Diseases and Health in Cattle Herds

Project report
Øget konkurrencekraft i landbruget gennem brug af kunstig intelligens - AP1C: Sygdomme og sundhed i kvægbesætninger
Data Science, Digital, SEGES
December 11, 2018

Summary

The objective of the project was to predict cow life expectancy using diseases and health related data from the DCDB database. Several machine learning models were used: Dummy regressor, Ridge, SVR, Elastic Net, Gradient Boosting Regressor, and Random Forest Regressor. To evaluate and pick the best model, a 10-fold cross validation and grid search analysis was used. Unfortunately, these models with the chosen feature sets were not able to make a acceptable predictions on life expectancy. Future improvements were discussed and refined with domain experts.

Table of Contents

Introduction

In the danish cattle database (DCDB) ("Kvægdatabasen" in danish), one of the largest agricultural databases, data related to health status and disease treatments of individual cattle herds have been collected for several years. Detailed registrations are required for herds that are part of a formalized healthcare process. Using machine learning (ML), we expect to be able to analyze and predict which individuals and herds are most at risk of health problems. Likewise, it will be possible to analyze the treatment process, enabling us to say more about which treatments are the most optimal in the individual situation. By combining this data with information on survival, release time, lactation, etc., it will be possible to predict the cows and herds where treatment is most profitable. The goal is to describe a decision support tool that can propose the most optimal decision in each case.

This project deals with predicting some future quantity related to an individual cattle at some point during its life course. Specifically, we attempt to predict the time to culling at the time of the first calving. We denote this time interval "cow life expectancy", as illustrated in Figure 1.

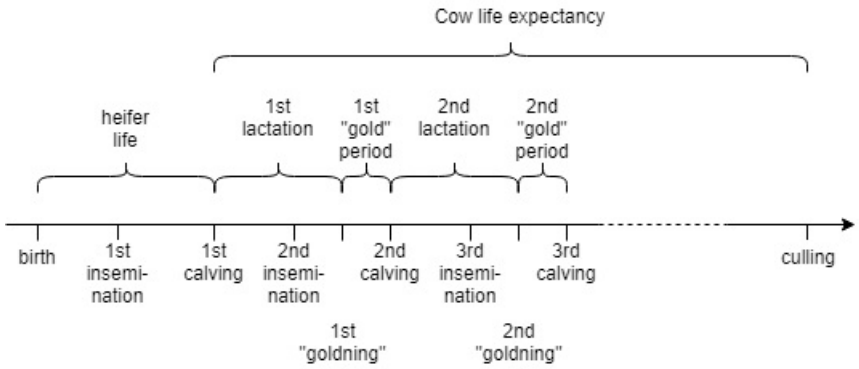


Figure 1: Timeline of the events in the typical cattle life course.

We utilize data from DCDB in combination with a set of common supervised ML methods to train multiple models for prediction of cow life expectancy. To develop such models, we consider the following problem statements:

- How can data from DCDB be transformed to features for the ML methods?
- How much training data is needed for the ML methods to provide accurate predictions?
- What level of uncertainty is present in the provided predictions?
- Which features are the most important ones in explaining the prediction?

Methods

In this chapter we describe the DCDB data and ML methods utilized in the project, along with our approach to evaluate the performance of the models.

Data collection

DCDB contains data regarding many different aspects of cattle breeding and dairy production. Thus, a cattle domain expert and an experts in the DCDB data model were both consulted in a series of workshops to obtain the relevant data for our prediction of cow life expectancy at the time of the first calving.

Initially, we made the assumption that a larger data set, both in terms of number of features and samples, would improve our later predictions. Thus, we use the largest data set possible based of the data in DCDB. Our data set contains registrations concerning not only the cow it self, but also its mother and father.

Our initial group of cows we could include in our data set were restricted by the following parameters:

- belong to the race Danish Holstein,
- born after the January 1'st 2000,
- had its first carving, and
- culled, slaughtered, or natural death before the 31'st of December 2017.

This restriction resulted in a total of 1.989.374 cows extracted from DCDB.

Our data analysis (see `./MVP2/initial_feature_set_analysis.ipynb`) shows:

- 1.967 cows, of the 1.989.374, have suspicious registrations that fail our validity checks - e.g. the instances where the first carving of the mother was registered to have happened after the birth of the cow. Removing these cows from the dataset resulted in a dataset of 1.987.407 cows.
- Only 2 cows, of the remaining 1.987.407, contains non-null values for all features. This is not enough samples for our ML models.
- We can see that many of the lksk features contain nulls. This is because the heifer or the mother do not have any registrations of sickness. We assume that no registrations are equal to the situation where the cow has never been sick. Thus, we can derive that the null values of these features can be assigned to 0.

We continued our analysis of nulls using the data set with the derived NaN values. In the count of non-null values below, it can be seen that the features with the fewest non-null values are `tomdvaegt`, `mor_huld_aendring_goldperiode` and `foedselsvaegt` with 687, 18.607, and 128.312, respectively.

Our data analysis (see `./MVP2/initial_feature_set_analysis.ipynb`) also shows:

- Removing `tomdvaegt` results in 1.576 samples containing non-null values for all features.
- Removing `tomdvaegt` and `mor_huld_aendring_goldperiode` results in 32.026 samples containing non-null values for all features.
- Removing `tomdvaegt`, `mor_huld_aendring_goldperiode`, and `foedselsvaegt` results in 330.957 samples containing non-null values for all features.
- This is enough samples for our goal in our task definition stating a minimum of 5.000 samples. However, we assume our ML models will perform better with more samples, thus we will continue our investigation to reduce the nulls.

The last workshop with the cattle experts resulted in the set of 24 features (shown in Table 1, 2, and 3) and one response variable (shown in Table 4). Note that, many of our features are derived from combining multiple data attributes in DCDB, and the feature types found in the columns `Feature type` are restrictions of the data based on the knowledge provided by cattle exports, as explained in the columns `Feature description`. The main purpose of the restrictions based on the domain knowledge is to ensure a better feature set with less noise. This is performed by restricting the range of values and remove all empty values (i.e. no NULL values) from the final feature set. These restrictions resulted in a total of 53.731 cows.

After analyzing the reason why our restrictions only left 53.731 of the initial 1.987.407 cows, we conclude that the registration of the birth weight feature name `foedselsvaegt` is sparse. Thus, we decide to collect two feature sets: one with the birth weight feature containing 53.731 samples and one without the birth weight feature containing 718.161 samples. Both data science and cattle domain experts agreed that these datasets are of reasonable extent and quality for use in machine learning models.

Feature name	**Feature type**	**Feature description**
`alder_insem`	integer(300;900)	The age of the cow in days at her first insemination. We restrict to ages between 300 and 900 days.
`alder_kaelvning`	integer(550;1.200)	The age of the cow in days at her first carving. We restrict to ages between 550 and 1.200 days.
`foedselsaar`	categorical(2000,...,2013)	The birth year of the cow. We restrict to cows born after 1999 and before 2014 and.
`foedselsdag`	categorical(1,...,31)	The day in the month when the cow was born.
`foedselsdagiugen`	categorical(0,...,6)	The week day when the cow was born.
`foedselsmd`	categorical(1,...,12)	The birth month of the cow.
`foedstrkode`	categorical(1,2,3,4), ordered	The ordered size (1 small - 4 large) of the cow at the time of its birth. We remove cows having the size 5 as this category have been repealed in DCDB as part of changing the scale at some point from a 5 level scale to a 4 level scale.
`insem_antal`	categorical(1,2,3,4,5+), ordered	Number of inseminations of the cow before her first carving.
`lksk_11plus`	bool	Boolean of whether the cow have any disease registrations, with the code no. 11, 12, 14, 15, 94, 95, and 179, from its birth to its first carving.
`lksk_2`	bool	Boolean feature like `lksk_11plus`, but for disease code no. 2.
`lksk_28plus`	bool	Boolean feature like `lksk_11plus`, but for disease code no. 28 and 51.
`lksk_41`	categorical(0,1-3,4+), ordered	The number of disease registrations where the cow have had pneumonia (i.e. code no. 41). This is not a boolean feature like `lksk_11plus` as we assume the number of times the cow has pneumonia will effect its life expectancy.
`lksk_53`	bool	Boolean feature like `lksk_11plus`, but for disease code no. 53.
`lksk_72`	bool	Boolean feature like `lksk_11plus`, but for disease code no. 72.
`foedselsvaegt`	float(25;60)	The birth weight of the cow (i.e. weight measured in the first 3 days after the birth) "Kviens vægt ved fødsel (målt på dagen eller op til 3 dage efter)". This attribute seems to include a lot of standard figures between 25 and 60 kg.

