was not used. An addition to previous research would be that the model could serve as a second opinion next to the estimations of the nurse. Data was provided by EscuLine. It consisted of data on personal and treatment characteristics collected when a client entered the HHC-system. The target variable was the hours of HHC in the second week of the care plan. Results show that the model using Random Forest Regressor gave the best predictions (R^2 = 0.64), which performed better than the nurses' estimations of the HHC hours in that week (R^2 = 0.42). Patient clusters as created by EscuLine did not improve the model. It was concluded that predicting hours of HHC shows promising possibilities for the future.

1. Problem Statement & Research Goal

The goal of this thesis project is to create a machine learning model that predicts the hours of home health care (HHC) a person needs in a week, based on data on personal and treatment characteristics, which can serve as a second opinion next to the estimation of the nurse. In 2018, 589.000 people in the Netherlands received HHC, and this amount is increasing yearly (Vektis, 2020; Volksgezondheidenzorg, 2020). At the same time, the shortage of HHC providers (further referenced as nurses) is increasing. It is expected that at the end of 2031, there will be a shortage of 14.000 people working in HHC (Prognosemodel Zorg en Welzijn, 2022). The combination of a shortage of nurses and an increasing amount of HHC clients leads to an unsustainable situation with regards to HHC provision (National Health Services, 2022). The nurses are getting more and more overworked, which leads to higher burn-out rates, which increases the shortage of nurses even more (Rachel, Francesco, 2018). Therefore, it is important to improve the scheduling of nurses. This could lead to a more efficient use of their time, which can decrease the amount of job stress perceived (Vijayan, 2017), and can thus diminish burn-out rates among nurses. To improve the scheduling of nurses, proper predictions of HHC hours needed are necessary. Nurses estimate the needed hours of HHC when a client enters the HHC-system, but this is done subjectively, based on their own experience. These estimations are therefore prone to human error (Samima and Sarma, 2021). A more objective approach to predicting hours of HHC could therefore be of great value. Scientific research, however, on prediction of hours of HHC is scarce, especially when taking machine learning into account. Previous research on predicting hours of HHC was mostly related to hours of care in a hospital setting or wasn't studied on an individual level. Novakovic (2022) wrote her thesis at Tilburg University on predicting the amount of HHC hours an individual needs in a week, based on data from surveys, demographic data and previous care received. An addition to this would be a model that predicts hours of HHC needed at the beginning of the care period, with no historical information on HHC received: this model could serve as a second opinion next to the estimation of needed hours of HHC made by the nurse. Thus, in this thesis, data on personal and treatment characteristics and on patient clusters is added instead of using data on previous care.

From a scientific point of view, the additional data on personal characteristics, treatment characteristics and patient clusters are expected to improve the performance of the already existing model created by Novakovic. Patient clusters are groups of patients with similar needs of care: examples of clusters are patients with cardiovascular or cognitive problems. Using these as predictors and comparing different clusters, can add important information to the model by letting the cluster interact with the other predictors. Using extra data sources that have not been used before in predicting hours of HHC fills a gap in literature: it could also lead to more research on this topic in the future. From a societal point of view, the results of this thesis can contribute to an improved scheduling of HHC workers. Machine learning algorithms can do their predictions based on thousands of datapoints, if not more. The algorithms are expected to learn more complicated patterns than humans can and are expected to do better predictions than the estimations of the nurse. The predictions of the model could then serve as a second opinion next to the estimation of the nurse, which can improve the scheduling of the nurse. As mentioned above, this can diminish the level of job stress perceived, and thus burn-out rate (Vijayan, 2017). This is important considering the increasing shortage of nurses (Prognosemodel Zorg en Welzijn, 2022).

Thus, the research question of this thesis is as follows: *To what extent can a machine learning model, which predicts hours of HHC a person needs in a week based on data on personal and treatment characteristics, outperform the estimation of hours of HHC a person needs in a week made by the nurse, using R^2?* The research question will be addressed with two sub-questions, of which the first one focuses on which algorithm to use: *When comparing Linear Regression, Random Forest Regressor and Extreme Gradient Boost, which machine learning model leads to the best R^2, MAE and MSE?* The second sub-question will focus on the role of patient clusters in the performance of the model: *Does a model using patient-clusters outperform the general model with regard to the prediction of hours of HHC, using R^2?*

Results show that Random Forest Regressor Regressor performs the best in predicting hours of HHC, with an R^2 of 0.64, MAE of 3.26 and MSE of 29.59. This beats the baseline, which had an R^2 of 0.42, MAE of 5.13 and MSE of 117.94. This means that the machine learning model outperforms the baseline in predicting hours of HHC a person needs in a week. Patient clusters did not seem to improve the performance of the model. There was also no link found between these clusters and the hours of received HHC. Overall, it can be concluded that machine learning can aid in improving the predictions of HHC a person needs in a week, which could eventually lead to a diminished rate of burn-out among nurses.

2. Literature Review

As mentioned before, research on prediction of HHC hours a person needs in a week is scarce. However, research on similar topics has been done before. Hansen, Hansen, Alstrup and Lioma (2017) used different ensembles of Logistic Regression and Random Forest Regressor classifiers to predict large increases in the hours of HHC that citizens in Copenhagen receive in a month. They state that this achieved an area under the receiver performance curve of 0.715. The most important features in their model were the number of large increases in HHC hours in the last three months, hospitalization of the citizen, sick care received by the citizen, home care received during the weekend and the age of the citizen. They did not have access to data on the actual conditions of the citizens and state that this is expected to improve the performance of the model. Madsen et al. (2020) also predicted differences in hours of HHC that citizens receive, but they did this by predicting the probability of a large deviation between granted and delivered HHC, using Naïve Bayes and Tree-Augmented Naïve Bayes (TANB) models. Their data consisted of general information about the citizens and data on delivery of care each week, but they also had no access to specific data on patient- and treatment characteristics. They found high levels of accuracy (around 0.90), for which the TANB-model performed the best. Similar to Hansen, Hansen, Alstrup and Lioma (2017) they state that using new data sources is expected to improve the model. De Korte et al. (2020) created a classification model to predict home care utilization, using "high" and "low" as categories. They used Random Forest Regressor and Gradient Boosting models, both of which gave an accuracy of 0.79. Their regression approach led to an R^2 of 0.32, using Random Forest Regressor. De Korte et al. state that their results must be evaluated in other home healthcare settings as well, to measure the generalizability of their model. Furthermore, they mention that using clusters to create groups of relevant variables might improve the model (de Korte et al., 2020).

A study on a similar topic is the one from Bertsimas, Pauphilet, Stevens and Tandon (2021), who predicted length of stay of hospital patients, based on data from electronic health records. They found that Logistic Regression and Gradient Boosted Trees gave the best results. Furthermore, they emphasize that decision trees are more interpretable and therefore user-friendly for healthcare workers. Mekhaldi et al. (2020) also predicted length of stay of hospital patients, using Random Forest Regressor and Gradient Boosting. Both these algorithms performed comparably, with high levels of R^2 (>0.92) and low levels of mean absolute error (MAE) (<0.45). Important to note is that Mekhaldi et al. made use of an artificial dataset provided by Microsoft. A dataset provided by a healthcare organization might lead to more and different predictors, which could impact the performance of the model (Mekhaldi et al., 2020). Alsinglawi et al. (2020) predicted length of hospital stay for a cluster of cardiovascular patients. They used Gradient Boosting Regressor and Random Forest Regressor Regressor, and compared these models with a deep neural network (DNN). The machine learning models scored comparably on R^2 (around 0.81) and MAE (around 1.95), however the DNN scored lower on R^2 (0.77) and higher on MAE (2.30).

Recent literature thus shows that Logistic Regression, Gradient Boosting and Random Forest Regressor are algorithms that appear to be performing well: however, which algorithm performed best in these studies differed. In Novakovic's thesis written on the same topic, Linear Regression, Random Forest Regressor and Extreme Gradient Boost (XGBoost) models were used, where Linear Regression gave the best results (Novakovic, 2022). Therefore, in my thesis, Linear Regression, Random Forest Regressor and XGBoost models will be implemented and compared to see which algorithm performs the best in the prediction of HHC hours. Furthermore, this literature review shows that there is a gap in availability of data sources. Treatment characteristics are expected to improve the performance of the model (Disley, Roy and Smith, 1993; Badosa et al., 2017), but haven't been incorporated in models predicting hours of HHC using machine learning before. In my thesis, I will thus be using data on treatment characteristics and patient clusters (see Section Methodology Experimental Setup).

3. Methodology & Experimental Setup

In this chapter, the methodology and experimental setup of this thesis will be presented.

