

Data Science Project

Team nr: 13	Student 1: Gonçalo Gonçalves IST nr: 99226
	Student 2: José Cruz IST nr: 99260
	Student 3: Jorge Santos IST nr: 99258
	Student 4: Matilde Heitor IST nr: 99284

The present document presents a template for the Data Science Project report. It specifies the mandatory format and suggests the structure to follow. All text with grey background shall be replaced with the analysis made over the datasets. Put your charts in the `images` folder, and set the name of the file in the `includegraphics` command, after uncommenting it.

CLASSIFICATION

1 DATA PROFILING

May be used to describe any useful observation about the data, and that was used in the current project. An example is the use of any domain knowledge to process the data or evaluate the results. **Shall not exceed 200 characters.**

For the second dataset, the services domain one, we didn't do much processing prior to the study of the data for most forms of analysis. We simply noticed that the `age` variable had values with the character "_" which we removed for it to become a numeric variable as it is. There were several anomalies in the values for some of the variables. We decided to keep these values for the profiling.

Data Dimensionality

Shall contain all relevant information and charts respecting to the data dimensionality perspective, such as the number of records and number of dimensions, and their impact on the following analysis. **Shall not exceed 500 characters.**

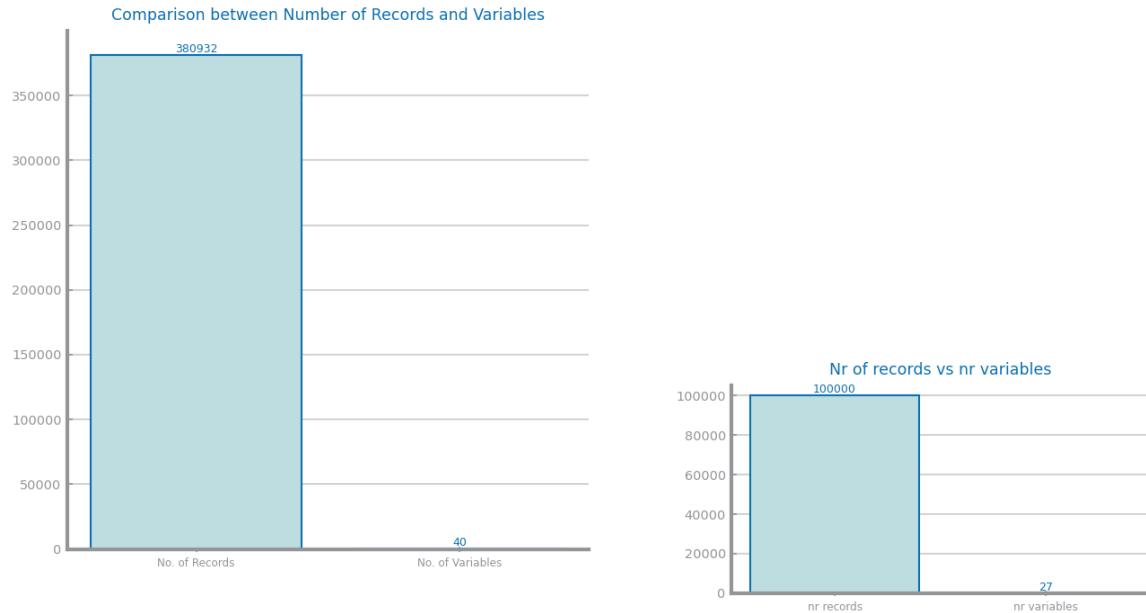


Figure 1: Nr Records x Nr variables for dataset 1 (left) and dataset 2 (right)

We can see that for both datasets there are much more records than variables, avoiding the curse of dimensionality.

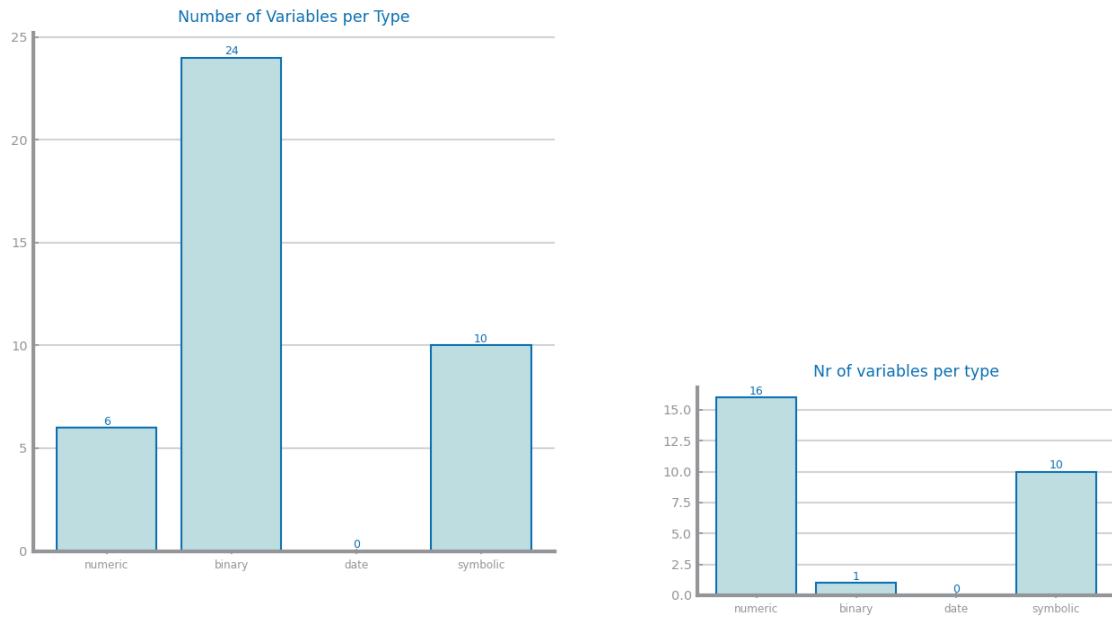


Figure 2: Nr variables per type for dataset 1 (left) and dataset 2 (right)

Neither dataset has *date* variables, meaning that there is no time frame associated with them or the records. The dataset about health has predominantly symbolic variables, especially symbolic. This is expected as it is harder to quantify clinical observations about the state of a person. In contrast, the services dataset has mainly numerical variables. These offer

more precision and are more easily obtained in the financial context the data is in.

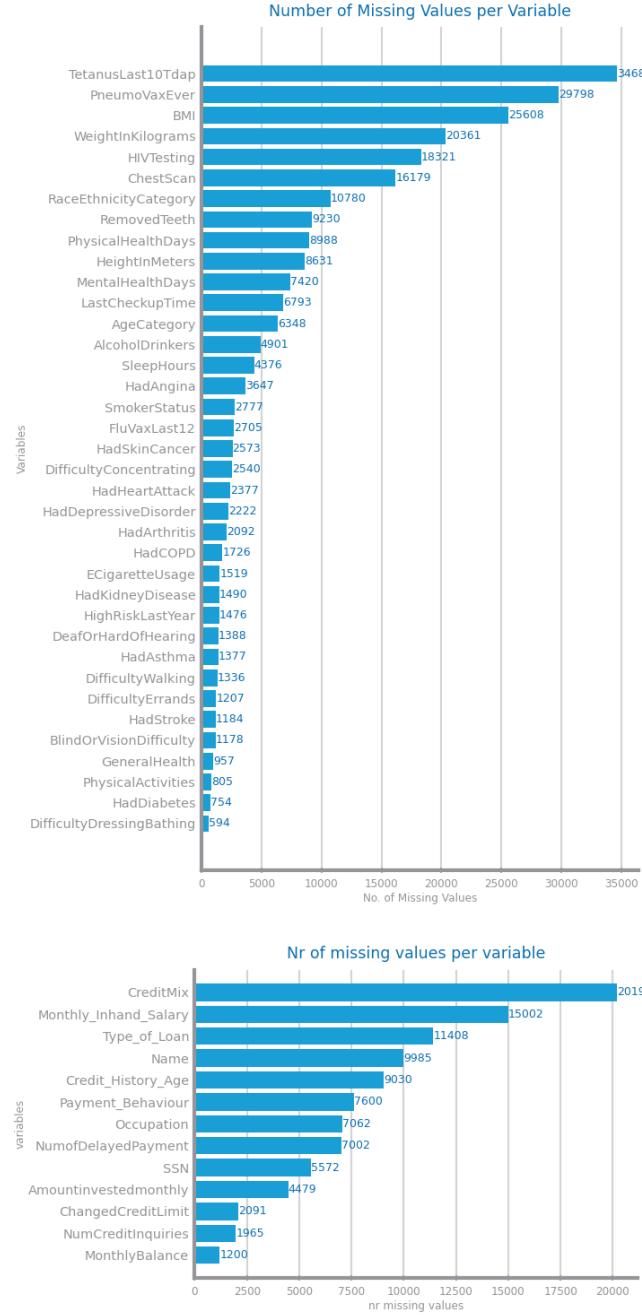


Figure 3: Nr missing values for dataset 1 (left) and dataset 2 (right)

For the health domain dataset, the number of records is 380,932. The variable with the most missing values has less than 10% of missing values. This indicates that some sort of imputation might be a viable option to address them. The second dataset has variables with a bigger ratio of missing values. However these ratios don't go over 20%, making it unlikely that it will be better to drop said variables.