Table 1: Features regarding the cow.

Feature name	**Feature type**	**Feature description**
`mor_forloebskode`	categorical(1,2,3-5), ordered	Progress code for the birth of the cow. We remove all cows where this value is 0, since this category have been repealed in DCDB.
`mor_kaelvningsnr`	categorical(1,2,3,4,5,6-10), ordered	The number the cow is in the carving sequence of its mother.
`mor_alder_foerste_kaelvning`	integer(550;1200)	The ages in days of the mother at her first carving.
`mor_goldperiode`	categorical(-1,0-28,29-	The length of the dry period (in danish "goldperiode") of the mother in days before she gives birth to the cow. Note that, the flag value (i.e. -1) represents the first born cows

	42,43-56,57-119)	(i.e. where the feature `mor_kaelvningsnr` equals 1).
`mor_lksk_12`	bool	Boolean of whether the mother have any disease registrations, with the code no. 12, in the period of one week before and three weeks after the birth of the cow.
`mor_lksk_21`	bool	Boolean feature like `mor_lksk_12`, but for disease code no. 21.
`mor_lksk_22`	bool	Boolean feature like `mor_lksk_12`, but for disease code no. 22.
`mor_lksk_91`	bool	Boolean feature like `mor_lksk_12`, but for disease code no. 91.

Table 2: Features regarding the mother of the cow.

Feature name	**Feature type**	**Feature description**
`far_NTM`	float(-inf;inf)	The normalized nordic total merit (NTM) value of the father of the cow at the point in time of the first carving of the cow. The NTM value from 2000 to 2013 is divided into two groups (one with mean around 0 and one with mean around 100) corresponding to two different approaches to register NTM. These values are not comparable. Neither are NTM values for different years. Thus, the NTM values must be normalized on a year basis, i.e. all data for 2000 must be normalized, all data for 2001 must be normalized, and so forth.

Table 3: Features regarding the father of the cow.

Response variable name	**Variable type**	**Variable description**
`ko_levealder`	integer(0;7200)	Number of days from the first carving to the culling (i.e. culled, slaughtered, or natural dead) of the cow. This is the response variable.

Table 4: Response variable for the life expectancy of the cow.

Feature set

The collect datasets described above have been encoded and standardized to create the final feature set. However, as the dataset comes in two variations (i.e. with and without the `foedseslvaegt` feature), two feature sets are crated `/projects/DHICH/MVP2/final_feature_set_2018_07_25_15_21_28_with_weights.parquet` and `/projects/DHICH/MVP2/final_feature_set_2018_07_25_15_21_28_no_weights.parquet`, respectively.

The feature set with weights (containing the `foedselsvaegt` feature) amount to 53.731 heifers whereas the feature set without weights amounts to 718.161 heifers.

Profiling

To during our data analysis and for our final analysis of the two feature sets, we utilize the Python package called `pandas_profiling`, as it presents a lot of common statistics and visualizations of the data for an initial exploratory analysis of the dataset.

The `pandas_profiling` HTML reports for the two feature sets `/projects/DHICH/MVP2/profiling_final_feature_set_2018_07_25_15_21_28_no_weights.html` and `/projects/DHICH/MVP2/profiling_final_feature_set_2018_07_25_15_21_28_with_weights.html` are displayed in Appendix A Feature set profiling results.

Models

Several ML models were used in this project. Initially the following simple ML methods were investigated:

- Dummy regressor,
- Ridge and
- SVR.

Then the more complex models were investigated:

- Elastic Net,
- Gradient Boosting Regressor,
- Random Forest Regressor.

All these methods were used as implemented in the `sklearn` Python package.

Performance evaluation

To evaluate and select the best ML model, we performed a 10-fold cross validation. For every ML model, the learning curve with an increasing number of samples was plotted. This process was repeated for each of the feature sets: 53.731 samples with the birth weight feature and 718.161 samples without the birth weight feature. Finallye, a grid search analysis was used to tune the models for the best performance of the different ML models.

Results

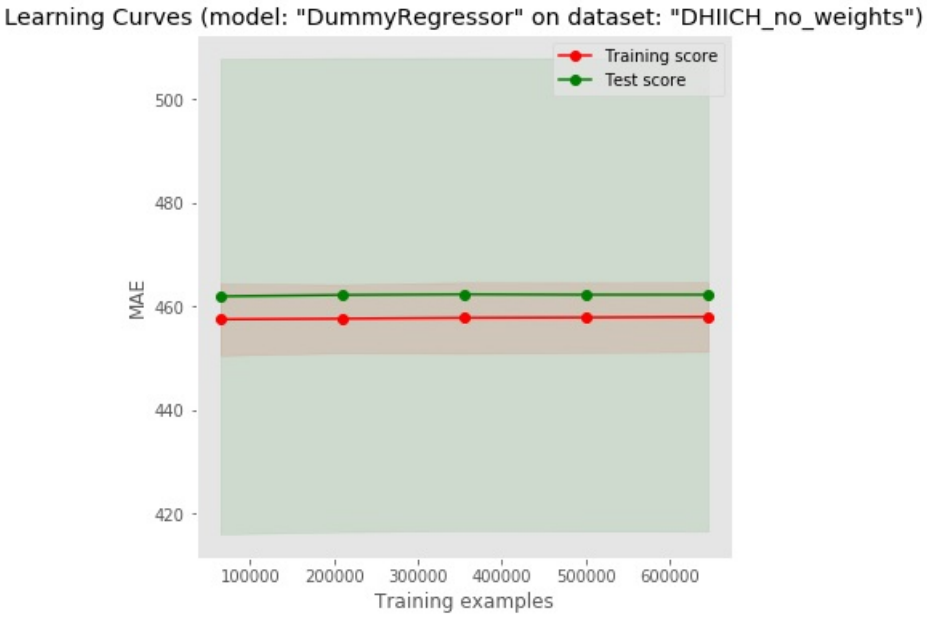
This section presents the results, as computed in the notebook `MVP4/DHICH_59_complex_models.ipynb`.

Learning curve

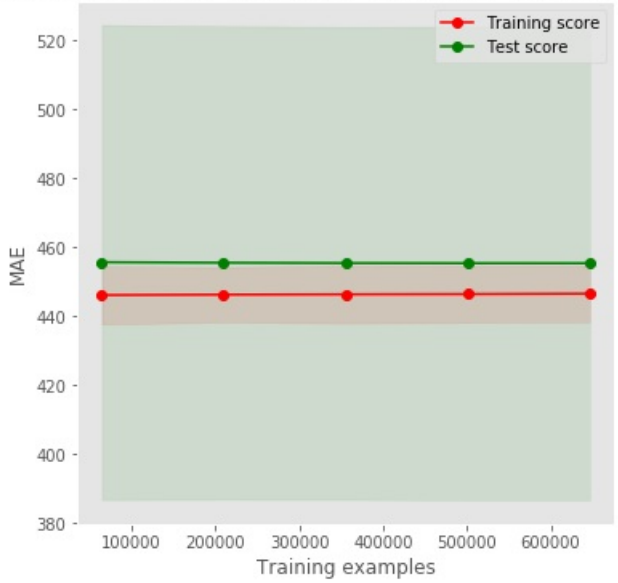
The following graphs show the learning curves (using 10-fold cross validation) for each utilized regression machine learning method based on it's default hyper-parameters, and for each of the two feature sets.