3.1 Description of home health care data

For this thesis, data was provided by EscuLine, a Business Intelligence and Healthcare Analytics company working for healthcare organizations in the Netherlands. The data was collected by nurses from Zorgbalans, a HHC organization in the Netherlands. The data consisted of several datasets, delivered as CSV-files. The datasets had to be linked and merged for usage. See Table 1 for a description of the raw datasets.

The data was derived from the Omaha System, a classification system used in HHC that records the health status, actions and measurements of a client (Omahasystem, 2021). The Omaha System consists of three components: (1) Problem Classification Scheme (42 problems based on four domains: environmental, psychosocial, physiological and health-related behavior domain), (2) Intervention Scheme (76 action areas based on four kind of actions: advise/instruct/supervise, treating and applying procedures, case management and monitoring) and (3) Problem Rating Scale for Outcomes (current and target scores, divided into three areas: knowledge, behavior and status, further referenced as 'Problem Rating Score'). See Appendix A for an extensive description of the Omaha System, including an overview of which Problem Areas and Intervention Schemes exist.

The data that was used, was collected at the start of the care plan. When a new client enters the HHC-system, the nurse creates a so-called 'care plan', which consists of info on which Problem Areas (and linked domains), Intervention Schemes (and linked categories) and Problem Rating Scores are relevant to the client. Furthermore, the nurse gives an estimation of the hours of HHC the client is likely to utilize. During the time that the client receives care, the nurse makes registrations of the hours of care the client has received. At so-called 'evaluation moments', the nurse evaluates the progress of the client with regards to the Problem Areas and updates the current Problem Rating Scores. The nurse can then also add new Problem Areas and Intervention Schemes to the care plan. Additionally, the client is assigned to a patient cluster. These nine clusters are derived from a KNN-model created by EscuLine and are based on the most dominant Problem Areas in that cluster: see Appendix B for an explanation of the cluster model and an overview of the possible patient clusters.

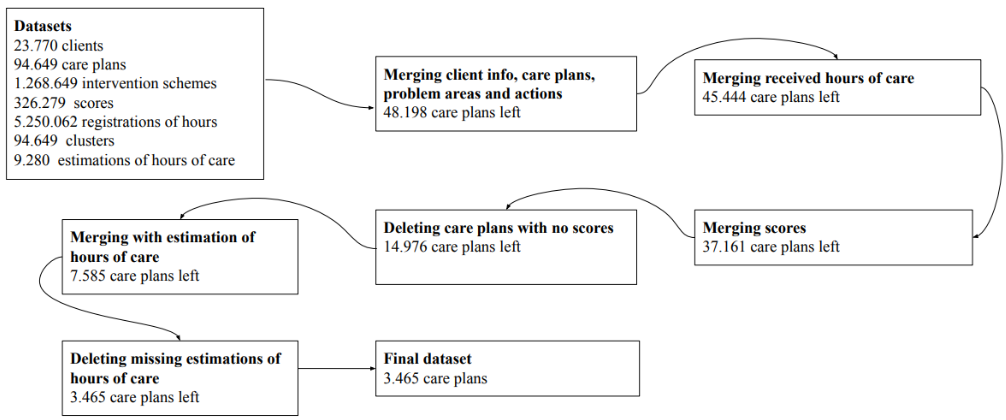
Table 1
Datasets and their descriptions

Dataset	Description
Demographic data	23.770 clients, linked to client ID
Care plans	94.649 care plans, linked to client ID
Intervention schemes	1.268.649 Intervention Schemes and Problem Areas, linked to care plan ID
Problem Rating Scale Scores	326.279 scores, linked to care plan ID
Minutes of care received	5.250.062 registrations of minutes of care, linked to client ID
Clusters	94.649 clusters, linked to care plan ID
Estimation of hours of care	9.280 estimations, linked to client ID

3.2 Data preprocessing

The datasets were merged before utilization, where each row in the final dataset represented a care plan. Each row in the data frame consisted of all Problem Areas and Intervention Schemes that were linked to the care plan in the time that the care plan was in place, even if some Problem Areas and Intervention Schemes were added later than at the start of the care plan: this choice was made as there was no data on when the Problem Areas and Intervention Schemes were added to the care plan. The same choice was made for the current Problem Rating Scores for Outcome: they had no time-stamp linked to them and there was only one current Problem Rating Score per Problem Area per care plan. It was assumed that this was the current score at the beginning of the care plan. Duplicate rows in each dataset were dropped and missing data was deleted, as in each case of missing data there was more than 50 % of data of that feature missing: missing data imputation would in this case lead to unreliable data (Jakobsen, Gluud, Wetterslev and Winkel, 2017). This was the case in missing Problem Rating Scores and missing estimations of hours of care: for some care plans, there was no data on scores and estimations, so these care plans were taken out as they were incomplete. Missing data imputation on these incomplete care plans was avoided as well, as this could lead to unreliable data. Clients of all ages were kept in the dataset, as this was preferred by EscuLine. One client could have multiple care plans that partly overlapped: however, if two care plans overlapped in the first two weeks, only the most recent one was kept. After merging all datasets together, there were n = 3.465 care plans left. See Figure 1 for the flowchart of the data preprocessing.

Figure 1
Flowchart of the data preprocessing



3.3 Feature extraction and exploratory data analysis

Multiple features were extracted to provide the model with relevant information. Information on marital status was transformed into a dummy for having a partner (value = 1) or no partner (value = 0): status 'unknown' was seen as having no partner. Extra features on age at the start of the care plan, amount of actions, amount of domains and amount of Problem Areas were added. The difference between current and target scores on the Problem Rating Scale for Outcomes was calculated, after which the current scores were discarded, as recommended by EscuLine: the difference between the current and target scores and the height of the target scores were expected to be more strongly related to hours of HHC than the current scores. Minutes of care received was transformed to hours of care, as to make the range of this variable smaller. The age of the client was transformed to age at the start of the care plan and was then standardized by subtracting the mean age. All variables on Problem Areas and Intervention Schemes were transformed to dummies for having the Problem Area/Intervention Scheme (value = 1) or not (value = 0). The feature on patient clusters was also transformed to ten dummies, for having the patient cluster (value = 1) or not (value = 0). See Appendix C for an overview of all the features in the final dataset.

After preprocessing the data, exploratory data analysis (EDA) was done. The characteristics on care moments per week and estimation of hours were calculated based on the subset of data for which this data was known. See Table 2 for the descriptive statistics of the clients in the dataset. As can be seen, the dataset consists of mostly elderly people. There are twice as many women in this dataset than men. This imbalance reflects the current elderly population: women grow older than men and utilize more health care than men (CBS, 2022). Furthermore, most clients are assigned to cluster 2, which is the patient cluster for 'other': there is no specific dominant Problem Area for this cluster.

Table 2
Descriptive statistics of the clients

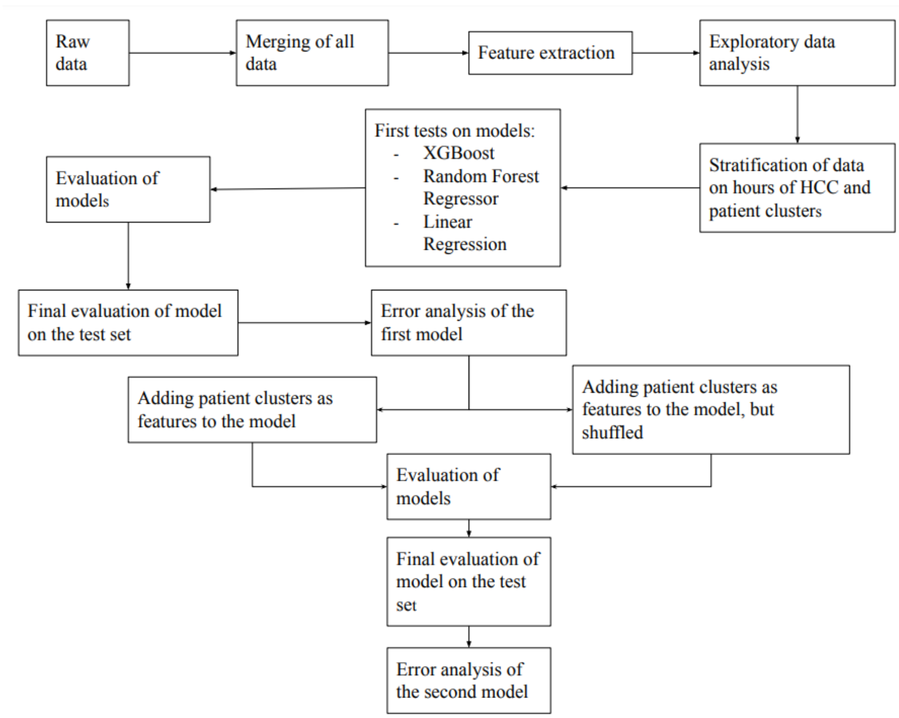
Characteristics	**Clients (*n* = *3.465*)**
Age (mean, SD)	80.25, 11.83
Gender (*n* (%))	
Male	1.248 (36.02)
Female	2.217 (63.98)
Marital status (*n* (%))	
Partner	632 (18.24)
No partner	2.833 (81.76)
Patient cluster (*n* (%))	
Cluster 0	319 (9.21)
Cluster 1	343 (9.89)
Cluster 2	1210 (34.9)
Cluster 3	194 (5.59)
Cluster 4	350 (10.10)
Cluster 5	325 (9.38)
Cluster 6	265 (7.65)
Cluster 7	49 (1.41)
Cluster 8	197 (5.69)
Cluster 9	213 (6.15)
Amount of Problem Areas (mean, SD)	3.97, 2.62
Estimation of HHC hours (mean, SD)	6.68, 10.77
Received HHC hours (mean, SD)	7.07, 9.23