Data Distribution

Shall contain all relevant information and charts respecting to the data distribution perspective, such as each variable distribution, type, domain and range. May be used to describe any useful observation about the data, and that was used in the current project. **Shall not exceed 500 characters.**

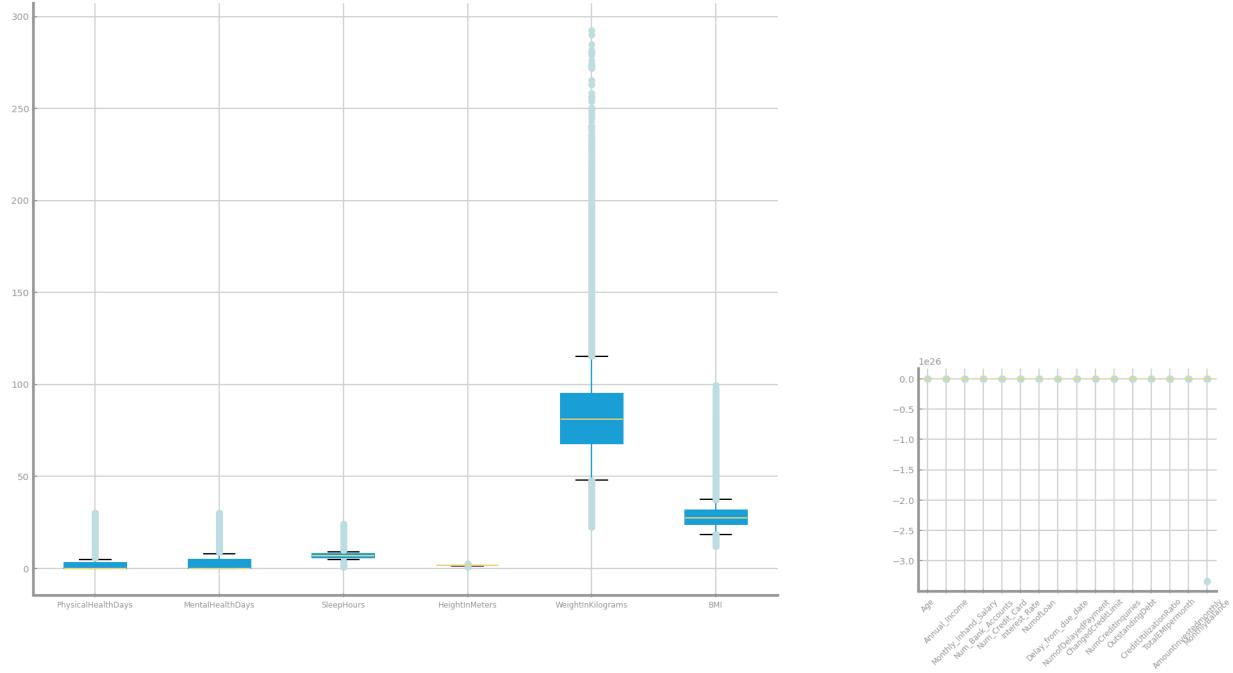


Figure 4: Global boxplots dataset 1 (left) and dataset 2 (right)

Global boxplots are rarely a good option to visualize data as they depend on the fact that all variables have the same scale. Given this, we can say, about the health dataset, that the variable with the biggest spread is *WeightInKilograms*. When examining the services dataset, the only thing we can tell is that the *MonthlyBalance* variable has a huge outlier. This completely distorts the overall representation of the data, rendering the figure nearly useless in terms of insights it could give us.

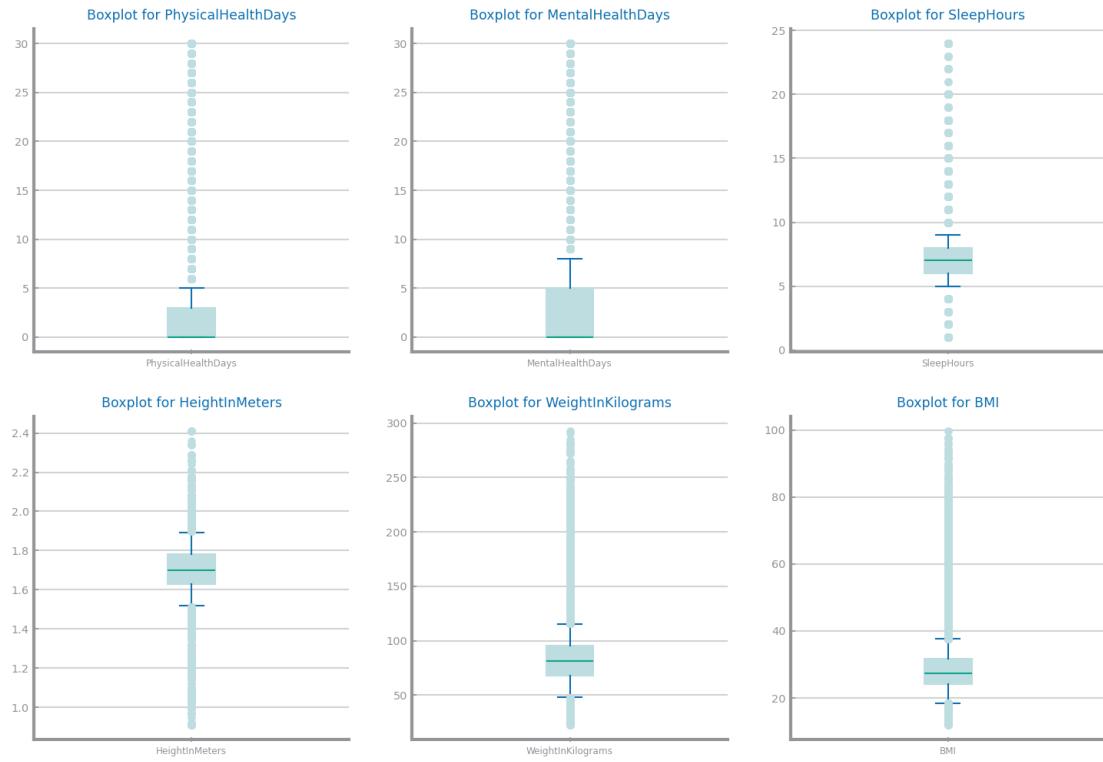
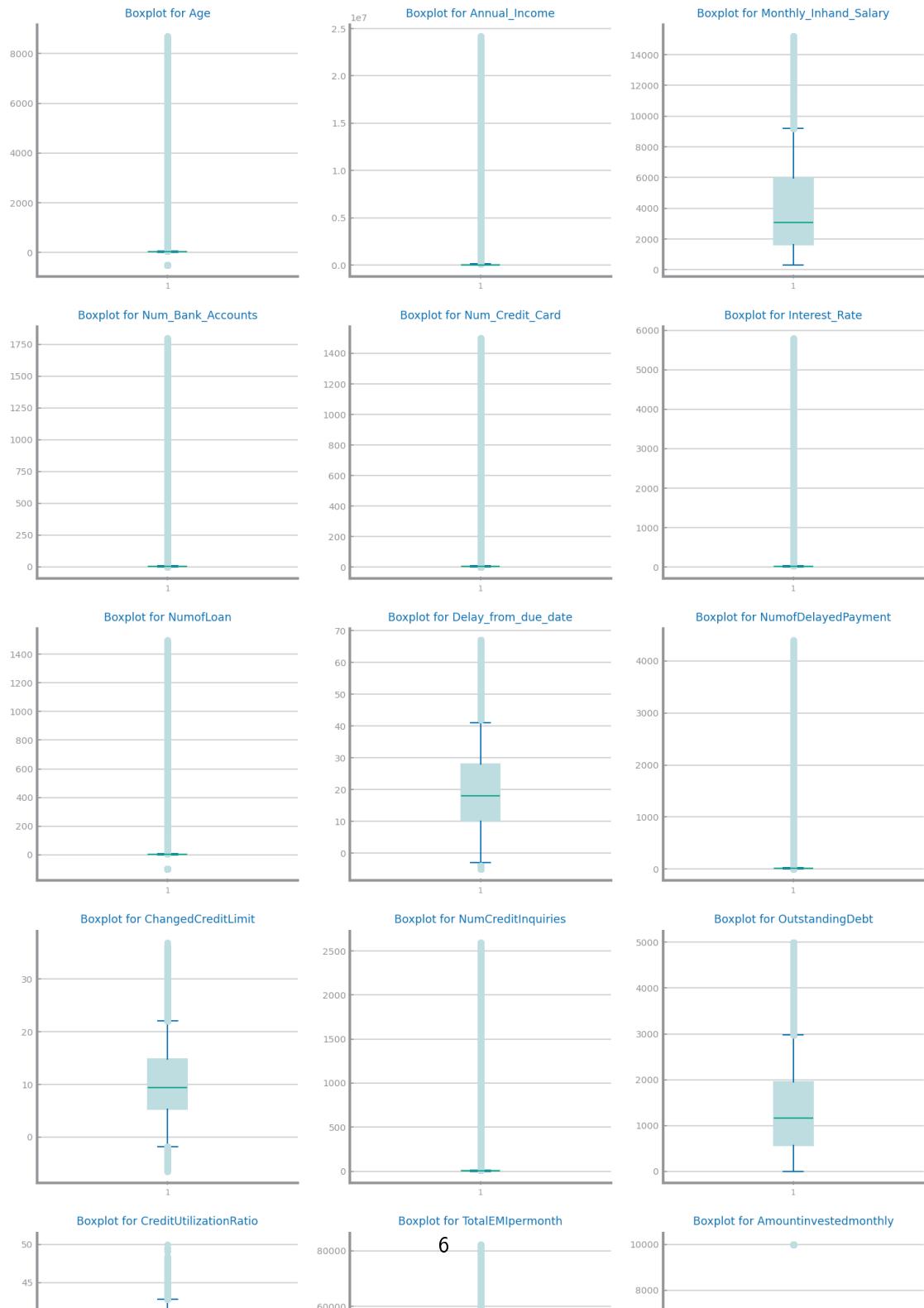


Figure 5: Single variables boxplots for dataset 1

From this we can conclude that all the numeric variables in this first dataset have quite high variability in their values.