In [19]:

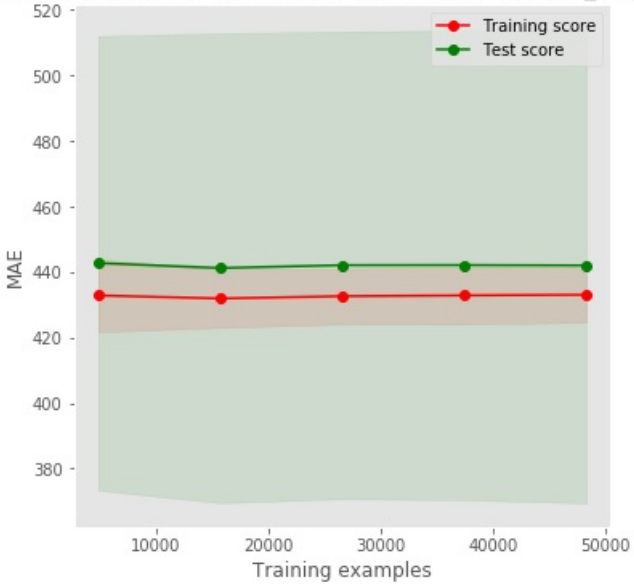
```
plot_learning_curves(df_learning_curve_local, score='MAE')
```



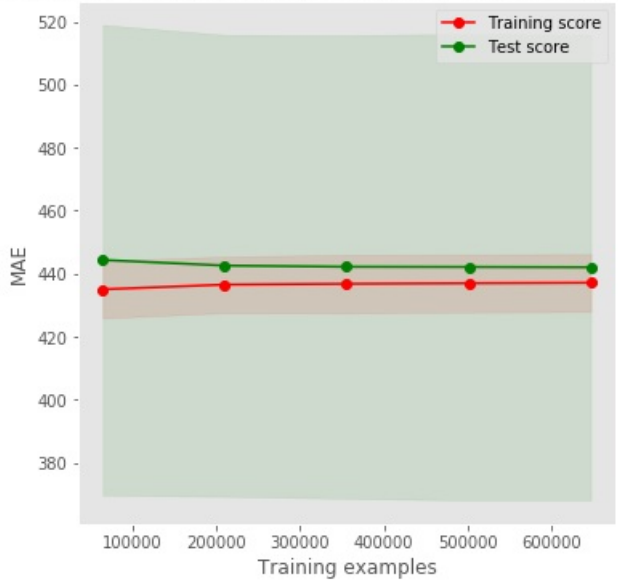
Learning Curves (model: "ElasticNet" on dataset: "DHIICH_no_weights")



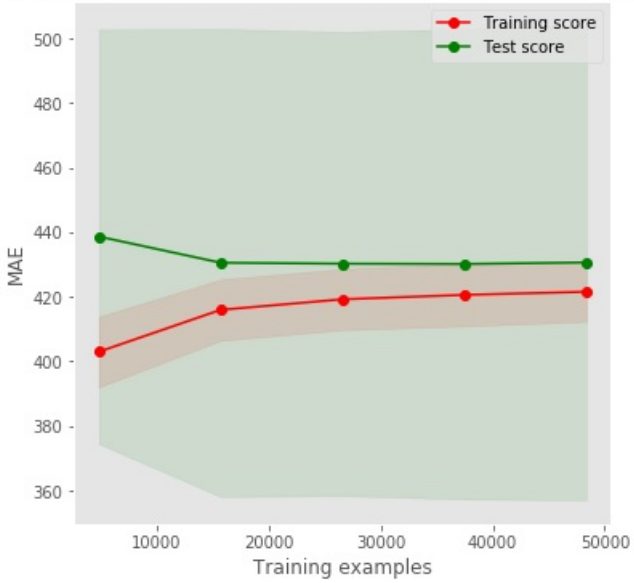
Learning Curves (model: "ElasticNet" on dataset: "DHIICH_with_weights")



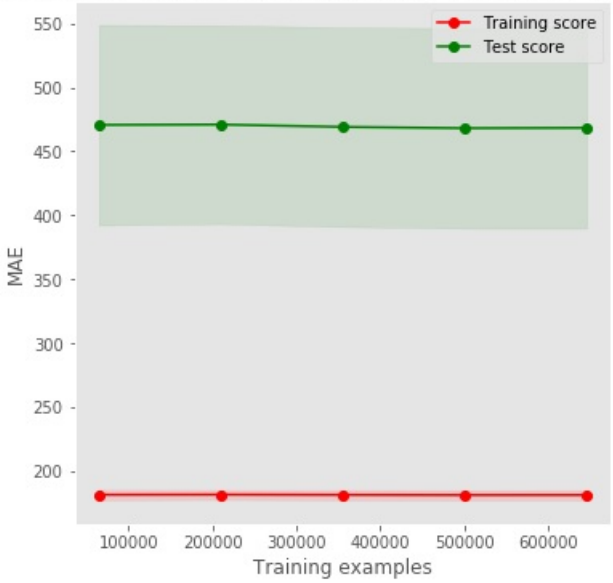
Learning Curves (model: "GradientBoostingRegressor" on dataset: "DHIICH_no_weights")



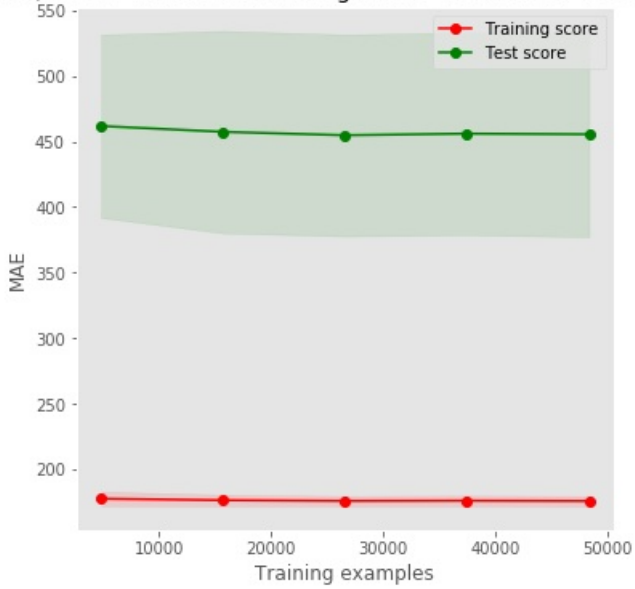
Learning Curves (model: "GradientBoostingRegressor" on dataset: "DHIICH_with_weights")

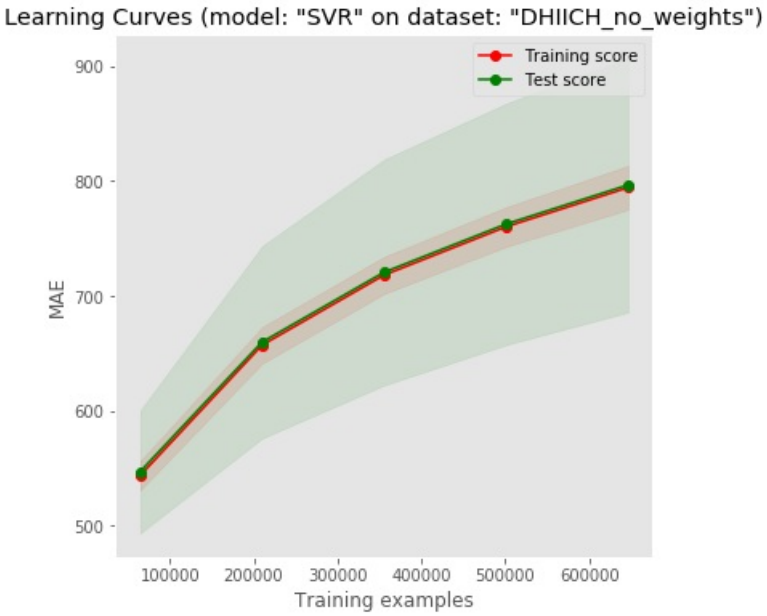


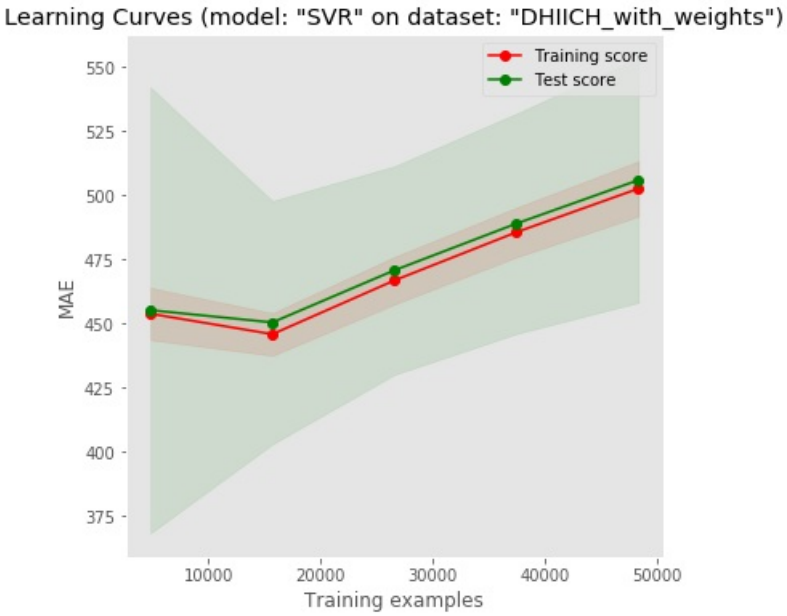
Learning Curves (model: "RandomForestRegressor" on dataset: "DHIICH_no_weights")



Learning Curves (model: "RandomForestRegressor" on dataset: "DHIICH_with_weights")







As is seen from the learning curves, our models show difficulty in learning a relation for cow life expectancy independently of number of training samples.