3.4 Experimental procedure

The experiments that are described below, were done to answer the sub-questions: see Figure 2 for a visualization of the experimental approach. The target variable is the hours of care received in the second week of the care plan: the second week was chosen as the hours of care in the first week are not reliable as this depends on when in that week the client starts their first moment of care. Beforehand, the data was split into a train-set containing 80% of the data and a test-set containing 20% of the data. To ensure that the test-set was representative of the train-set, the test set was stratified on both hours of received HHC and patient clusters. This was done by binning the hours of received HHC from low to high, based on interquartile ranges (25% and 75% of the value), and by keeping the distribution of the patient clusters the same in these bins and thus across the train- and test set. This was done so that the train- and test set could be kept the same for both research questions, which makes it easier to compare the predictions.

The first sub question was answered by comparing the evaluation metrics of the algorithms after 5-fold cross-validation. The best model was chosen based on the evaluation metrics: R^2, MAE and MSE. After the best model was chosen, a final evaluation on the test set was done. Feature importance (either using Random Forest Regressor Feature Importance or SelectKBest, depending on which model was chosen) was evaluated. Error analysis was done by creating scatterplots and residual plots. For the second sub question, the best algorithm of the first sub question was chosen. The feature on patient clusters was added in two different ways: once by adding it normally and once by adding it randomly shuffled, so that the model can interact with the clusters but the clusters don't contain relevant information. The models were initialized and trained again. The evaluation metrics after 5-fold cross-validation were compared, after which the best model was chosen using R^2, MAE and MSE. The final evaluation of the model on the test set was done. Afterwards, error analysis on the model and on the individual patient clusters was done, again by creating scatterplots and residual plots.

Figure 2
Flowchart of the experimental approach



3.5 Algorithms

In this thesis, the performance of three different algorithms were compared, which were chosen based on literature review (see section: 3. Literature Review). The performance of Linear Regression, Random Forest Regressor and XGBoost will be evaluated. Below, the strong points of each of the three algorithms are described.

3.5.1 Linear Regression

Linear Regression is one of the most common and easy to interpret machine learning algorithms, which makes it a suitable algorithm for health care. In Linear Regression, the dependent variable is predicted with one or multiple predictors (Maulud and Abdulazeez, 2020). The prediction of the dependent variable is based on coefficients which are calculated for every independent variable. Linear Regression is the continuous variant of Logistic Regression, an algorithm that was successfully used in previous studies on similar topics (see section: 3. Literature Review). Linear Regression also gave the best predictions in the thesis by Novakovic on the same topic (Novakovic, 2022). It was thus hypothesized to perform well in this thesis.

3.5.2 Random Forest Regressor

Random Forest Regressor is a supervised machine learning algorithm that is often used in regression problems (Sruthi, 2021). In Random Forest Regressor, multiple decision trees are randomly created. Each tree is trained independently, which generates multiple evaluation metrics values. The Random Forest Regressor takes the average of the predictions of these trees (Graw, Wood, Phrampus, 2021). The strong point of Random Forest Regressor is that it has the possibility to use bootstrapping, an ensemble technique in which each model is generated using resamples of a dataset, where the samples are created using row sampling (Sruthi, 2021; Frost, 2022). Bootstrapping leads to an estimation of uncertainty, for which confidence intervals can be computed: this gives more information on the estimation of the Random Forest Regressor. Next to bootstrapping, the process of the Random Forest Regressor algorithm and its outputs are seen as easy to interpret for outsiders, as it can be visualized more easily (Antoniadis, Lambert-Lacroix Poggi, 2021). This makes Random Forest Regressor a suitable algorithm, as it allows for easy interpretation by the nurses.

3.5.3 XGBoost

XGBoost consists of a comparable method of generating decision trees to Random Forest Regressor. However, there is an important difference between the two algorithms. Random Forest Regressor creates multiple random trees in parallelization, whereas XGBoost does so in a sequential way. In doing so, errors of the previous tree are taken into account in the next tree by using a convex loss function and a penalty term (Mello, 2020). This leads to improved trees over time, which can lead to better predictions. Furthermore, XGBoost works well on structured, tabular data (Mello, 2020), and is therefore hypothesized to perform well in this thesis.

3.6 Hyperparameter tuning

Hyperparameter tuning for Random Forest Regressor and XGBoost was done with Randomized Search, where a fixed number of parameter settings is evaluated from a previously specified grid (ScikitLearn, 2022). The best hyperparameter settings were chosen as the optimal model. Randomized Search was chosen instead of Grid Search (in which all parameter settings are evaluated) as the computational costs of Grid Search were too high with regards to the external server of EscuLine. Furthermore, not all possible settings of XGBoost could be evaluated. The higher the parameters on amount of estimations and maximum depth of the tree, the higher the computational costs of running XGBoost. As all code for this thesis had to be run on the external server of EscuLine, the high costs of XGBoost led to practical problems with regards to speed of the server for the other employees of EscuLine. This is why the values chosen for the hyperparameters of XGBoost were lower. For Random Forest Regressor, for the hyperparameter 'bootstrap' only the option in which bootstrapping would be implemented was evaluated. This choice was made as the dataset is relatively small and bootstrapping would lead to the most optimal use of the dataset. For Linear Regression, both values of the only hyperparameter were evaluated. The hyperparameter tuning grids used for Randomized Search can be found in Appendix D, together with the hyperparameter settings that were used in the final models.

3.7 Evaluation criteria

In line with recommendations by EscuLine, R^2 was used as the main evaluation criterium. R^2 was chosen as this measurement is seen as easier to interpret, which is important with regards to the healthcare workers who will be working with the final model. Additionally, MAE and MSE were calculated. The values of these metrics were compared to the values of the baseline. The baseline used in this thesis was the estimation of HHC hours that a nurse does at the beginning of the care plan. These hours were already known: creating the baseline was thus done by calculating R^2, MAE and MSE between the estimated HHC hours of the nurse and the received HHC hours. A sidenote on the data on this baseline is that the estimated hours are not closely related to actual received hours of HHC. While in some cases this might be because of unexpected changes in need of HHC, it might also be because of practical mistakes in the estimations of hours. Even so, this baseline is relevant as it shows the downsides of predicting hours of HHC in practice: the registration of estimation of hours of HHC is done by humans and this can go wrong.

3.8 Implementation

All experimental procedures described above were performed using the 3.8.8. version of Python, on EscuLine's server. Libraries that were used during the experimental procedures are *ScikitLearn*, *XGBoost*, *Statistics*, *Matplotlib*, *Seaborn*, *Pandas* and *Numpy*.

4. Results

In this section, the performance of the models described in chapter 3.5 will be presented

4.1 Results first sub question

When comparing Linear Regression, Random Forest Regressor and Extreme Gradient Boost, which machine learning model leads to the best R^2 , MAE and MSE?

When comparing the evaluation metrics of Linear Regression, Random Forest Regressor and XGBoost, it is found that all three methods perform better than the baseline: see Table 3 for the R^2, MAE and MSE of the algorithms and baseline. XGBoost performed the best, whereas Linear Regression performed the worst. When comparing XGBoost and Random Forest Regressor, it is seen that XGBoost performed only slightly better than Random Forest Regressor. The computational costs of XGBoost, however, were so much higher than that of Random Forest Regressor that it significantly slowed down the external server of EscuLine. For example, in the training phase, the hyperparameter tuning of XGBoost took several days, during which the external server was also slow for other employees of EscuLine. This led to practical difficulties in working with XGBoost. This is why, even though the performance is a bit lower than that of XGBoost, Random Forest Regressor is chosen as the best model in predicting hours of HHC.