Analysing these boxplots, we come to the conclusion that more than half the numeric variables in this dataset have massive variation. This means that a more aggressive approach during data preparation for the outlier values might be needed. Knowing this, it is important to remember that an outlier is a point that is far away from the regular ones. Given that for all but the *MonthlyBalance* variable all the points are continuous, some care must be taken in their treatment.

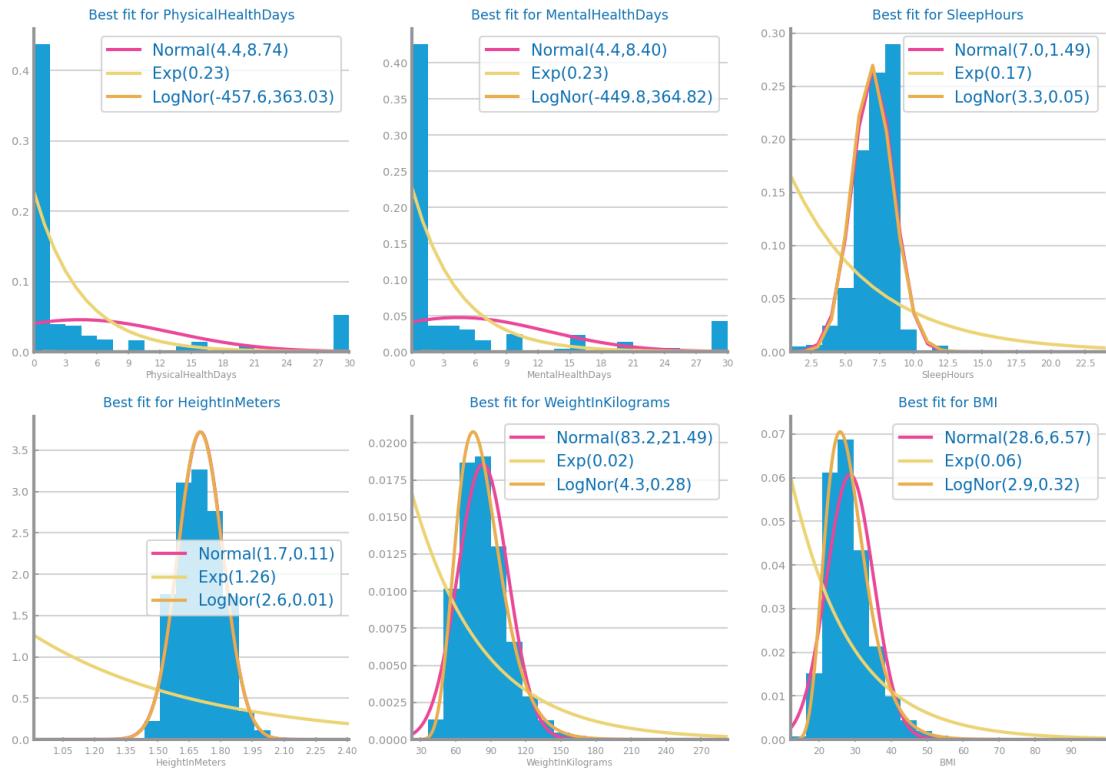
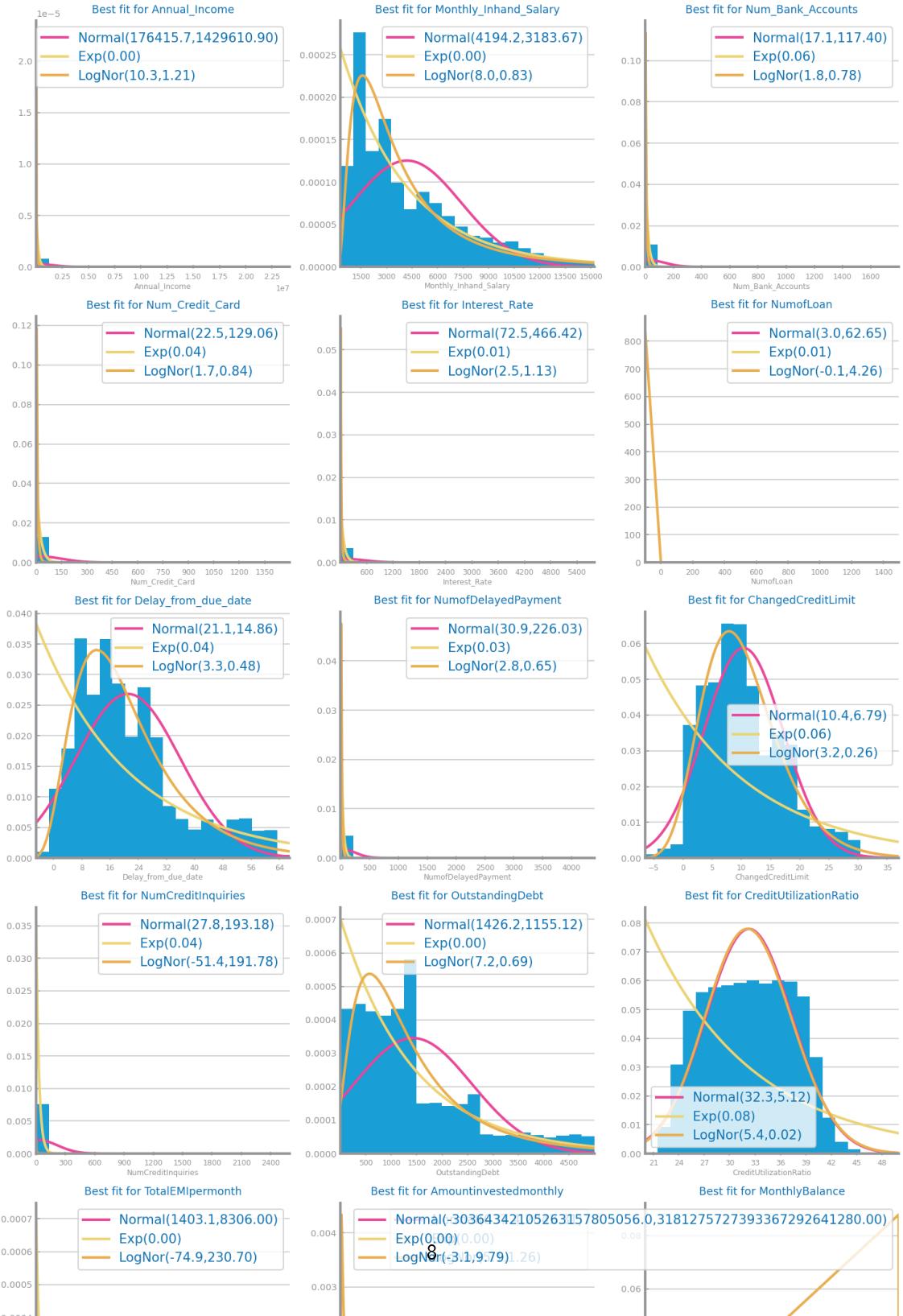


Figure 7: Histograms for dataset 1

From this visual representation we can extract that a majority of numerical variables follow a normal distribution quite well. The others have a shape closer to that of the exponential distribution. Maybe a normalization of the last would be advantageous. This would enable the application of models that rely on the data following normal distributions later on.



The second dataset once again suffers from having very distant outliers. The histograms from which we can infer more information either follow a normal distribution or a lognormal one. The latter can easily be transformed to a normal one if needed for some modelation.

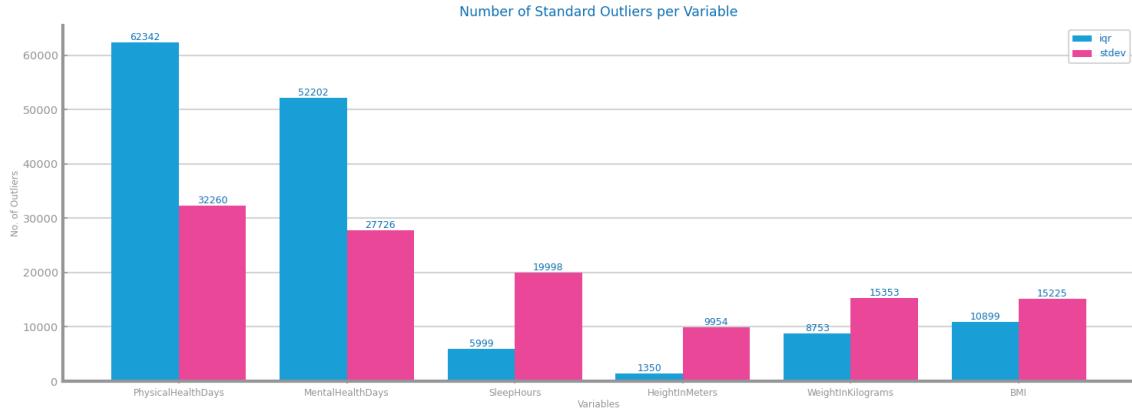


Figure 9: Outliers study dataset 1

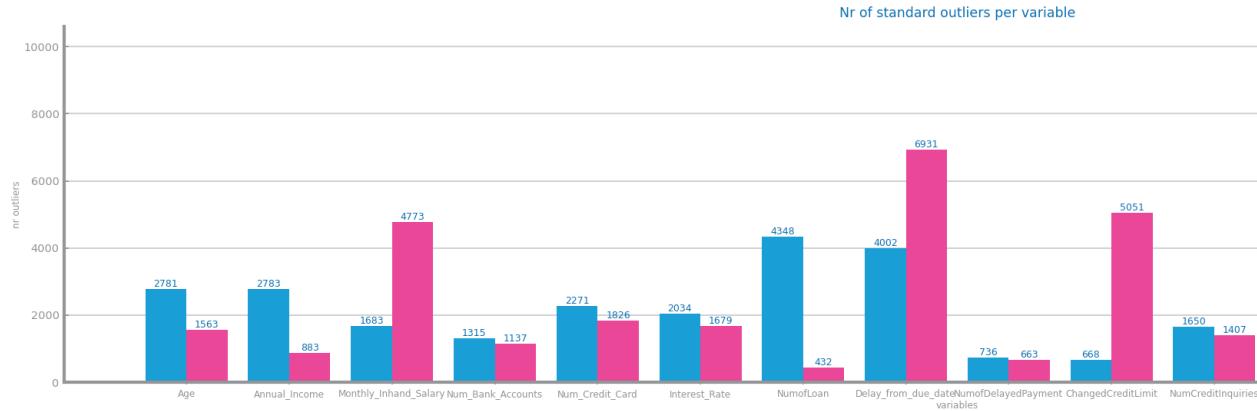


Figure 10: Outliers study dataset 2

For the outliers study per variable we used the IQR factor and the standart deviation one. The IQR factor was of 1.5 and the stddev was of 2. IQR is more robust to extreme values, making it more desiarable for the study of the second dataset. The stddev method should only be used for variables that follow a normal distribution.