Hyper parameter optimization

A hyper-parameter optimization analysis was done using grid search with 10-fold cross validation and the results are shown in the following two tables below sorted for the best 10 results of absolute mean error and regression coefficient R^2, respectively. Note that all samples of the `DHIICH_with_weights` feature set were used for the grid search, where as only 50.000 samples of the `DHIICH_no_weights` feature set were used, with the purpose of decreasing the training time, as the learning curves did not show better predictions when increasing the amount of training samples.

In [5]:

```
performance_local.dataset_name.unique()
```

Out[5]:

```
array(['DHIICH_no_weights', 'DHIICH_with_weights'], dtype=object)
```

In [4]:

```
# Present the top 10 of mean_test_MAE
performance_local.sort_values('mean_test_MAE', ascending=True)[['model_name', 'dataset_name', 'params', 'mean_fit
_time', 'mean_train_MAE', 'mean_train_R2', 'mean_test_MAE', 'mean_test_R2']].reset_index(drop=True).head(20)
```

Out[4]:

	model_name	dataset_name	params	mean_fit_time	mean_train_MAE	mean_train_R2	mean_test_MAE	mean_test_R2
0	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	335.061606	417.978533	0.106897	422.611638	0.088579
1	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 4, "max_feature...	74.764557	416.525639	0.112597	422.664280	0.088690
2	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 5, "max_feature...	97.415801	411.194276	0.132107	422.703975	0.087838
3	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	50.230664	419.890137	0.100203	422.911129	0.088048
4	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 4, "max_feature...	14.262906	418.884390	0.102512	423.102289	0.085536
5	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 5, "max_feature...	14.717444	416.010219	0.113020	423.168017	0.085089
6	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 5, "max_feature...	17.379867	415.113184	0.117230	423.227098	0.086102
7	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	67.752193	418.204575	0.106739	423.248248	0.086833
8	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	74.016918	417.772393	0.108368	423.256027	0.086588
9	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	1586.707799	416.207886	0.113278	423.299208	0.085419
10	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	33.036591	422.145559	0.091493	423.347427	0.086438
11	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	100.961638	409.369245	0.139844	423.367134	0.085408
12	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	10.125166	421.323937	0.092285	423.433014	0.083655
13	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 4, "max_feature...	11.594028	419.453527	0.099432	423.506052	0.083292
14	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	83.764208	410.539042	0.135341	423.515260	0.085100
15	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	9.308869	421.997796	0.087060	424.109392	0.078760
16	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	321.066999	408.881740	0.141103	424.155871	0.081392
17	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	499.657488	407.116453	0.148044	424.296266	0.080797
18	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	7.805923	423.314611	0.082670	424.321853	0.078431
19	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 4, "max_features":...	69.445510	418.272806	0.118382	424.401672	0.091501

In [6]:

```
# Present the top 10 of mean_test_R2
performance_local.sort_values('mean_test_R2', ascending=False)[['model_name', 'dataset_name', 'params', 'mean_fit
_time', 'mean_train_MAE', 'mean_train_R2', 'mean_test_MAE', 'mean_test_R2']].reset_index(drop=True).head(20)
```

Out[6]:

	model_name	dataset_name	params	mean_fit_time	mean_train_MAE	mean_train_R2	mean_test_MAE	mean_test_R2
0	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 4, "max_features":...	69.445510	418.272806	0.118382	424.401672	0.091501
1	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 3, "max_features":...	53.503541	421.563558	0.104720	424.620030	0.091075
2	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 2, "max_features":...	43.160100	420.031245	0.110959	424.880955	0.090290
3	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 5, "max_features":...	97.576650	413.117762	0.139586	424.615226	0.090058
4	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 2, "max_features":...	51.108872	419.584533	0.112963	424.904898	0.089916
5	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 2, "max_features":...	29.969853	423.752433	0.095410	424.988930	0.089882
6	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 5, "max_features":...	11.895541	417.432422	0.119944	424.489895	0.089630
7	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 4, "max_features":...	11.086340	420.383471	0.108086	424.750606	0.089474
8	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 5, "max_features":...	14.024892	416.705039	0.123602	424.819741	0.089246
9	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 4, "max_feature...	74.764557	416.525639	0.112597	422.664280	0.088690
10	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 2, "max_features":...	327.236822	417.098000	0.122449	424.879097	0.088683
11	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 2, "max_feature...	335.061606	417.978533	0.106897	422.611638	0.088579
12	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 3, "max_features":...	7.853469	422.743297	0.097683	424.959301	0.088189
13	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 4, "max_features":...	8.905023	420.910123	0.105288	424.831389	0.088181
14	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 3, "max_feature...	50.230664	419.890137	0.100203	422.911129	0.088048
15	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "huber", "max_depth": 5, "max_feature...	97.415801	411.194276	0.132107	422.703975	0.087838
16	RandomForestRegressor	DHIICH_with_weights	{"max_depth": 5, "max_features": "auto", "n_es...	24.105792	424.242869	0.093300	425.677921	0.087601
17	RandomForestRegressor	DHIICH_with_weights	{"max_depth": 5, "max_features": "auto", "n_es...	15.609729	424.266480	0.093217	425.682586	0.087521
18	GradientBoostingRegressor	DHIICH_with_weights	{"loss": "ls", "max_depth": 3, "max_features":...	59.480790	412.470751	0.141586	425.289802	0.087488
19	RandomForestRegressor	DHIICH_with_weights	{"max_depth": 5, "max_features": "auto", "n_es...	6.448678	424.339440	0.092997	425.716822	0.087437

Discussion

Generally, the learning curves do not show a progress on the performance of the models for different sample sizes. The mean absolute error is decreasing for the larger dataset and R^2 is also increasing for the larger dataset, however only slightly, and unfortunately not to an acceptable level. The grid search analysis results tables suggest that the Gradient Boosting Regressor performs the best, but still with a coefficient of determination R^2 below 0.1. Thus, we cannot conclude that the model is capable of explaining cow life expectancy. The mean absolute error is measured in days. With a deviation in mean absolute error exceeding one year, we conclude that the model trained with the selected features is not able to predict the life expectancy to an acceptable level. For the current results of the grid search analysis, we decided to use a feature set of 50.000 samples based on the insignificant improvement over the learning curves. Overall, we must conclude that it is not clear which features are the most important to explain the prediction because of the uncertainty of the results. Future improvements to the features are necessary and the most important topic of future work. Suggestions for such feature improvements are presented in the Future work section.

Future work

In a meeting with a cattle domain expert, we looked at the features and suggested improvements to the feature set. Additionally, we looked into rethinking the problem such that in the future, better results are obtained when training the ML models. The main suggestions are:

Handling of the (non) I.I.D. problem:

- Include `besætningsID` or group the dataset by `besætningsID`, Practically, we may consider `fodselsbesaetning` or `kaelvningsbesaetning`. This requires a significant change to the data extraction.
- Limit data to a shorter time period, e.g. only heifers born in the most recent two years.

Change of response:

- Change the current regression problem to a classification problem. Binary classification: more or less than 3 lactations (which is considered an economic break even point).
- Multi class classification: 1, 2, 3, 4, 5, 6 or more lactations.
- Consider a completely different problem.
- Binary classification: Does the heifer survive the period from 6 months to first lactation?
- This requires a significant change to the data extraction and data munging steps.

Change of features:

- Consider removing mother related illness codes or simply all illness codes.
- Alternatively, include `besætningsID` to differentiate between registration practices.
- Consider training models based on a single feature at a time and picking the best performing features based on such an analysis.
- Pick only features that have a high correlation with the response.
- Simplify the `fodstrkode` and `lksks` by making them binary features.
- Pick the most important features based on domain knowledge of feature quality.
- Remove features with highly skewed distributions.

Highest priority changes:

- Limit time span of data.
- Change from regression to classification problem.
- Pick most important features based on domain knowledge.
- Pick most important features based on highest correlation / best "modelling power".

Tasks from the Kanban board future work list:

- Split dataset into evaluation and test set, and standardize the test set only based on the standardizer learned on the evaluation set.
- Only train models on the heifers born in 2011-2013.
- Re-run MVP4 learning curve and grid search on improved features.
- Assess feature importance.
- Use XGBoost.
- Overload sklearn fit and predict/score estermator methods to save all trained models and save all train and test predicts (for later computation of performance matrix).

Appendix A Feature set profiling results

Feature set with birth weights included.