Table 3
Test results of the three algorithms for sub research question 1

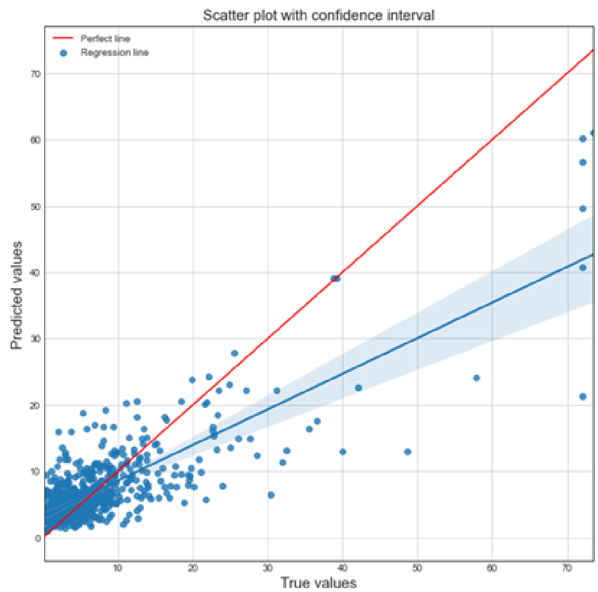
Algorithm	R^2	MAE	MSE
Baseline	0.42	5.13	117.94
Random Forest Regressor	0.64	3.26	29.59
XGBoost	0.66	3.19	27.40
Linear Regression	0.51	3.89	39.73

In Figure 3, the scatterplot of the received hours of HHC (further referenced as true values) vs. the predicted hours of HHC (further referenced as predicted values) is displayed. The red line is the line where true and predicted values are equal: the blue line shows the regression line fit on the true vs. predicted values, where the shaded area denotes the confidence interval. As can be seen, the confidence interval gets more broad the higher the estimated hours of HHC get: this shows that the model is more uncertain of high predictions than of lower predictions. This could be explained by having significantly less data on higher hours of care, which makes it more difficult for the model to learn these patterns. With more data on higher hours of care, the predictions are expected to improve. This statement is also supported by the scatterplots in Appendix E. In these scatterplots, the colors of the dots are dependent on the presence of various Intervention Schemes and Problem Areas. It can be seen that for rare Problem Areas and Intervention Schemes, the difference between true and predicted values is high. Examples of rare Problem Areas are ‘hearing’, ‘income/finance’ and ‘interpersonal relationships’ (Appendix E, Scatterplots E1-E5). Examples of rare Intervention Schemes are ‘finance’ and ‘specimen collection’. In contrast, for common Problem Areas and Intervention Schemes, the predictions are better. Examples are the Problem Areas ‘medication’, ‘nutrition’ and ‘personal care’, or the Intervention Schemes ‘continuity of care’, ‘skin care’ and ‘medication administration’ (Appendix E, Scatterplots E6-E11). Thus, it can be concluded that more data on these rare Problem Areas and Intervention schemes and more data on higher hours of HHC could improve the predictive performance of the model.

In Figure 3, an interesting pattern is also seen for the care plans which have true values of around 71 hours: these care plans all have 12 or more Problem Areas, which is only the case in 0.3% of care plans. The predictions for these care plans might be off as the model did not have enough data to learn the strong link between a high amount of Problem Areas and more than 70 hours of care per week. Furthermore, the model often predicts around 18 hours and 20-22 hours of HHC, while the true values of these care plans differ. When looking at these care plans, it is seen that they all have ‘medication’ and ‘personal care’ as Problem Areas and that the average amount of Problem Areas is a bit lower for the care plans which have predictions of 18 hours in comparison to the care plans which have predictions around 22 hours. It could be that the model predicts an average amount of HHC hours based on ‘medication’, ‘personal care’ and amount of Problem Areas, but has difficulties with distinguishing different care plans if there are no remarkable other patterns in their data.

Figure 3

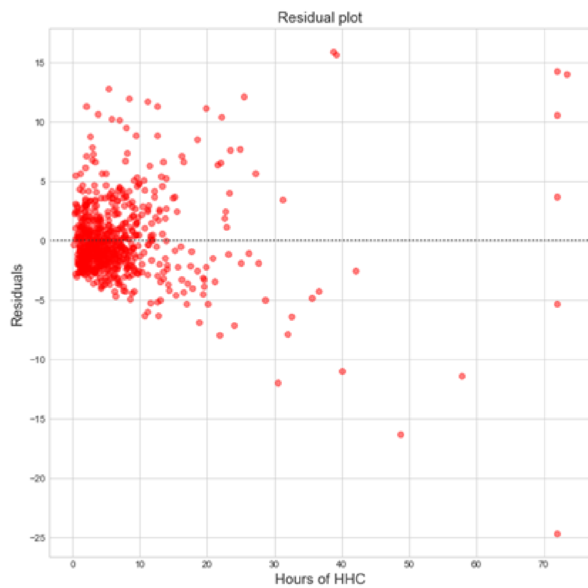
Scatterplot of the true vs. predicted hours of HHC with confidence interval



Furthermore, feature importance was calculated: see the graph in Appendix F. The most important features were, by far, amount of Problem Areas, amount of Intervention Schemes, age at the start of the care plan, target signal score 'sleep and resting patterns' and the Intervention Schemes 'care in the last moments of life', 'medication prescription' and 'supplies'. The target signal score 'sleep and resting patterns' is especially important in patients with dementia, a group of patients that get high hours of care. 'Care in the last moments of life' is a common Intervention Scheme for palliative care, which is also related to higher hours of care. Next to this, the scatterplots of Intervention Schemes 'Medication prescription' and 'supplies' vs. true and predicted values (Appendix E, Figure E.12 and E.13) also show that these are related to higher hours of care. It thus seems to be that the model gives more importance to features that are linked with higher hours of care.

In Figure 4, the residuals of each individual prediction is displayed. This shows the same as can be concluded from Figure 2: the higher the received hours of HHC, the bigger the residuals. The plot shows again that this could be linked to the amount of data there is: the more data, the lower the residuals. Some care plans which have a low amount of hours of HHC, so for which there is enough data, still have high residuals. Explanations for this could lie in the data. It might be that the nurse did not fill in the care plan in the right way, or that the nurse added comments to the care plans which were not part of my dataset.

Figure 4
Residual plot of the predicted hours of HHC



Overall, it can be concluded that Random Forest Regressor performs better than the baseline on predicting hours of HHC. The residuals increase the higher the true values go, where Random Forest Regressor has the tendency to give predictions that are too low. It is expected that, in the case of more data on higher hours of HHC, the model learns these patterns better and thus performs better.

4.2 Results second sub question

Does a model using patient-clusters outperform the general model with regard to the prediction of hours of HHC, using R^2 ?
After adding data on patient clusters to the model, it is found that the performance of the model decreases slightly in comparison to the model without data on patient clusters: see Table 4. The error increases slightly and the R2 decreases. After shuffling the data on the clusters, the performance of the model decreases, which shows that the clusters do contain information that is not completely random: otherwise the performance would have been exactly the same. The model without clusters and the model with clusters containing relevant information, perform almost exactly the same, however the model without clusters still performs the best.

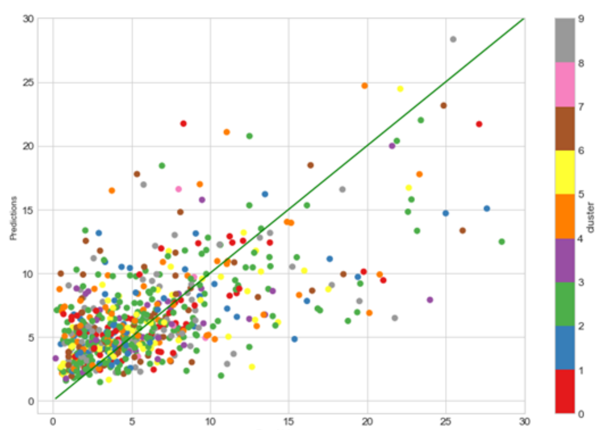
Table 4
Test results of models for the second sub question

Model	**R^2**	**MAE**	**MSE**
Baseline	0.42	5.13	117.94
Model without clusters	0.64	3.26	29.59
Model with clusters containing relevant information	0.63	3.27	29.81
Model with clusters containing irrelevant information	0.62	3.34	30.93

When comparing the feature importance for both models, the dummy feature for cluster 2 is seen as the most important cluster in both models. This could be because cluster 2 is the most common cluster in the dataset. In Figure 5, the scatterplot of true vs. predicted values for the model with clusters containing relevant information is displayed. The clusters are scattered across the plot, which shows that the clusters do not seem to be linked to hours of HHC. This is interesting, as the clusters have different dominant Problem Areas, that differ in mean hours of HHC. It would then be expected that the patient clusters are also linked to HHC hours. An explanation might lie in that patient clusters are created based on data on current Problem Rating Scores, and these features already exist in the model. It thus seems to be better to use the Scores and Problem Areas themselves than a construct of them. Furthermore, it seems that the clusters might not be that reliable yet, as the model with clusters containing irrelevant information performs almost equally well.