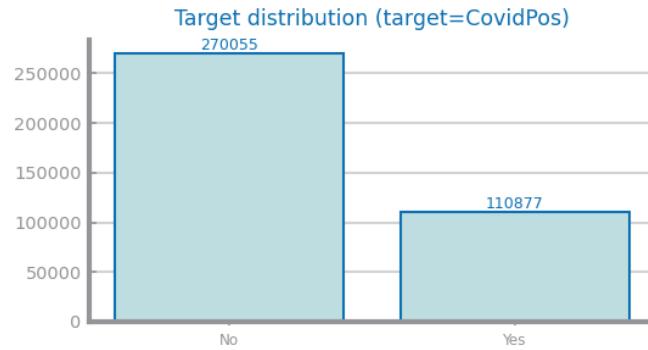


Figure 11: Class distribution for dataset 1



Figure 12: Class distribution for dataset 2

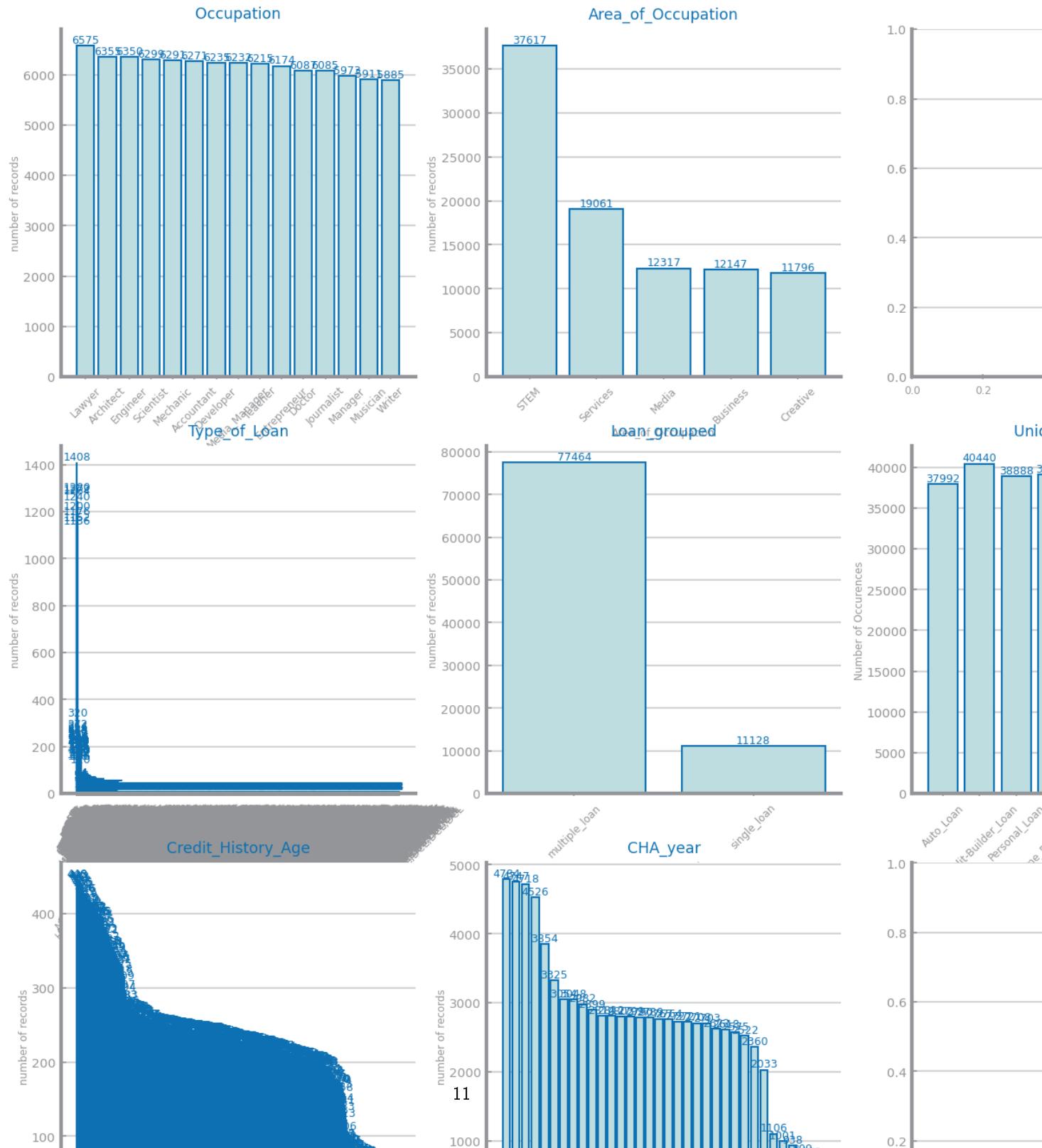
The class distribution is similar in both datasets, a 25%-75%. This means that probably some balancing will be required during the data preparation phase.

Data Granularity

Shall contain all relevant information and charts respecting to the data granularity perspective, such as the impact of different granularities considered for each variable. May present additional taxonomies if needed. **Shall not exceed 500 characters.**

Figure 13: Granularity analysis for dataset 1

Granularity study for Occupation, Credit_History_Age, Payment_Behaviour and Type



We only did a granularity study for symbolic variables with reasonably clear taxonomies that could then aid us in the variable encoding process. Some of the variables clearly had too much different values, this study allowed us to know if descending in the taxonomy hierarchy would not significantly alter the distribution of these variables. Notable examples of this are the *Type_of_Loan* and *Credit_History_Age* variables.

Data Sparsity

Shall contain all relevant information and charts respecting to the data sparsity perspective, such as domain coverage and correlation among variables. **Shall not exceed 500 characters.** For both datasets, we used variable encoding to transform the symbolic variables into numeric ones for the correlation analysis.

Figure 16: Sparsity analysis for dataset 2

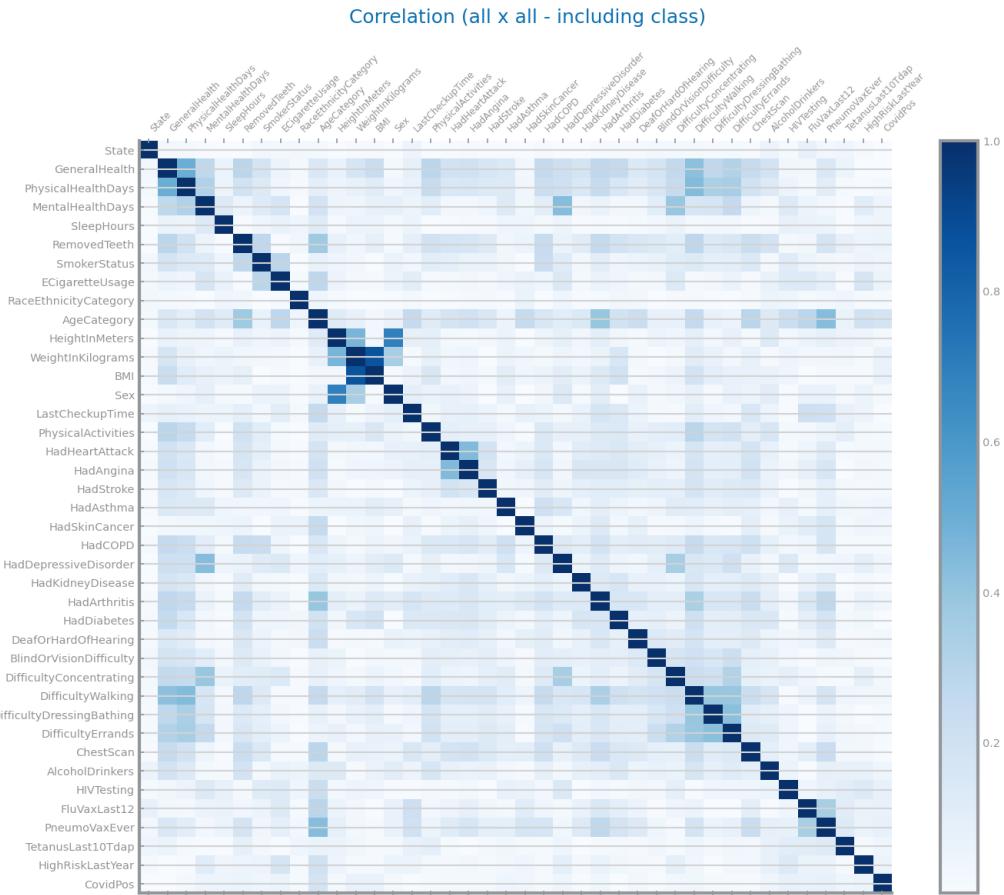


Figure 17: Correlation analysis for dataset 1

We can clearly observe that there are not many variables with high correlations, and the ones with higher ones aren't relevant - The gender of a person and their height, their BMI and their weight.