In [11]:

```
from IPython.display import HTML
HTML('profiling_final_feature_set_2018_07_25_15_21_28_with_weights.html')
```

Out[11]:

Overview

Dataset info

Number of variables	77
Number of observations	53731
Total Missing (%)	0.0%
Total size in memory	7.2 MiB
Average record size in memory	140.0 B

Variables types

Numeric	13
Categorical	0
Boolean	64
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

Warnings

- `foedselsaar` has 2721 / 5.1% zeros **Zeros**
- `insem_antal_cat` has 30375 / 56.5% zeros **Zeros**
- `lksk_41_cat` has 49508 / 92.1% zeros **Zeros**
- `mor_forloebskode_cat` has 41951 / 78.1% zeros **Zeros**
- `mor_kaelvningsnr_cat` has 28026 / 52.2% zeros **Zeros**
- `scaled_far_ntm` has 4051 / 7.5% zeros **Zeros**

Variables

alder_insem
Numeric

Distinct count	547
Unique (%)	1.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	-0.14352
Minimum	-2.6928
Maximum	5.6558
Zeros (%)	0.0%

Toggle details

alder_kaelvning
Numeric

Distinct count	608
Unique (%)	1.1%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	-0.10784
Minimum	-2.7288
Maximum	4.3542

Zeros (%) 0.0%

Toggle details

dyr_id
Numeric

Distinct count	53731
Unique (%)	100.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	1008900000
Minimum	1001092816
Maximum	1015264220
Zeros (%)	0.0%

Toggle details

foedselsaar
Numeric

Distinct count	14
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	7.363
Minimum	0
Maximum	13
Zeros (%)	5.1%

Toggle details

foedselsdag_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.034077

True	1831	
(Missing)		51900

Toggle details

foedselsdag_10
Boolean

Distinct count	2

Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033649

True	1808	
(Missing)		51923

Toggle details

foedselsdag_11
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032774

True	1761	
(Missing)		51970

Toggle details

foedselsdag_12
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032495

True	1746	
(Missing)		51985

Toggle details

foedselsdag_13
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032998

True	1773	
(Missing)		51958

Toggle details

foedselsdag_14
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032998

True	1773	
(Missing)		51958

Toggle details

foedselsdag_15
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031937

True	1716	
(Missing)		52015

Toggle details

foedselsdag_16
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033295

True	1789	
(Missing)		51942

Toggle details

foedselsdag_17
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033575

True	1804	
(Missing)		51927

Toggle details

foedselsdag_18
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031602

True	1698	
(Missing)		52033

Toggle details

foedselsdag_19
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.03229

True	1735	
(Missing)		51996

Toggle details

foedselsdag_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032551

True	1749	
(Missing)		51982

Toggle details

foedselsdag_20
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033165

True	1782	
(Missing)		51949

Toggle details

foedselsdag_21
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.030876

True	1659	
(Missing)		52072

Toggle details

foedselsdag_22
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033147

True	1781	
(Missing)		51950

Toggle details

foedselsdag_23
Boolean

Distinct count 2

Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.034449

True	1851	
(Missing)		51880

Toggle details

foedselsdag_24
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033072

True	1777	
(Missing)		51954

Toggle details

foedselsdag_25
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032421

True	1742	
(Missing)		51989

Toggle details

foedselsdag_26
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032923

True	1769	
(Missing)		51962

Toggle details

foedselsdag_27
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033463

True	1798	
(Missing)		51933

Toggle details

foedselsdag_28
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033482

True	1799	
(Missing)		51932

Toggle details

foedselsdag_29
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.03082

True	1656	
(Missing)		52075

Toggle details

foedselsdag_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031416

True	1688	
(Missing)		52043

Toggle details

foedselsdag_30
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.029331

True	1576	
(Missing)		52155

Toggle details

foedselsdag_31
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.019635

True	1055	
(Missing)		52676

Toggle details

foedselsdag_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032793

True	1762	
(Missing)		51969

Toggle details

foedselsdag_5
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032719

True	1758	
(Missing)		51973

Toggle details

foedselsdag_6
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033184

True	1783	
(Missing)		51948

Toggle details

foedselsdag_7
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032905

True	1768	
(Missing)		51963

Toggle details

foedselsdag_8
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033351

True	1792	
(Missing)		51939

Toggle details

foedselsdag_9
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032607

True	1752	
(Missing)		51979

Toggle details

foedselsdagiugen_0
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.1469

True	7893	
(Missing)		45838

Toggle details

foedselsdagiugen_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.13998

True	7521	
(Missing)		46210

Toggle details

foedselsdagiugen_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14338

True	7704	
(Missing)		46027

Toggle details

foedselsdagiugen_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14241

True	7652	
(Missing)		46079

Toggle details

foedselsdagiugen_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14305

True	7686	
(Missing)		46045

Toggle details

foedselsdagiugen_5
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14377

True	7725	
(Missing)		46006

Toggle details

foedselsdagiugen_6
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14051

True	7550	
(Missing)		46181

Toggle details

foedselsmd_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

Mean 0.090097

True	4841	
(Missing)		48890

Toggle details

foedselsmd_10
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.094173

True	5060	
(Missing)		48671

Toggle details

foedselsmd_11
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.093838

True	5042	
(Missing)		48689

Toggle details

foedselsmd_12
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.09434

True	5069	
(Missing)		48662

Toggle details

foedselsmd_2
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.074985

True	4029	
(Missing)		49702

Toggle details

foedselsmd_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.078149

True	4199	
(Missing)		49532

Toggle details

foedselsmd_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.073514

True	3950	
(Missing)		49781

Toggle details

foedselsmd_5
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.070834

True	3806	
(Missing)		49925

Toggle details

foedselsmd_6
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.070313

True	3778	
(Missing)		49953

Toggle details

foedselsmd_7
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.080531

True	4327	
(Missing)		49404

Toggle details

foedselsmd_8
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.089855

True	4828	
(Missing)		48903

Toggle details

foedselsmd_9
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.089371

True	4802	
(Missing)		48929

Toggle details

foedselsvaegt
Numeric

Distinct count	84
Unique (%)	0.2%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	1.1962e-14
Minimum	-2.8098
Maximum	3.2608
Zeros (%)	0.0%

Toggle details

foedstrkode
Numeric

Distinct count	4
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.10324
Minimum	-2.1825
Maximum	2.0456

Zeros (%)	0.0%

Toggle details

insem_antal_cat
Numeric

Distinct count	5
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.76072
Minimum	0
Maximum	4
Zeros (%)	56.5%

Toggle details

ko_levealder
Numeric

Distinct count	2809
Unique (%)	5.2%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	871.27
Minimum	0
Maximum	4694
Zeros (%)	0.1%

Toggle details

lksk_11plus
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.012265

True	659
(Missing)	53072

Toggle details

lksk_2
Boolean

Distinct count	2

Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0036106

True	194
(Missing)	53537

Toggle details

lksk_28plus
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.064637

True	3473
(Missing)	50258

Toggle details

lksk_41_cat
Numeric

Distinct count	3
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.092628
Minimum	0
Maximum	2
Zeros (%)	92.1%

Toggle details

lksk_53
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.012693

True	682
(Missing)	53049

Toggle details

lksk_72
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0014517

True	78
(Missing)	53653

Toggle details

mor_alder_foerste_kaelvning
Numeric

Distinct count	595
Unique (%)	1.1%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	-0.10237
Minimum	-2.936
Maximum	4.2785
Zeros (%)	0.0%

Toggle details

mor_forloebskode_cat
Numeric

Distinct count	3
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.23884
Minimum	0
Maximum	2
Zeros (%)	78.1%

Toggle details

mor_goldperiode_0
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.5216

True	28026
(Missing)	25705

Toggle details

mor_goldperiode_1
Boolean

Distinct count	2
Unique (%)	0.0%

Missing (%)	0.0%
Missing (n)	0
Mean	0.014182

True	762	
(Missing)		52969

Toggle details

mor_goldperiode_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.066033

True	3548	
(Missing)		50183

Toggle details

mor_goldperiode_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.20619

True	11079	
(Missing)		42652

Toggle details

mor_goldperiode_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.19199

True	10316	
(Missing)		43415

Toggle details

mor_kaelvningsnr_cat
Numeric

Distinct count	6
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.96896
Minimum	0
Maximum	5
Zeros (%)	52.2%