Figure 5

Scatterplot of true vs. predicted values, based on clusters



When looking at the individual clusters and its residuals, it is seen that some clusters are on average being over-predicted and some are under-predicted: see Table 5. The predictions of cluster 6 are the best, while the predictions of cluster 3 are the worst. The most dominant Problem Area of cluster 6 is 'skin', which is one of the most common Problem Areas: this could explain why this cluster performs the best, as the model had enough data on this Problem Area. For cluster 3 'cognition' is the most dominant Problem Area. This Problem Area is linked to patients in dementia, that get, as mentioned in research question 1, high hours of care. This is also seen in the relatively high mean hours of HHC for cluster 3, as shown in Table 5. The residuals of this cluster might be the highest because the cluster isn't developed enough yet, or because, as mentioned before, the construct of the cluster itself performs worse than the Problem Areas themselves. It is also interesting to see that the mean residual of cluster 2, which is the cluster for all clients whose Problem Areas do not fit the other clusters, is relatively low: this again shows that the clusters do not seem to be well-developed yet. Otherwise, the mean residual for cluster 2 would be expected to be higher, as this cluster is based on a group of random Problem Areas.

Table 5
Mean residuals and standard deviations and mean hours per cluster

Cluster (*n*)	Residuals	Mean hours of HHC
Cluster 0 (65)	-0.29	6.47
Cluster 1 (69)	0.31	7.76
Cluster 2 (243)	0.17	7.04
Cluster 3 (39)	0.73	8.99
Cluster 4 (70)	-0.51	6.93
Cluster 5 (65)	0.43	6.64
Cluster 6 (53)	-0.02	8.14
Cluster 7 (9)	-0.47	5.80
Cluster 8 (39)	-0.21	6.33
Cluster 9 (42)	0.60	7.47

Overall, it can be stated that this model does not outperform the general model with regard to the prediction of hours of HHC. Therefore, the best model in predicting hours of HHC is the Random Forest Regressor model without patient clusters.

5. Discussion

In this thesis, the prediction of hours of HHC using machine learning was studied, using data on personal and treatment characteristics. The goal was to create a model that leads to predictions of hours of HHC that could serve as a second opinion next to the nurse's estimation.

5.1 Discussion of sub-questions

The first sub-question dealt with choosing the best algorithm for predicting hours of HHC. It was found that XGBoost leads to the best results in predicting hours of HHC and Linear Regression performed the worst. Due to the high computational costs of XGBoost, however, Random Forest Regressor was chosen as the best model. This led to an R^2 of 0.64, which outperformed the baseline, that had an R^2 of 0.42. Errors in predictions were linked to low amounts of data: the higher the hours of HHC predicted, for which there were few care plans, the higher the uncertainty of the model. This was supported by the finding that the predictions for rare Problem Areas and Intervention Schemes had higher residuals than those of common Problem Areas and Intervention Schemes. Important features were also linked to higher hours of care. The results of the sub-question are partly in line with previous research. That Random Forest Regressor performed well, was expected and is in line with research by De Korte et al. (2020), for whom Random Forest Regressor also performed well. Mekhaldi et al. (2020) stated that for their study, on predicting length of hospital stay, Random Forest Regressor and Gradient Boosting performed comparably: this was also the case in this thesis. However, the bad performance of Linear Regression was not expected: in Novakovic's (2022) thesis on the same topic, Linear Regression was found to be the best algorithm. An explanation could lie in that in this thesis, much more data was used which led to more features. These features could have interacted with each other, which decreases the performance of linear regression. Furthermore, Linear Regression might have been too simple for this many features: it assumes that the dependent variable and predictors are linearly related to each other, but this doesn't have to be the case (Flom, 2022).

The next step in this research was the addition of patient clusters to the model. The results showed that the addition of patient clusters did not lead to better results. The patient clusters were also found not to be linked to hours of HHC. Furthermore, the model with patient clusters that contained irrelevant information performed comparably to the model with clusters that contained relevant information: this showed that the clusters might not be that reliable. These findings were not in line with the expectation. De Korte et al. (2020) stated that creating clusters of relevant variables might improve the model. Furthermore, the addition of these clusters were seen as extra data on personal characteristics, which was also expected to improve the performance of the model. That the results were not as expected, might be because of several reasons. For example, the patient clusters were created based on Problem Rating Scores of each Problem Area. If patient clusters were created based on both Problem Areas and the mean hours of HHC received per Problem Area was also taken into account in creating the clusters, the patient clusters might have performed better in predicting HHC hours. Furthermore, the patient clusters are still under development: clusters that are fully developed might perform better as they are expected to form a better construct than the current clusters. Also, the information that the patient clusters contain, is already part of the data: the clusters are based on the Problem Rating Scores, and these Scores are also features in the model. If the features on Problem Rating Scores were not part of the model, then addition of the patient clusters might have led to improvement of the predictions. This research question thus emphasizes the need of developing the patient clusters further before using them in predicting HHC hours.

In previous research (Hansen, Hansen, Alstrup and Lioma, 2017; Madsen et al., 2020), using more data sources was recommended to improve the performance of the model: this study showed that this indeed leads to improvement. The model also performed better than the estimation of the nurse, however it is important to be careful with these conclusions. The estimation of the nurse were not directly linked with the planned hours: the estimations of the nurse were done per activity of care, while some activities were registered in the received hours that were not part of the estimations. Therefore, with the current data, it cannot be concluded whether the model truly performs better than the estimations of the nurse, because the data on the estimations of the nurse was not always correct. The same could be said for the registration of hours of HHC received: these registrations were done by the nurses themselves, which raises the chances of the registrations containing human error. Thus, even after improving the model, it is important to emphasize the role of the model as a second opinion. In the end, the estimation of the nurse should be leading and not the estimation of the model, which should only serve as a second opinion.

5.2 Contribution to current literature

The results of this research contributes to current literature in several ways. This research adds to previous studies on HHC utilization, such as the one by De Korte et al. (2020), who stated that Random Forest Regressor was the best performing algorithm in predicting home care utilization. They stated that their results should be generalized in other home health care settings as well, which was done in this thesis. Furthermore, using extra data sources was a recommendation often done in previous studies (Hansen, Hansen, Alstrup and Lioma, 2017; Madsen et al., 2020). This study emphasized the value of using extra data as features, as it was one of the first studies using data on personal and treatment characteristics in predicting HHC hours. Overall, this study fills a bit of the gap in research on predicting HHC hours. However, this thesis had several limitations, which lead to recommendations for future research.

5.3 Limitations and recommendations for future research

This study only used data from one HHC organization, for which there wasn't much data available. In future research it is therefore recommended to combine data from several organizations, so that more data is available: this is expected to strengthen the model and make it more robust, as there would be more diverse data. Furthermore, data collection in this study was done by nurses. They were schooled to register the data in the Omaha system in a uniform way, but filling in the data can only be standardized to a certain degree: the data could have been biased depending on which nurse filled it in. The data could also contain human errors, which could have biased the results. For future studies it is therefore recommended to discuss data quality with the HHC organization, to see if and where the chances of human error and bias are higher. Additionally, nurses could add comments in open-text fields if they wanted to register something that did not fit the given options of the Omaha system, however this data was not available. For future research, it would be interesting to take this into account as well. Lastly, for future research it is recommended to study the performance of XGBoost: this algorithm seemed to perform well but I did not have enough computational power to work with it.