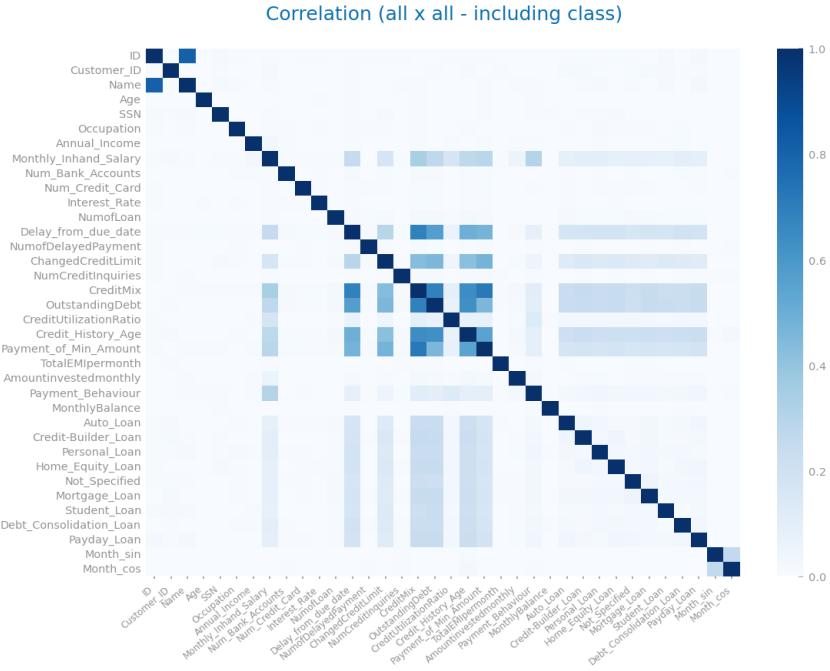


Figure 18: Correlation analysis for dataset 2

In the services dataset we can infer a bit more information, especially about the behaviour of people and what leads them to having a good or bad credit score. Despite this not all high correlations are relevant, per example, the *Name* and *ID* variables.

2 DATA PREPARATION

Variables Encoding

Shall contain all relevant information respecting to the transformation of variables. The list of variables under each one of the transformations, shall be presented. If not applied explain the reason for that, based on data characteristics. **Shall not exceed 500 characters for each dataset.**

Dataset 2 - Services

In this step we converted all non numerical variables into numerical ones. For the *ID*, *Customer_ID*, *SSN* and *Name* were trasnformed into numbers, preserving their uniqueness. For variables that had order to their values we used ordinal linear encoding, these were *Credit_Score*, *Credit_Mix*, *Payment_of_Min_Amount*, *Credit_History_Age* and *Payment_Behaviour*. For the *Month* we used cyclic encoding. We reduced the granularity of the *Occupation* variable, turning it more broad. For the *Type_of_Loan* variable we used dummification.

Missing Value Imputation

Shall contain all relevant information and charts respecting to missing values imputation, such as the choices made and the impact of the different approaches on modelling results. Shall also clearly reveal the approach selected to proceed

with the processing. If not applied explain the reason for that, based on data characteristics. **Shall not exceed 500 characters.**

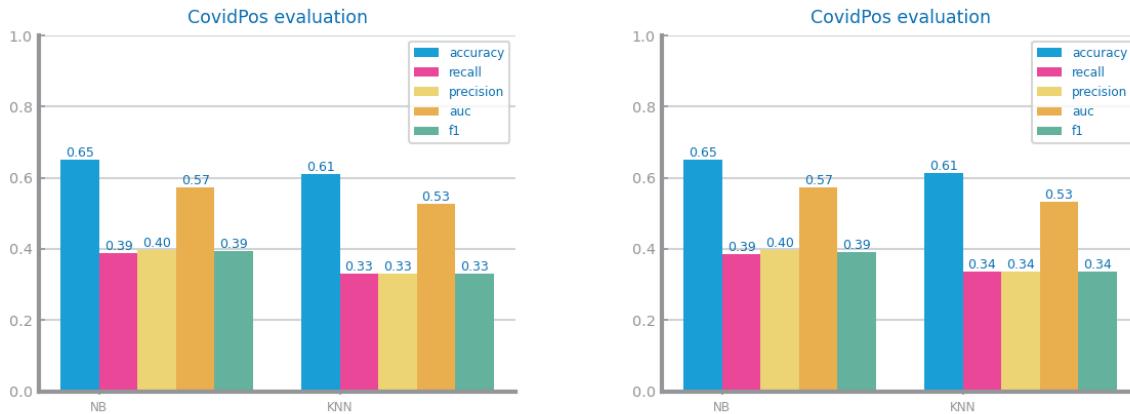


Figure 19: Missing values imputation results with different approaches for dataset 1

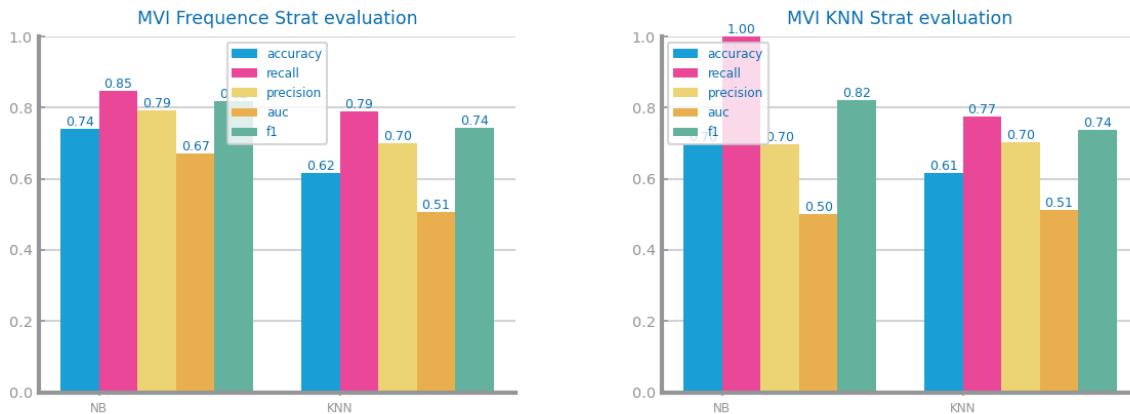


Figure 20: Missing values imputation results with different approaches for dataset 2

Firstly, we drop all the records with more than 90% of missing values, removing 11.482 entries. Then, we applied two different filling strategies, frequent and KNN, both had very similar results after evaluation. However, the frequent strategy came out on top and we chose it to fill the rest of the missing values.

Outliers Treatment

Shall contain all relevant information and charts respecting to outliers imputation, such as the choices made and the impact of the different approaches on modelling results. Shall also clearly reveal the approach selected to proceed with the processing. If not applied explain the reason for that, based on data characteristics. **Shall not exceed 500 characters.**