Toggle details

mor_lksk_12
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0011911

True	64
(Missing)	53667

Toggle details

mor_lksk_21
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.028196

True	1515
(Missing)	52216

Toggle details

mor_lksk_22
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.025646

True	1378
(Missing)	52353

Toggle details

mor_lksk_91
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0010608

True	57
(Missing)	53674

Toggle details

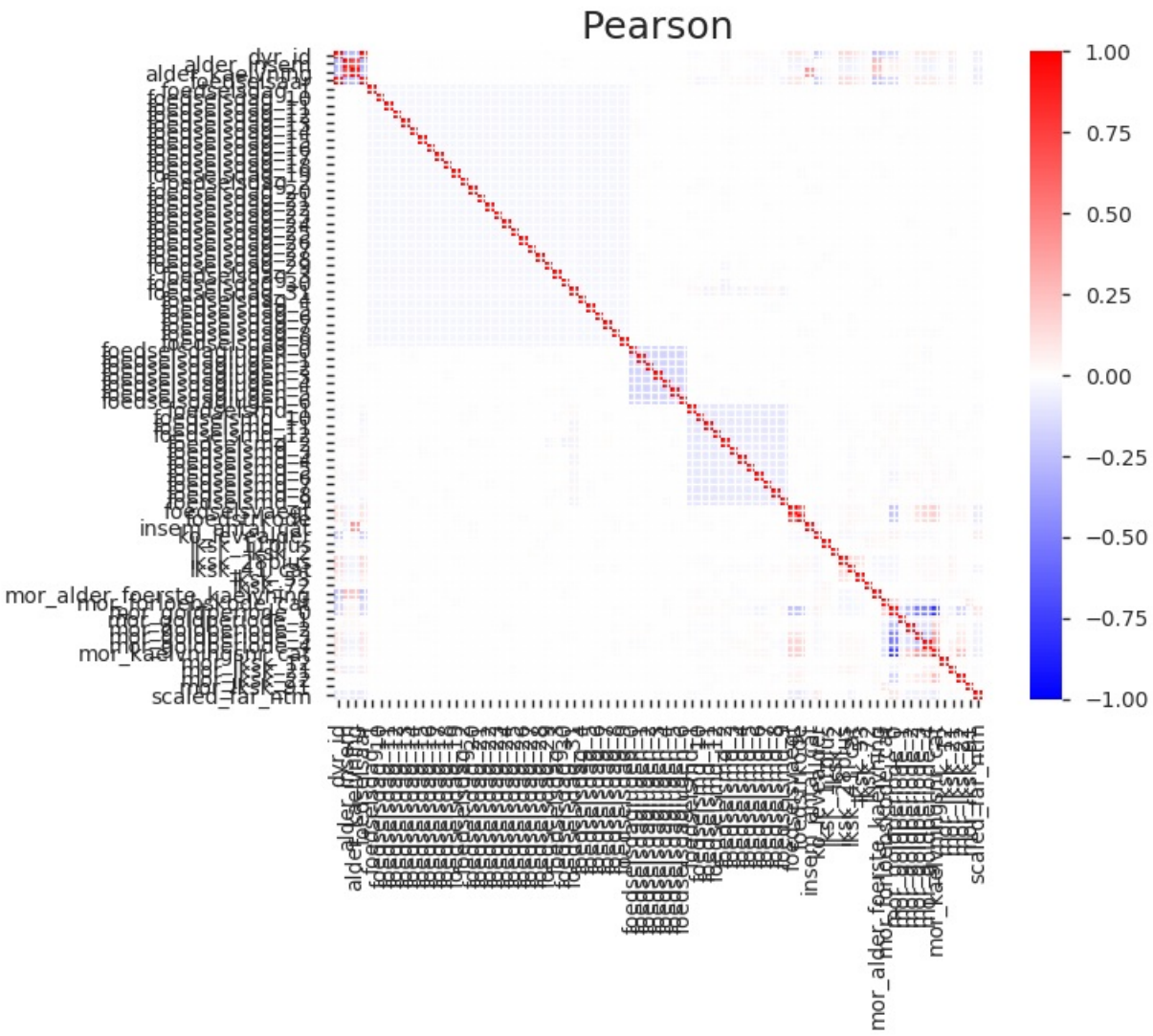
scaled_far_ntm
Numeric

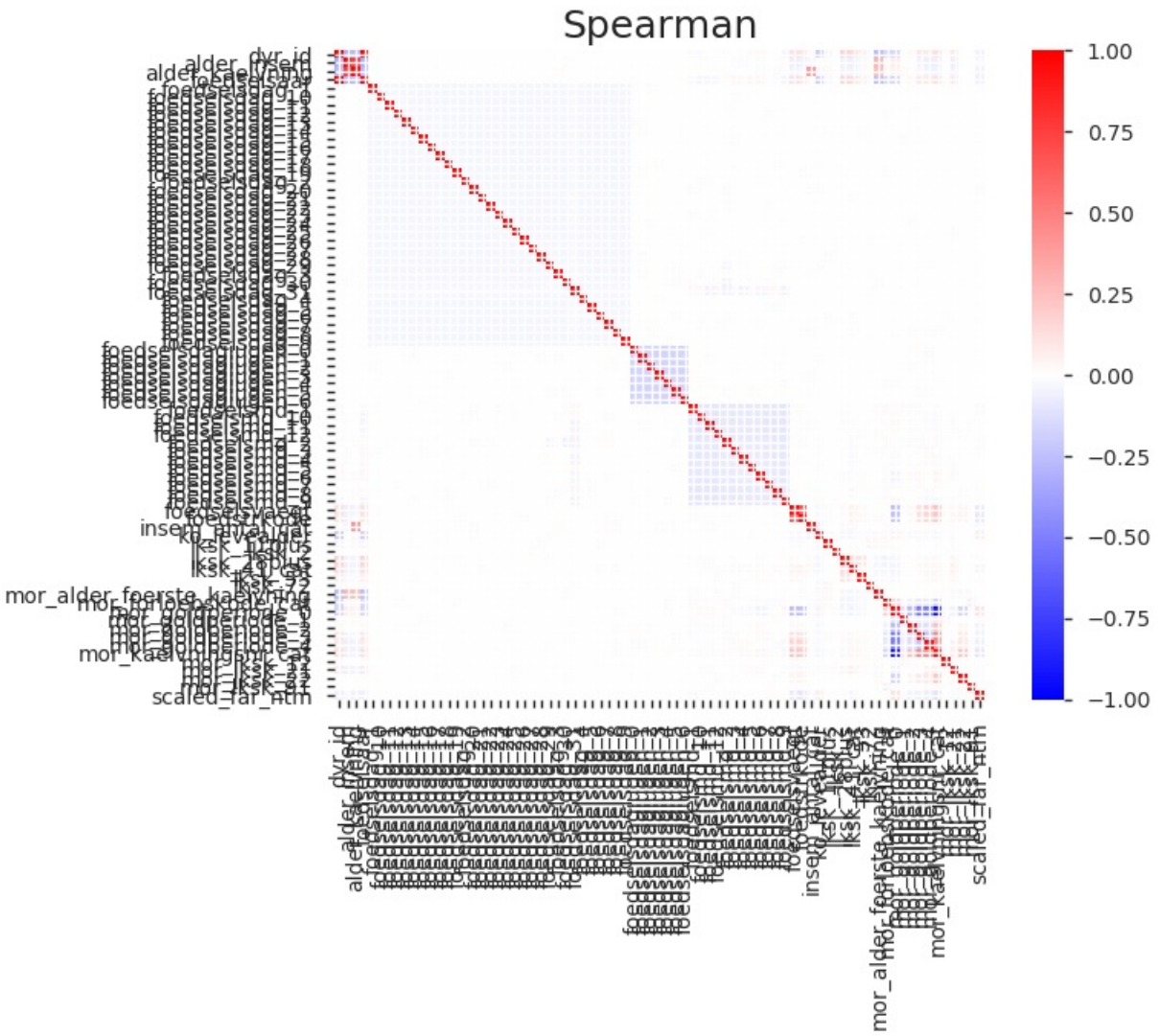
Numeric

Distinct count	381
Unique (%)	0.7%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	-0.1947
Minimum	-15.75
Maximum	9.8333
Zeros (%)	7.5%

Toggle details

Correlations





Sample

dyr_id	alder_insem	alder_kaelvning	foedselsaar	foedselsdag_1	foedselsdag_10	foedselsdag_11	foedselsdag_12	foedselsdag_13	foedselsdag_14	foe
1001092816	0.332847	-0.025748	0	False	False	False	False	False	False	Fal
1001092817	-0.899842	-1.095921	0	False	False	False	False	False	False	Fal
1001095854	-0.241474	-0.555318	0	True	False	False	False	False	False	Fal
1001096403	-1.179999	-0.919398	0	False	False	False	False	False	False	Fal
1001110424	0.150745	0.250070	0	False	False	False	False	False	False	Fal

Feature without birth weights

In [12]:

```
HTML('profiling_final_feature_set_2018_07_25_15_21_28_no_weights.html')
```