6. Conclusion

This study focused on creating a machine learning model which predicts hours of HHC needed per week, based on data on personal and treatment characteristics. This was done by comparing the evaluation metrics of three different algorithms, after which the role of patient clusters in predicting hours of HHC was studied. The results showed that Random Forest Regressor led to the best results, with an R^2 of 0.64, an MAE of 3.26 and an MSE of 29.59. This outperformed the baseline, which was the estimation of the nurse. The residuals on lower hours of care were smaller than residuals on higher hours of care, which seemed to be linked to the amount of data available. The addition of patient clusters did not improve the model. The patient clusters used were not found to be linked to hours of HHC and they were not seen as reliable: it is thus recommended to use different patient clusters or develop them further before using them as features for prediction of hours of HHC. The ability of this model to predict hours of HHC at the beginning of the care period of a new client, is valuable for HHC organizations. They could use this to improve the scheduling of their nurses, which could lead to less job-stress and therefore reduce burn-out rates among nurses. Furthermore, it gives more insight in the utilization of care. Future research could combine data from several healthcare organizations to make a more robust model. Furthermore, it would be interesting to take the logistics of HHC scheduling into account as well, such as travel time, routing and working shifts. The prediction of HHC hours is shown to have promising possibilities, which hopefully leads to many more studies on this topic in the future.

7. References

Alsinglawi, B., Alnajjar, F., Mubin, O., Novoa, M., Alorjani, M., Karajeh, O., Darwish, O. (2020, July). Predicting length of stay for cardiovascular hospitalizations in the intensive care unit: Machine learning approach. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (pp. 5442-5445). IEEE.

Antoniadis, A., Lambert-Lacroix, S., Poggi, J. M. (2021). Random Forest Regressors for global sensitivity analysis: A selective review. *Reliability Engineering System Safety, 206*, 107312. https://doi.org/10.1016/j.ress.2020.107312

Badosa, E. L., Tahull, M. B., Casas, N. V., Sangrador, G. E., Méndez, C. F., Meseguer, I. H., González, À.I., Urdiales, R.L., de Oca Burguete, F.J., Molas, M.T. and Farré, C.V (2017). Hospital malnutrition screening at admission: malnutrition increases mortality and length of stay. *Nutrition hospitalaria, 34(4),* 907-913.

Bernard, S., Heutte, L., Adam, S. (2009, June). Influence of hyperparameters on random forest accuracy. In *International workshop on multiple classifier systems (pp. 171-180).* Springer, Berlin, Heidelberg.

Bertsimas, D., Pauphilet, J., Stevens, J., Tandon, M. (2021). Predicting inpatient flow at a major hospital using interpretable analytics. *Manufacturing Service Operations Management.* https://doi.org/10.1287/msom.2021.0971

CBS (2022). Ouderen. Retrieved April 23rd, 2022 from https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/ouderen

de Korte, M. H., Verhoeven, G. S., Elissen, A. M. J., Metzelthin, S. F., Ruwaard, D., Mikkers, M. C. (2020). Using machine learning to assess the predictive potential of standardized nursing data for home healthcare case-mix classification. *European Journal of Health Economics, 21(8),* 1121–1129. https://doi.org/10.1007/s10198-020-01213-9

Disler, P. B., Roy, C. W., Smith, B. P. (1993). Predicting hours of care needed. *Archives of Physical Medicine and Rehabilitation, 74(2),* 139-143. https://doi.org/10.5555/uri:pii:000399939390352B

Flom (2022). The disadvantages of linear regression. Retrieved May 16th, 2022 from https://sciencing.com/disadvantages-linear-regression-8562780.html

Frost, Jim (202). ntroduction to Bootstrapping in Statistics with an Example. Retrieved May 15th, 2022 from https://statisticsbyjim.com/hypothesis-testing/bootstrapping/

Graw, J. H., Wood, W. T., Phrampus, B. J. (2021). Predicting global marine sediment density using the Random Forest Regressor regressor machine learning algorithm. *Journal of Geophysical Research: Solid Earth,* 126(1). https://doi.org/10.1029/2020JB020135

Hansen, C., Hansen, C., Alstrup, S., Lioma, C. (2017, November). Smart city analytics: Ensemble-learned prediction of citizen home care. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 2095-2098). https://doi.org/10.1145/3132847.3133101

Jakobsen, J. C., Gluud, C., Wetterslev, J., Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology, 17(1),* 1-10. https://doi.org/10.1186/s12874-017-0442-1

Lee, T., Mills, M. E. (2000). Analysis of patient profile in predicting home care resource utilization and outcomes. *JONA: The Journal of Nursing Administration, 30(2)*, 67-75.

Madsen, A. L., Olesen, K. G., Løvschall, H. L., Søndberg-Jeppesen, N., Jensen, F., Lindblad, M., Mogensen, L. L., Christensen, T. S. (2020, February). Prediction of High Risk of Deviations in Home Care Deliveries. In *International Conference on Probabilistic Graphical Models* (pp. 281-292). PMLR.

Maulud, D., Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends, 1(4),* 140-147.

Mekhaldi, R. N., Caulier, P., Chaabane, S., Chraibi, A., Piechowiak, S. (2020, April). Using Machine Learning Models to Predict the Length of Stay in a Hospital Setting. In *World Conference on Information Systems and Technologies (pp. 202-211).* Springer, Cham.

Mello (2020). XGBoost: Theory and Practice. Retrieved May 16th, 2022 from https://towardsdatascience.com/xgboost-theory-and-practice-fb8912930ad6

National health services (2022). NHS has reached tipping point, warn healthcare leaders. Retrieved May 14th, 2022 from https://www.nhsconfed.org/news/nhs-has-reached-tipping-point-warn-healthcare-leaders

Novakovic, M. (2022). Predicting the Number of Home Health Care Hours an Older Person Needs in the Netherlands [Master's Thesis, Tilburg University]. School of Humanities and Digital Sciences, Tilburg University.

Omaha System Support (2021). Het Omaha Classificatiesysteem. Retrieved March 1st, 2022, from https://www.omahasystem.nl/omaha-classificatiesys teem.html

Prognosemodel Zorg en Welzijn (2022). Dashboard Branches – Thuiszorg – Nederland. Retrieved March 1st, 2022 from https://prognosemodelzw.databank .nl/dashboard/dashboard-branches/thuiszorg

Rachel, H., Francesco, S. (2018). Factors associated with and impact of burnout in nursing and residential home care workers for the elderly. *Acta Bio Medica: Atenei Parmensis, 89,* 60. https://doi.org/10.23750/abm.v89i7-S.7830

Samima, S., Sarma, M. (2021). PSPHERE: person specific human error estimation. *Journal of Reliable Intelligent Environments,* 1-32. https://doi.org/10.1007/s40860-021-00146-1

Scikit-learn developers (2022). sklearn.model-selection.RandomizedSearchCV. Retrieved May 7th, 2022 from https://scikit-learn.org/stable/modules/generated/ sklearn.model-selection.RandomizedSearchCV.html

Sruthi, E. R. (2021). Understanding Random Forest. Retrieved May 16th, 2022 from https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Vektis (2020). Feiten en cijfers over wijkverpleging. Retrieved March 1st, 2022 from https://www.vektis.nl/intelligence/publicaties/factsheet-wijkverpleging: :text=32%20duizend%20mensen%20meer%20meter%20304 %20duizend%20per%20maand.

Vijayan, M. (2017). Impact of job stress on employees job performance in Aavin, Coimbatore. *Journal of Organisation and Human Behaviour, 6(3),* 21.

Volksgezondheid En Zorg (2020). Bevolking | Vergrijzing. Retrieved March 1st, 2022 from https://vzinfo.nl/bevolking/vergrijzing

8. Appendices

8.1 Appendix A: Omaha Classification System

The Omaha Classification System consists of three components: Problem Classification Schemes (Problem Areas), Intervention Schemes and Problem Rating Scales for Outcomes. In the table below, the possible Problem Classification Schemes are listed. After that, the Intervention Schemes and Problem Rating Scales for Outcomes are explained.