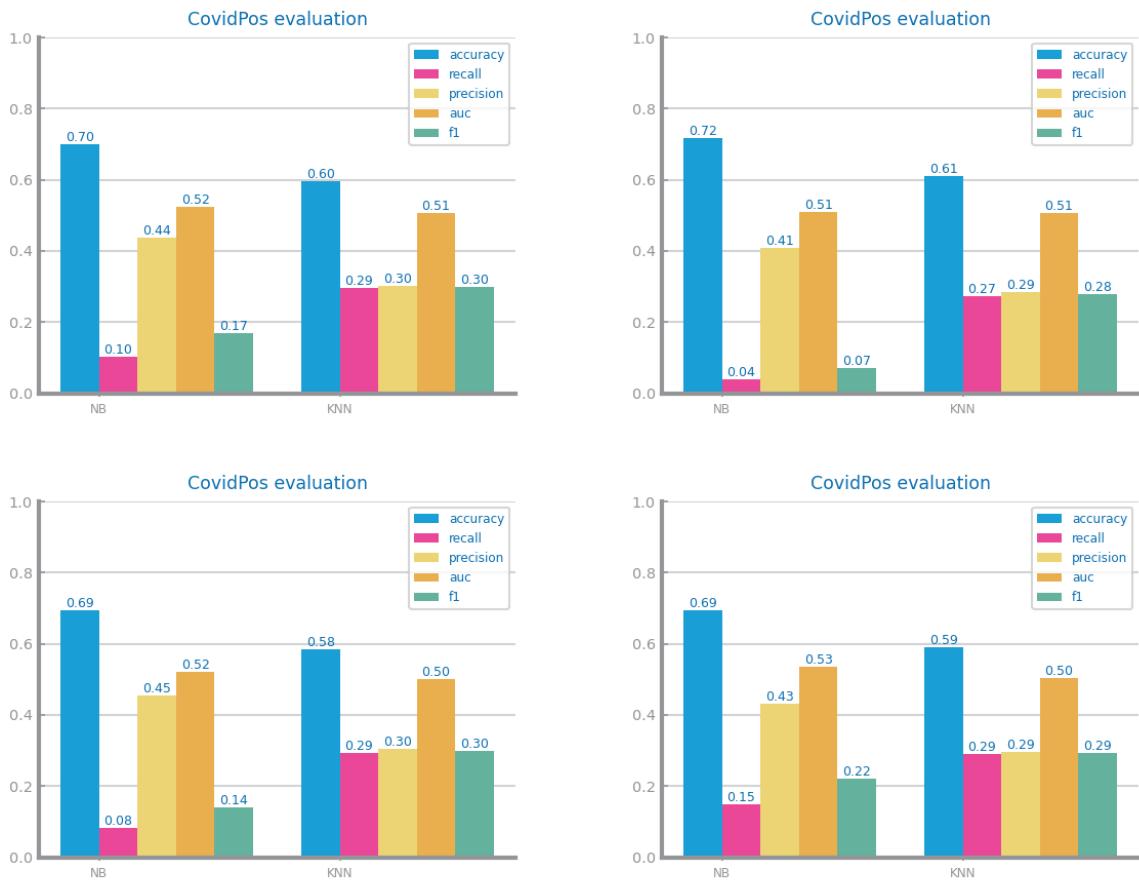


Figure 21: Outliers imputation results with different approaches for dataset 1

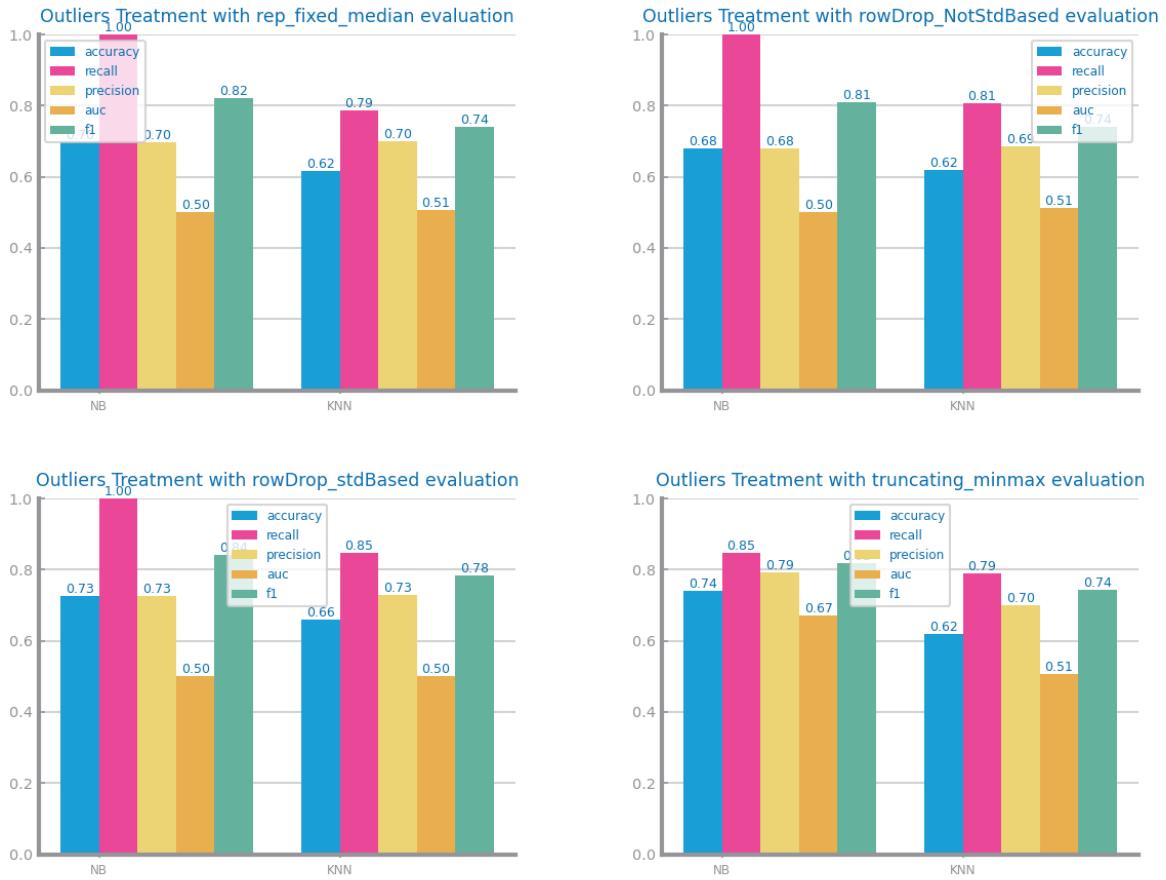


Figure 22: Outliers imputation results with different approaches for dataset 2

We applied four different strategies: replacing, truncating or dropping (either std based or iqr based) them. They all attained similar results but the best evaluation values were the result of dropping std based, therefore this was our choice to move forward.

Scaling

Shall contain all relevant information and charts respecting to scaling transformation, such as the choices made and the impact of the different approaches on modelling results. Shall also clearly reveal the approach selected to proceed with the processing. If not applied explain the reason for that, based on data characteristics. **Shall not exceed 200 characters.**

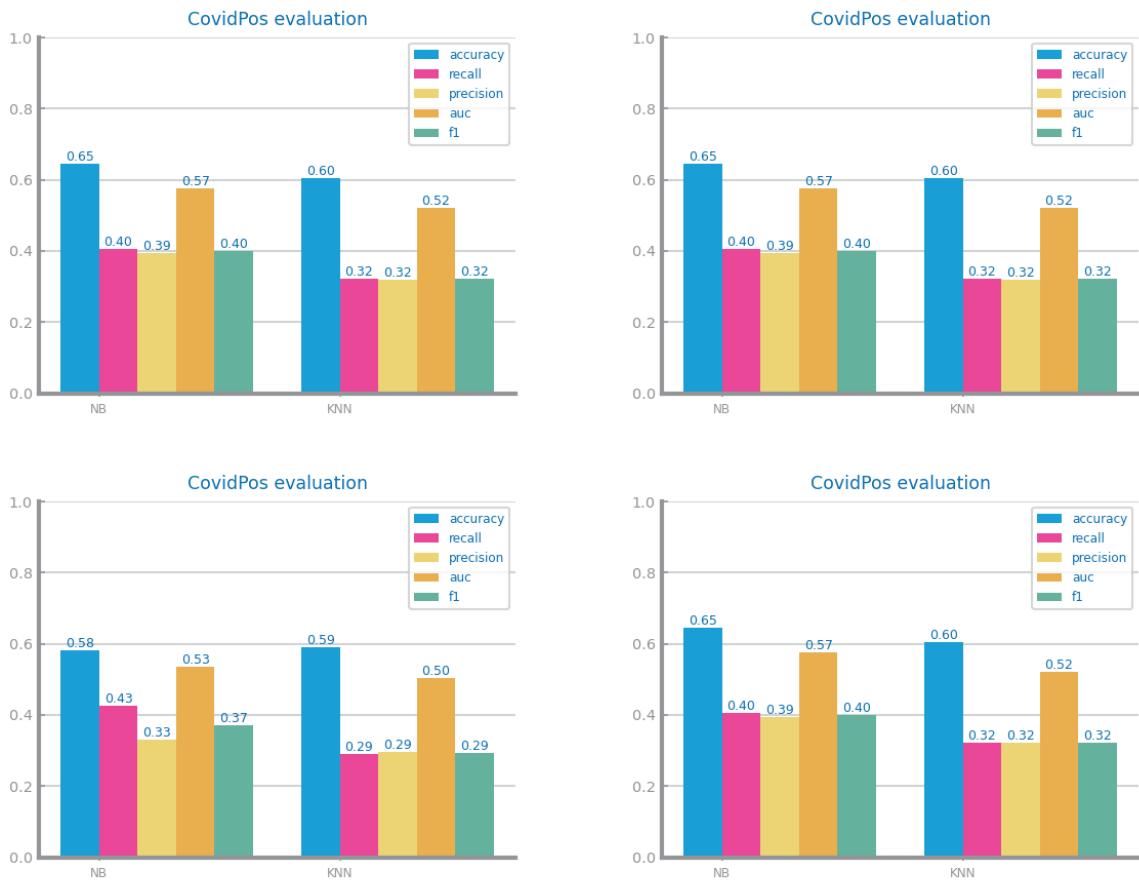


Figure 23: Scaling results with different approaches for dataset 1

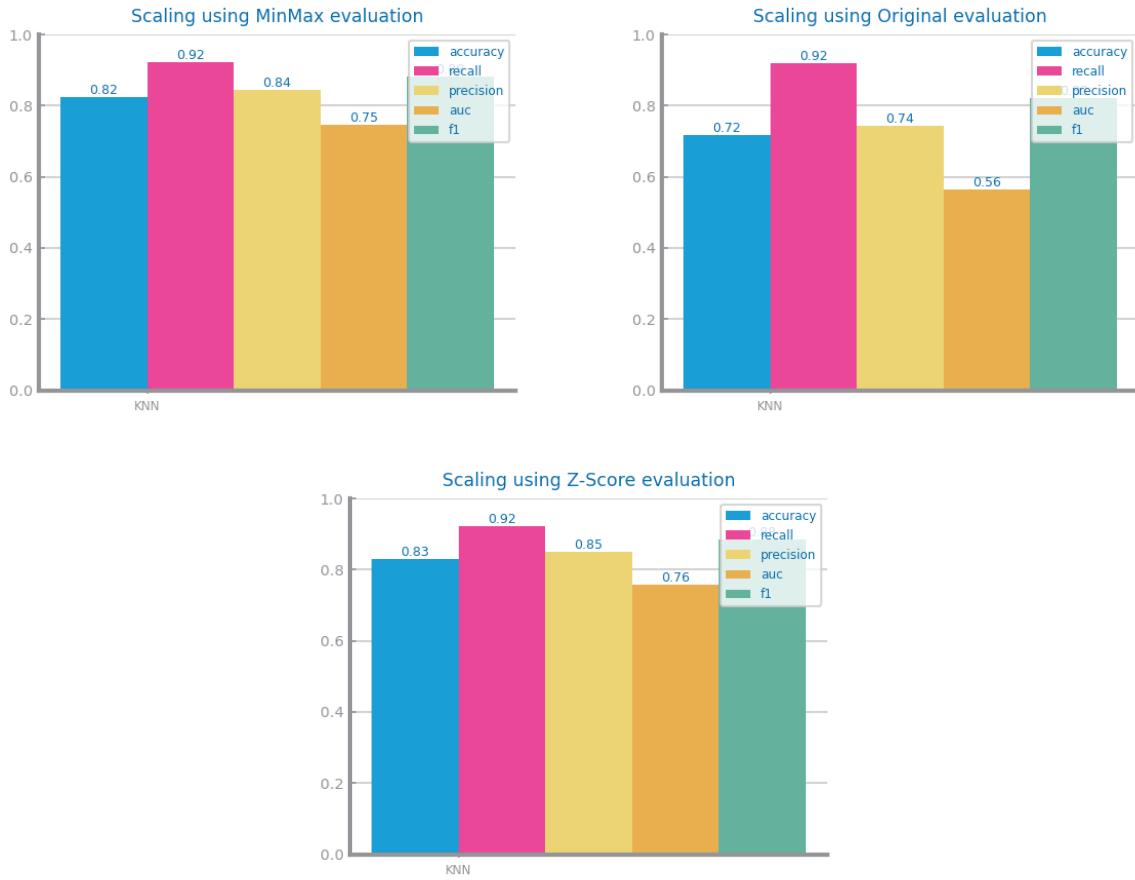


Figure 24: Scaling results with different approaches for dataset 2

In scaling, we compared the a MinMax and Z-Score approaches, seeing which would give us better improvement, if any. Z-Score had the best evaluation results therefore it was chosen to proceed with the data preparation.