Out[12]:

Overview

Dataset info

Number of variables	76
Number of observations	718161
Total Missing (%)	0.0%
Total size in memory	90.4 MiB
Average record size in memory	132.0 B

Variables types

Numeric	12
Categorical	0
Boolean	64
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

Warnings

- `foedselsaar` has 43028 / 6.0% zeros **Zeros**
- `insem_antal_cat` has 414008 / 57.6% zeros **Zeros**
- `lksk_41_cat` has 679832 / 94.7% zeros **Zeros**
- `mor_forloebskode_cat` has 533309 / 74.3% zeros **Zeros**
- `mor_kaelvningsnr_cat` has 427649 / 59.5% zeros **Zeros**
- `scaled_far_ntm` has 58317 / 8.1% zeros **Zeros**

Variables

alder_insem

Numeric

Distinct count	601
Unique (%)	0.1%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	-6.2245e-15
Minimum	-2.7489
Maximum	5.6558
Zeros (%)	0.0%

Toggle details

alder_kaelvning

Numeric

Distinct count	651
Unique (%)	0.1%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	2.404e-15
Minimum	-2.817
Maximum	4.3542
Zeros (%)	0.0%

Toggle details

dyr_id
Numeric

Distinct count	718159
Unique (%)	100.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	1008300000
Minimum	1001091827
Maximum	1016596816
Zeros (%)	0.0%

Toggle details

foedselsaar
Numeric

Distinct count	14
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	6.7489
Minimum	0
Maximum	13
Zeros (%)	6.0%

Toggle details

foedselsdag_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.034015

True	24428	
(Missing)		693733

Toggle details

foedselsdag_10
Boolean

Distinct count	2
Unique (%)	0.0%

Missing (%)	0.0%
Missing (n)	0
Mean	0.034432

True	24728	
(Missing)		693433

Toggle details

foedselsdag_11
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031848

True	22872	
(Missing)		695289

Toggle details

foedselsdag_12
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033682

True	24189	
(Missing)		693972

Toggle details

foedselsdag_13
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031968

True	22958	
(Missing)		695203

Toggle details

foedselsdag_14
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033229

True	23864	
(Missing)		694297

Toggle details

foedselsdag_15
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033516

True	24070	
(Missing)		694091

Toggle details

foedselsdag_16
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032814

True	23566	
(Missing)		694595

Toggle details

foedselsdag_17
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033066

True	23747	
(Missing)		694414

Toggle details

foedselsdag_18
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033565

True	24105	
(Missing)		694056

Toggle details

foedselsdag_19
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032106

True	23057	
(Missing)		695104

Toggle details

foedselsdag_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032621

True	23427	
(Missing)		694734

Toggle details

foedselsdag_20
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.034427

True	24724	
(Missing)		693437

Toggle details

foedselsdag_21
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032636

True	23438	
(Missing)		694723

Toggle details

foedselsdag_22
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033253

True	23881	
(Missing)		694280

Toggle details

foedselsdag_23
Boolean

Distinct count	2

Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033026

True	23718	
(Missing)		694443

Toggle details

foedselsdag_24
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032912

True	23636	
(Missing)		694525

Toggle details

foedselsdag_25
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0332

True	23843	
(Missing)		694318

Toggle details

foedselsdag_26
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.033118

True	23784	
(Missing)		694377

Toggle details

foedselsdag_27
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032619

True	23426	
(Missing)		694735

Toggle details

foedselsdag_28
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.03353

True	24080	
(Missing)		694081

Toggle details

foedselsdag_29
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.030064

True	21591	
(Missing)		696570

Toggle details

foedselsdag_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031785

True	22827	
(Missing)		695334

Toggle details

foedselsdag_30
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.029532

True	21209	
(Missing)		696952

Toggle details

foedselsdag_31
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.018545

True	13318	
(Missing)		704843

Toggle details

foedselsdag_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032519

True	23354	
(Missing)		694807

Toggle details

foedselsdag_5
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032376

True	23251	
(Missing)		694910

Toggle details

foedselsdag_6
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032589

True	23404	
(Missing)		694757

Toggle details

foedselsdag_7
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032275

True	23179	
(Missing)		694982

Toggle details

foedselsdag_8
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.032984

True	23688
(Missing)	694473

Toggle details

foedselsdag_9
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.031746

True	22799
(Missing)	695362

Toggle details

foedselsdagiugen_0
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14477

True	103965
(Missing)	614196

Toggle details

foedselsdagiugen_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14263

True	102430
(Missing)	615731

Toggle details

foedselsdagiugen_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14292

True	102643
(Missing)	615518

Toggle details

foedselsdagiugen_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14299

True	102691	
(Missing)		615470

Toggle details

foedselsdagiugen_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14271

True	102489	
(Missing)		615672

Toggle details

foedselsdagiugen_5
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14134

True	101504	
(Missing)		616657

Toggle details

foedselsdagiugen_6
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.14264

True	102439	
(Missing)		615722

Toggle details

foedselsmd_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

Mean 0.089544

True	64307	
(Missing)		653854

Toggle details

foedselsmd_10
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.095816

True	68811	
(Missing)		649350

Toggle details

foedselsmd_11
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.092563

True	66475	
(Missing)		651686

Toggle details

foedselsmd_12
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.094885

True	68143	
(Missing)		650018

Toggle details

foedselsmd_2
Boolean

Distinct count 2
Unique (%) 0.0%
Missing (%) 0.0%
Missing (n) 0
Mean 0.075355

True	54117	
(Missing)		664044

Toggle details

foedselsmd_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0752

True	54006	
(Missing)		664155

Toggle details

foedselsmd_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.069778

True	50112	
(Missing)		668049

Toggle details

foedselsmd_5
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.069778

True	50112	
(Missing)		668049

Toggle details

foedselsmd_6
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.069548

True	49947	
(Missing)		668214

Toggle details

foedselsmd_7
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.079154

True	56845	
(Missing)		661316

Toggle details

foedselsmd_8
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.093389

True	67068	
(Missing)		651093

Toggle details

foedselsmd_9
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.09499

True	68218	
(Missing)		649943

Toggle details

foedstrkode
Numeric

Distinct count	4
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	4.6012e-13
Minimum	-2.1825
Maximum	2.0456
Zeros (%)	0.0%

Toggle details

insem_antal_cat
Numeric

Distinct count	5
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.72619
Minimum	0
Maximum	4

Zeros (%) 57.6%

Toggle details

ko_levealder
Numeric

Distinct count	3779
Unique (%)	0.5%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	893.34
Minimum	0
Maximum	5324
Zeros (%)	0.1%

Toggle details

lksk_11plus
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.015593

True	11198
(Missing)	706963

Toggle details

lksk_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0030871

True	2217
(Missing)	715944

Toggle details

lksk_28plus
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.03965

True	28475

(Missing)	689686

Toggle details

lksk_41_cat
Numeric

Distinct count	3
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.062774
Minimum	0
Maximum	2
Zeros (%)	94.7%

Toggle details

lksk_53
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0069748

True	5009
(Missing)	713152

Toggle details

lksk_72
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0016974

True	1219
(Missing)	716942

Toggle details

mor_alder_foerste_kaelvning
Numeric

Distinct count	647
Unique (%)	0.1%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	2.9228e-14
Minimum	-2.936
Maximum	4.2785
Zeros (%)	0.0%

Toggle details

mor_forloebskode_cat
Numeric

Distinct count	3
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.2778
Minimum	0
Maximum	2
Zeros (%)	74.3%

Toggle details

mor_goldperiode_0
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.59548

True	427649
(Missing)	290512

Toggle details

mor_goldperiode_1
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.013441

True	9653
(Missing)	708508

Toggle details

mor_goldperiode_2
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.061029

True	43829
(Missing)	674332

Toggle details

mor_goldperiode_3
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.17652

True	126771
(Missing)	591390

Toggle details

mor_goldperiode_4
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.15353

True	110259
(Missing)	607902

Toggle details

mor_kaelvningsnr_cat
Numeric

Distinct count	6
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.82986
Minimum	0
Maximum	5
Zeros (%)	59.5%

Toggle details

mor_lksk_12
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.0011975

True	860
(Missing)	717301

Toggle details

mor_lksk_21
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.022434

True	16111
(Missing)	702050

Toggle details

mor_lksk_22
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.022279

True	16000
(Missing)	702161

Toggle details

mor_lksk_91
Boolean

Distinct count	2
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0
Mean	0.00092876