Table A.1
Domains and related Problem Classification Schemes

Domains	Problem Classification Schemes
Psychosocial domain	Communication with community resources, Social contact, Role change, Interpersonal relationship, Spirituality, Grief, Mental health, Sexuality, Caretaking/parenting, Neglect, Abuse, Growth and development
Environmental domain	Income, Sanitation, Residence, Neighbor-hood/workplace safety
Physiological domain	Hearing, Vision, Speech and language, Oral health, Cognition, Pain, Consciousness, Skin, Neuro-musculo-skeletal function, Respiration, Circulation, Digestion-hydration, Bowel function,Urinary function, Reproductive function, Pregnancy, Postpartum, Communi-cable/infectious condition
Health-related behaviors domain	Nutrition, Sleep and rest patterns, Physical activity, Personal care, Substance use, Family planning, Health care supervision, Medication regimen

Intervention Schemes are ‘actions’ that are used to describe the practices that the nurses do, to improve the health of client, diminish the decline of the client and/or provide comfort in the palliative state (OmahaSystem, 2022). The actions can be divided in four categories (OmahaSystem, 2022):

- Teaching, Guidance and Counselling: providing information, assisting in decision-making, encouraging responsibility for self-care
- Treatments and Procedures: ‘technical’ activities such as wound care
- Case Management: activities surrounding the service delivery, such as coordination and encouraging the individual to make use of appropriate resources
- Surveillance: detection, monitoring

The possible Intervention Schemes are the following: anatomy/physiology, anger management, behavior modification, bladder care, bonding/attachment, bowel care, cardiac care, caretaking/parenting skills, cast care, communication, community outreach worker services, continuity of care, coping skills, day care/respite, dietary management, discipline, dressing change/wound care, durable medical equipment, education, employment, end-of-life care, environment, exercises, family planning care, feeding procedures, finances, gait training, genetics, growth/development care, home, homemaking/housekeeping, infection precautions, interaction, interpreter/translator services, laboratory findings, legal system, medical/dental care, medication action/side effects, medication administration, medication coordination/ordering, medication prescription, medication set-up, mobility/transfers, nursing care, nutritionist care, occupational therapy care, ostomy care, other community resources, paraprofessional/aide care, personal hygiene, physical therapy care, positioning, recreational therapy care, relaxation/breathing techniques, respiratory care, respiratory therapy care, rest/sleep, safety, screening procedures, sickness/injury care, signs/symptoms-mental/emotional, signs/symptoms-physical, skin care, social work/counseling care, specimen collection, speech and language pathology care, spiritual care, stimulation/nurturance, stress management, substance use cessation, supplies, support group, support system, transportation, wellness and other.

The Problem Rating Scale for Outcomes consists of three scores per Problem Area: one score for the level of knowledge a client has about his/her problem, one score for the level of appropriateness of the client's behavior towards the problem and one score for the severity (status) of the problem. The scores are rated from 1, which means there is no knowledge, no appropriate behavior and a high level of severity, to 5, which means there is excellent knowledge, the client's behavior is appropriate and the level of severity is low.

8.2 Appendix B: patient clusters EscuLine

The patient clusters created by EscuLine are formed based on Problem Rating Outcome Scores. They are identified by the most dominant Problem Areas in that cluster. The clusters are still being developed and are not seen as definite.

Table B.1
Patient clusters and dominant Problem Areas

Patient clusters	Dominant Problem Areas
Cluster 0	Pain, personal care
Cluster 1	Nutrition
Cluster 2	Other
Cluster 3	Cognition
Cluster 4	Circulation, NO medication
Cluster 5	Medication
Cluster 6	Skin, NO medication
Cluster 7	Health care supervision, NO medication
Cluster 8	Urinary tract function
Cluster 9	Neural/muscular/skeleton functioning

8.3 Appendix C: Features final dataset

In Table C.1 an overview of all features in the final dataset is presented.

Table C.1
Variables in the final dataset

Content	Features and info
ID's	Care plan ID and Client ID
Info on client	Age at start of the care plan and dummy variables for gender, marital status, patient clusters
Info on care plan	Begin week and end week of the care plan and dummy variables for Problem Areas and Intervention Schemes
Hours of care	Hours of received care in week 2 of the care plan
Scores	Three scores per Problem Area, for target scores and difference score between target and current scores. When the client does not have the Problem Area, the scores are 0
Extracted features	Amount of Problem Areas, Intervention Schemes and domains.

8.4 Appendix D: Hyperparameters algorithms

See Table D.1 and D.2 for the hyperparameter tuning grid of Randomized Grid Search of respectively Random Forest Regressor and XGBoost. In Table D.3 and D.4, the final hyperparameters of respectively Random Forest Regressor and XGBoost can be found. The only hyperparameter of Linear Regression was 'positive', for which the values 'False' and 'True' were evaluated: 'True' was chosen.

Table D.1
Hyperparameter tuning grid of Randomized Grid Search for Random Forest Regressor

Hyperparameters	Value
Bootstrap	True
Max_features	'auto', 'sqrt', 'log2'
N_estimators	5, 10, 200, 250, 300
Max_depth	None, 5, 7.5, 10, 12.5, 15
Min_samples_leaf	1, 2, 3, 4, 5

Table D.2
Hyperparameter tuning grid of Randomized Grid Search for XGBoost

Hyperparameters	Value
Objective	'reg:squarederror'
Booster	'gbtree','gblinear'
Learning_rate	0.1 (chosen as 0.01 was too computationally expensive)
Max_depth	7, 10 ,15, 20
Min_child_weight	10, 15, 20, 25]
Colsample_bytree	0.8, 0.9, 1
N_estimators	300, 400, 500, 600
Reg_alpha	0.5, 0.2, 1
Reg_lambda	2, 3, 5
Gamma	1, 2, 3

Table D.3
Hyperparameters of Random Forest Regressor

Hyperparameters	Value
Bootstrap	True
Max_features	'auto'
N_estimators	250
Max_depth	None
Min_samples_leaf	1

Table D.4
Hyperparameters of XGBoost

Hyperparameters	Value
Objective	'reg:squarederror'
Booster	'gbtree'
Learning_rate	0.1
Max_depth	10
Min_child_weight	10
Colsample_bytree	0.8
N_estimators	300
Reg_alpha	0.5
Reg_lambda	5
Gamma	3

8.5 Appendix E: Scatterplots sub research question 1

First, the scatterplots of rare Problem Areas and Intervention Schemes are presented. Then, the scatterplots of common Problem Areas and Intervention Schemes are presented. The color of the dots are dependent on the Problem Areas and Intervention Schemes, where the dark color means that the Problem Area or Intervention Scheme is not present, and the lighter color means that the Problem Area or Intervention Scheme is present.