Balancing

Shall contain all relevant information and charts respecting to balancing transformation, such as the choices made and the impact of the different approaches on modelling results. Shall also clearly reveal the approach selected to proceed with the processing. If not applied explain the reason for that, based on data characteristics. **Shall not exceed 500 characters.**

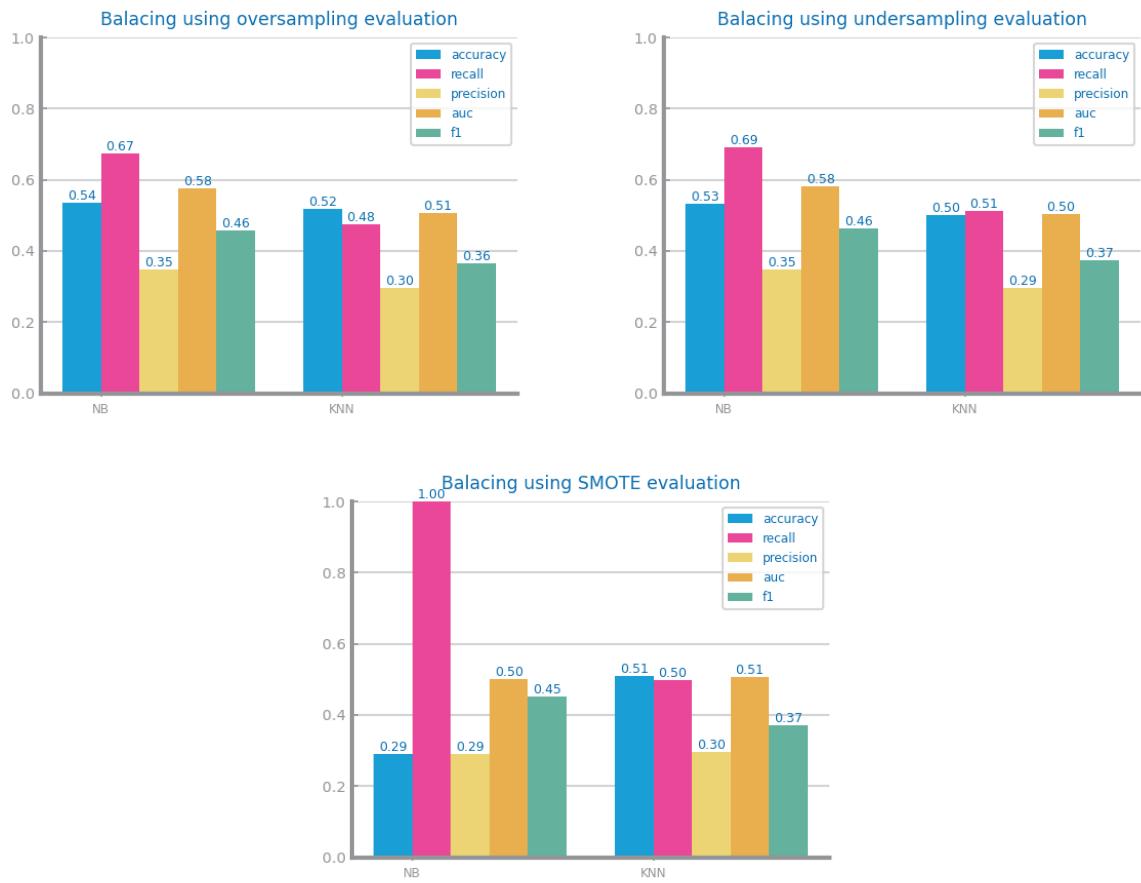


Figure 25: Balancing results with different approaches for dataset 1

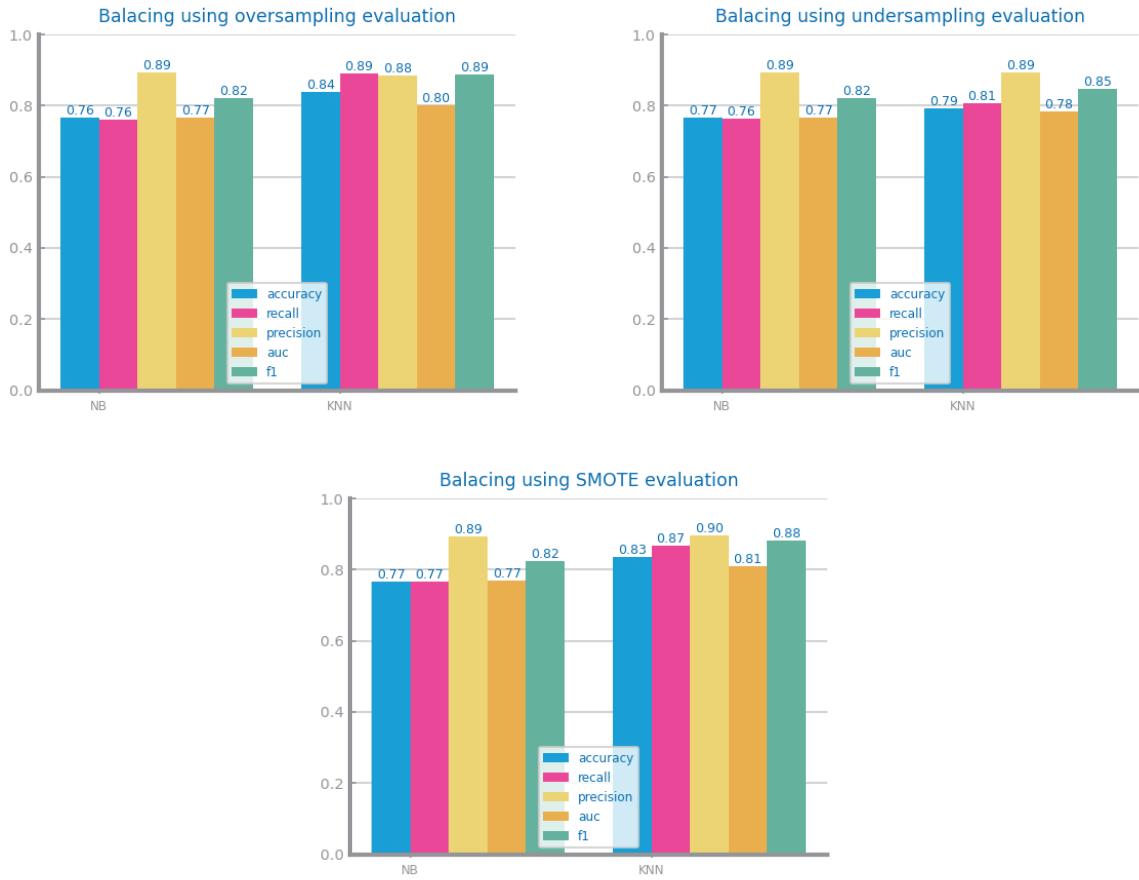


Figure 26: Balancing results with different approaches for dataset 2

Feature Selection

Shall contain all relevant information and charts respecting to feature selection based on filtering out redundant (based on correlation) and relevant (based on variation) variables. The different choices and their impact on the modelling results shall be presented and explained. Should also clearly reveal the approach selected to proceed with the processing. All explanations shall be based on data characteristics. **Shall not exceed 500 characters.**

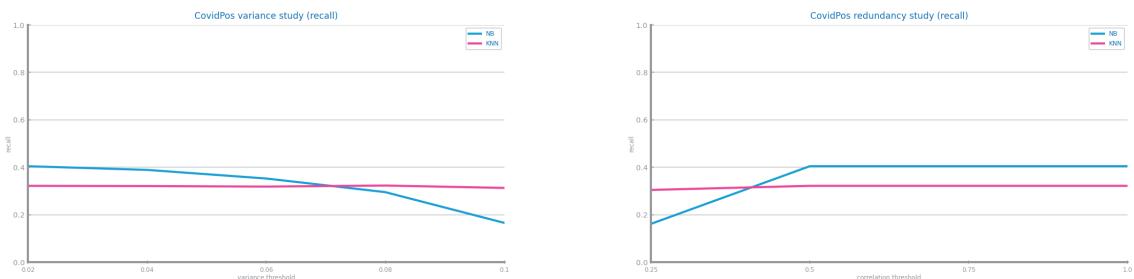


Figure 27: Feature selection of redundant variables results with different parameters for dataset 1



Figure 28: Feature selection of redundant variables results with different parameters for dataset 2

Figure 29: Feature selection of relevant variables results with different parameters for dataset 1 (variance study)