True	667
(Missing)	717494

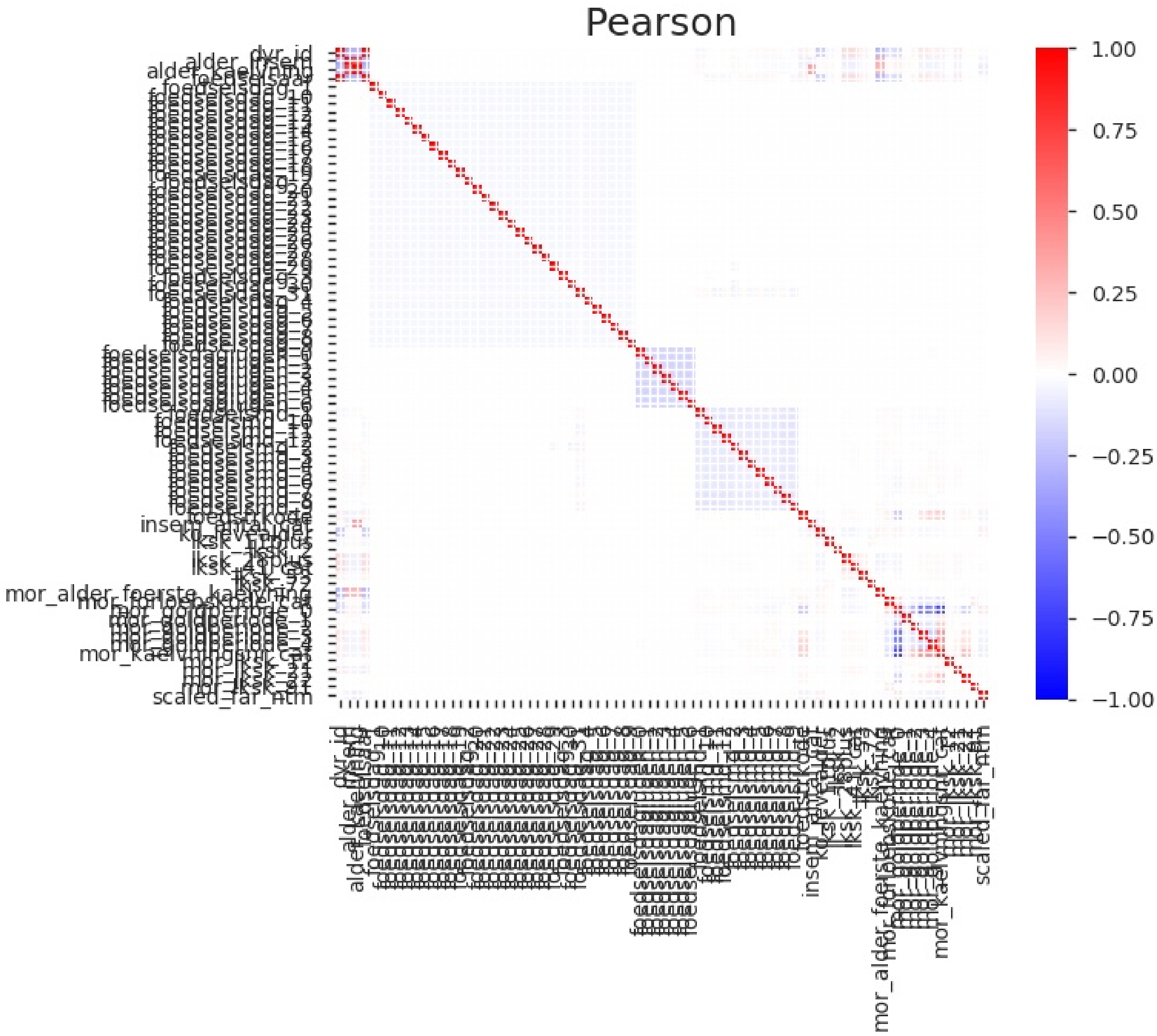
Toggle details

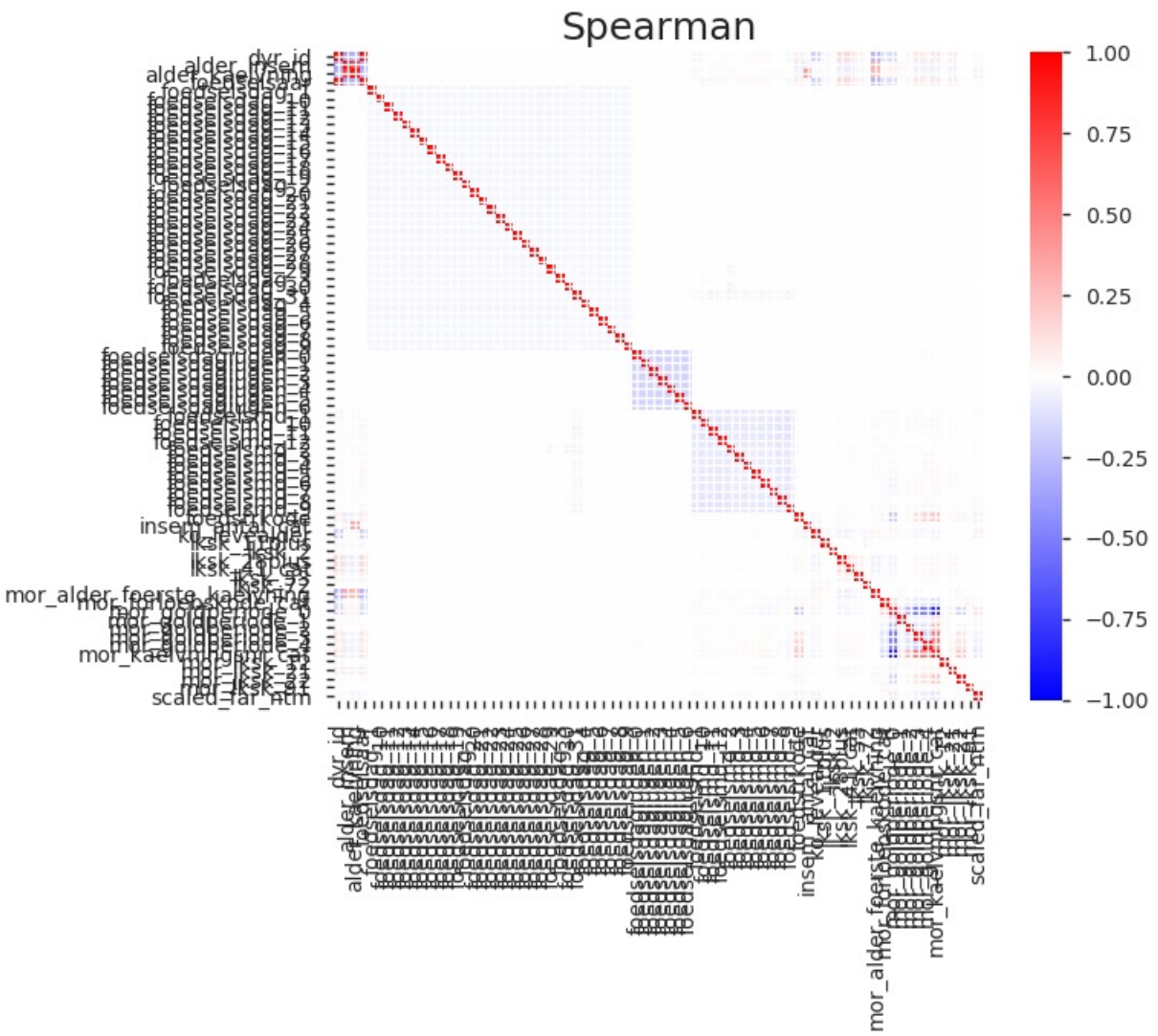
scaled_far_ntm
Numeric

Distinct count	516
Unique (%)	0.1%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	-0.12862
Minimum	-17
Maximum	10.091
Zeros (%)	8.1%

Toggle details

Correlations


Pearson
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
dvr_id
mor_alder_foerste_kaelvning
scaled_far_ntm



Sample

dyr_id	alder_insem	alder_kaelvning	foedselsaar	foedselsdag_1	foedselsdag_10	foedselsdag_11	foedselsdag_12	foedselsdag_13	foedselsdag_14	foe
1001091827	0.010667	-1.173150	0	True	False	False	False	False	False	Fal
1001092285	0.304831	-0.091944	0	True	False	False	False	False	False	Fal
1001092634	-0.913850	-0.897332	0	False	False	False	False	False	False	Fal
1001092700	-0.213459	-0.522220	0	True	False	False	False	False	False	Fal
1001092724	0.949192	2.864824	0	False	False	False	False	False	False	Fal