Rare Problem Areas and Intervention Schemes

Figure E.1
Problem Area Hearing

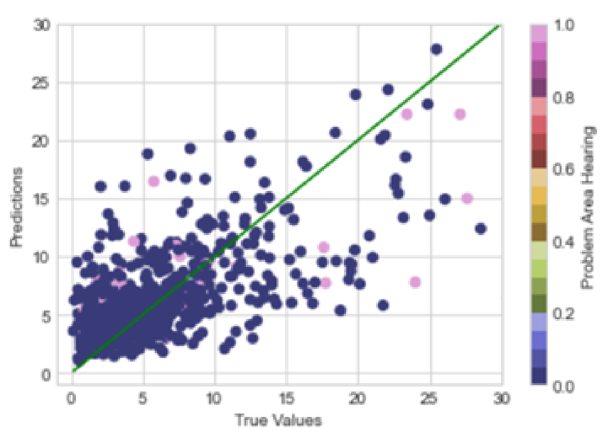


Figure E.2
Problem Area Income/Finance

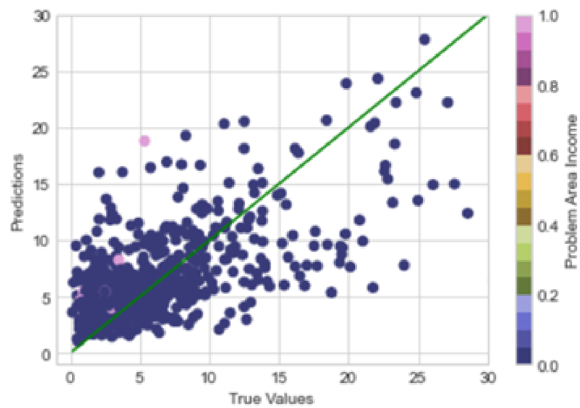


Figure E.3
Problem Area Interpersonal relationships

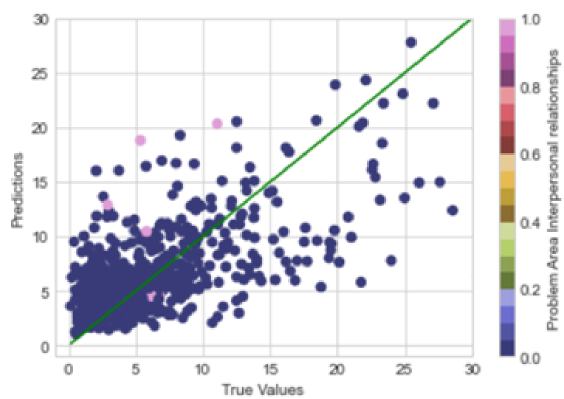


Figure E.4
Intervention Scheme finance

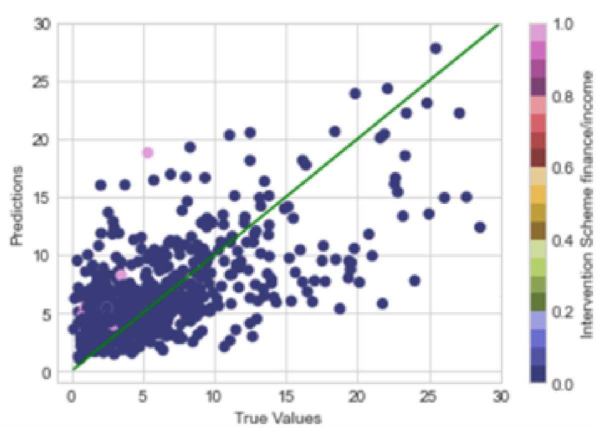
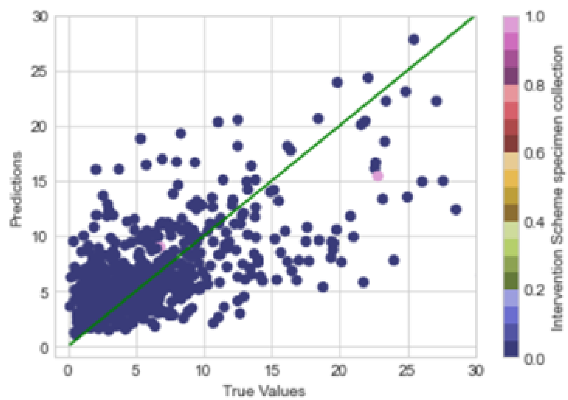


Figure E.5
Intervention Scheme specimen collection



Common Problem Areas and Intervention Schemes

Figure E.6
Problem Area Medication

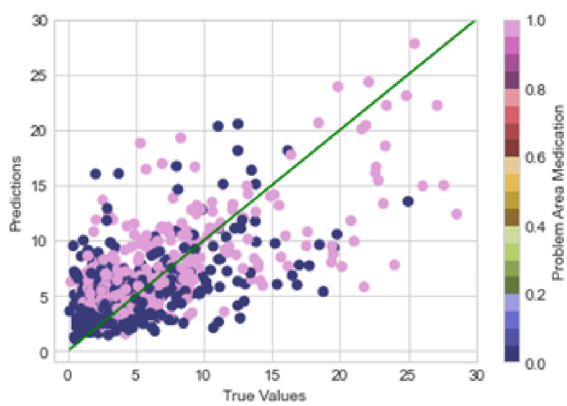


Figure E.7
Problem Area Nutrition

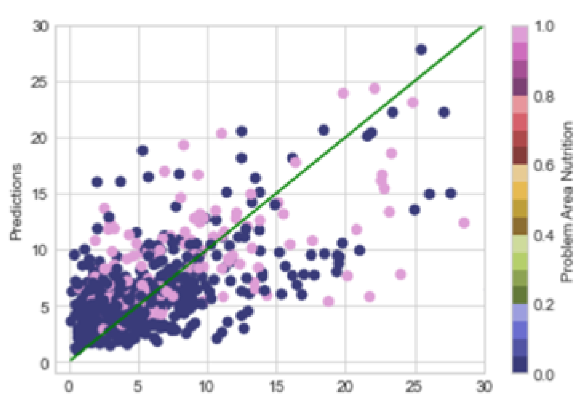


Figure E.8
Problem Area Personal care

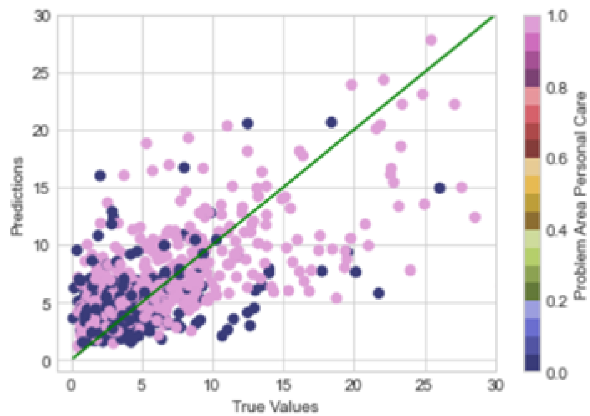


Figure E.9
Intervention Scheme continuity of care

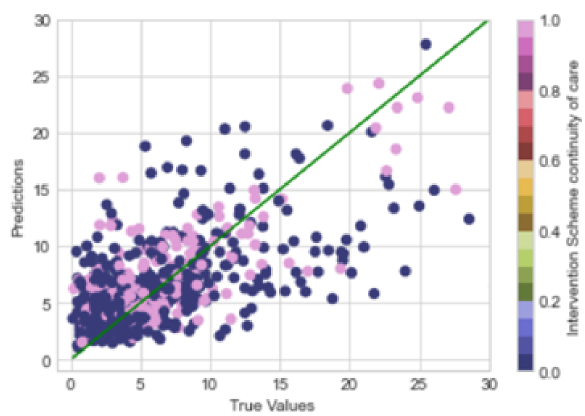


Figure E.10
Intervention Scheme skin care

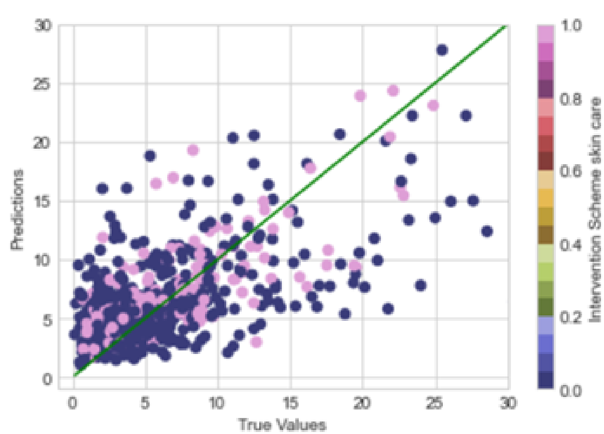
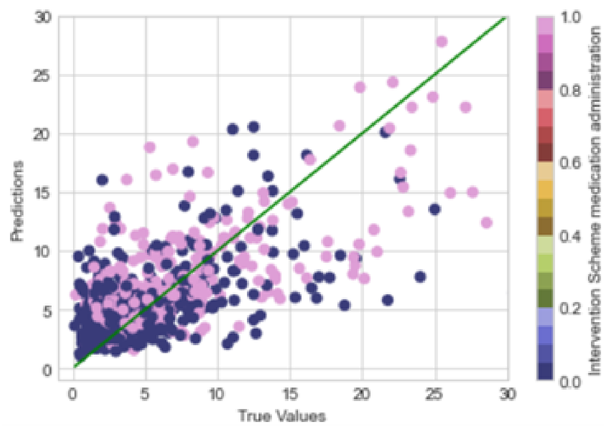


Figure E.11

Intervention Scheme medication administration



Scatter plots 'Medication prescription' and 'supplies'

Figure E.12

Intervention scheme medication prescription

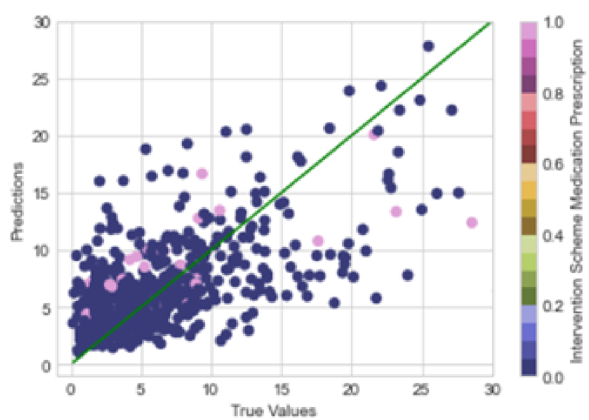
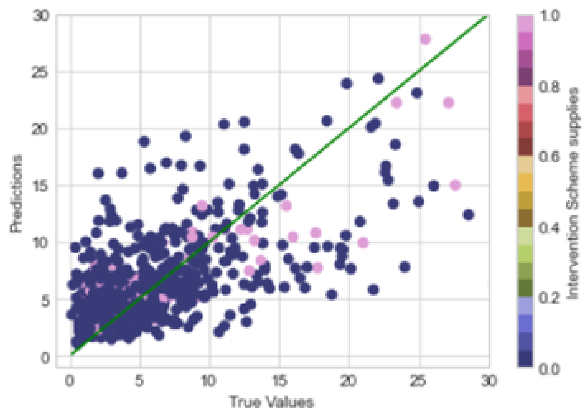


Figure E.13
Intervention scheme supplies



8.6 Appendix F: Feature importance sub research question 1

Here, a graph of the feature importance of sub research question 1 as calculated with Random Forest Feature Importance is presented.

Figure F1

Feature importance of the model of sub research question 1