Figure 30: Feature selection of relevant variables results with different parameters for dataset 2 (variance study)

Feature Extraction (optional)

Shall contain all relevant information and charts respecting to feature extraction, in particular PCA. The different choices and their impact on the modelling results shall be presented and explained. **Shall not exceed 200 characters.**

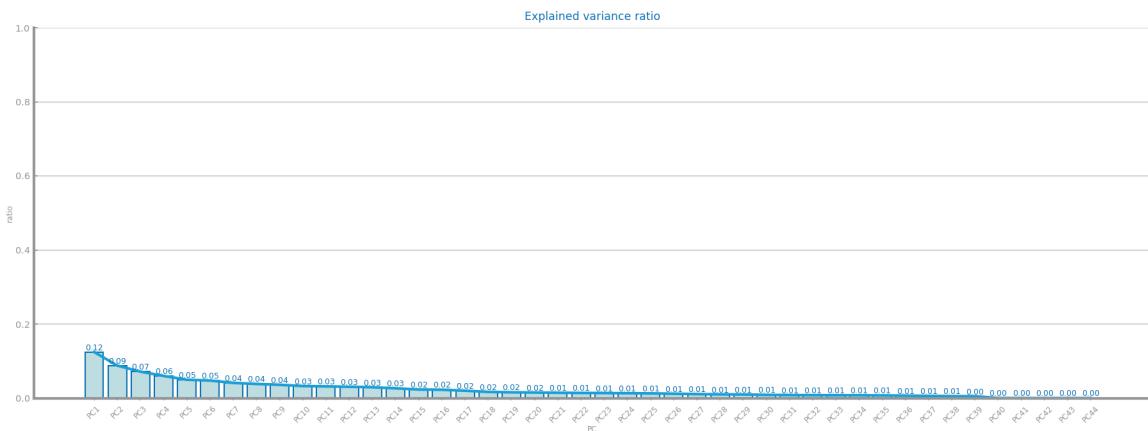


Figure 31: Principal components analysis and feature extraction results for dataset 1

Figure 32: Principal components analysis and feature extraction results for dataset 2

Additional Feature Generation (optional)

Shall contain all relevant information and charts respecting to feature generation. The different choices and their impact on the modelling results shall be presented and explained. Shall summarise all variables generated and the formula used to derive them (in a table). **Shall not exceed 200 characters.**

Figure 33: Feature generation results for dataset 1

Figure 34: Feature generation results for dataset 2

3 MODELS' EVALUATION

Shall be used to point out any important decision taken during the training, including training strategy and evaluation measures used. **Shall not exceed 500 characters.**

Naïve Bayes

Shall be used to present the results achieved with each one of Naïve Bayes implementations, comparing and proposing explanations for them. If any of the implementations is not used, a justification for it shall be presented. Shall be used to present the evaluation of the best model achieved. **Shall not exceed 300 characters.**

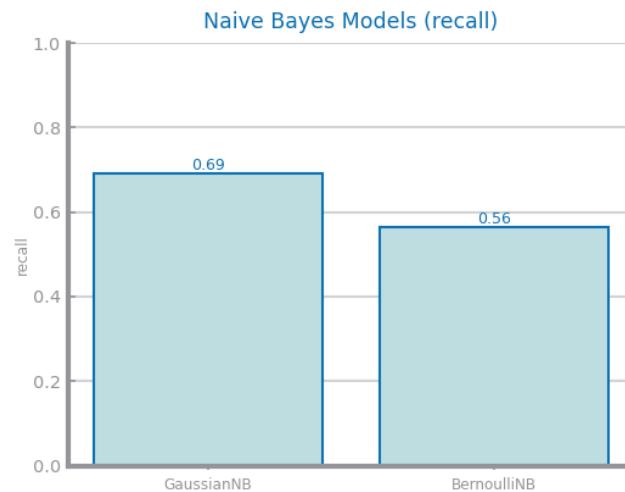


Figure 35: Naïve Bayes alternatives comparison for dataset 1

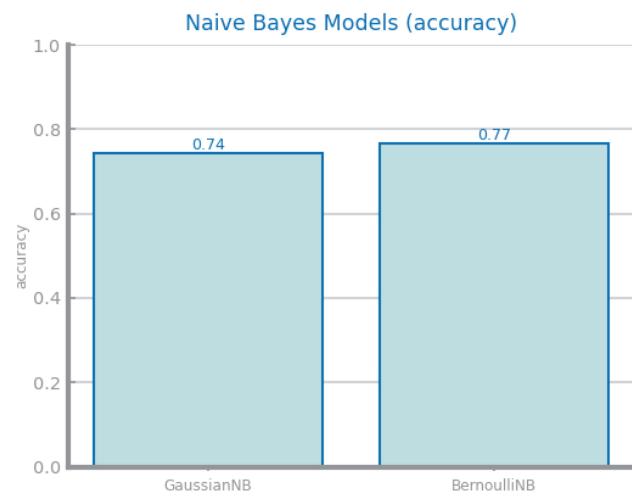
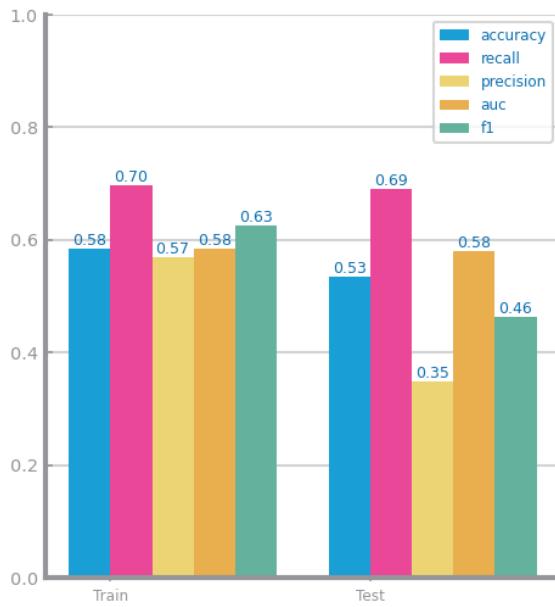
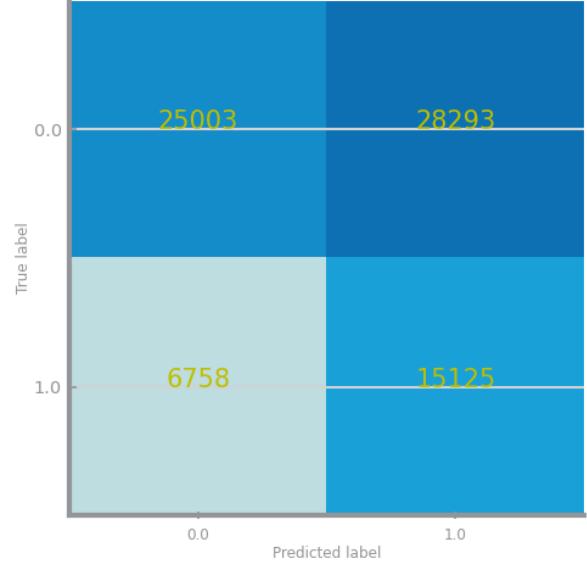


Figure 36: Naïve Bayes alternative comparison for dataset 2

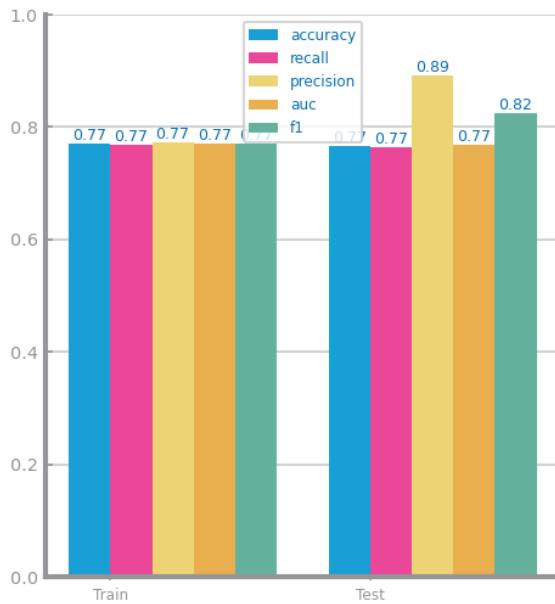
Best recall for GaussianNB



Confusion matrix



Best accuracy for BernoulliNB



Confusion matrix

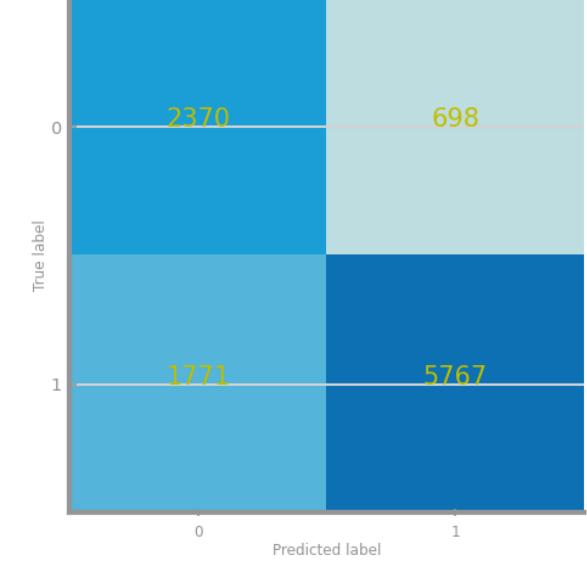


Figure 37: Naïve Bayes best model results for dataset 1 (left) and dataset 2 (right)

KNN

Shall be used to present the results achieved through different similarity measures and KNN parameterisations. The results shall be compared and explanations for them shall be presented. The justification for the chosen similarity measures shall

be presented. Shall be used to address the overfitting phenomenon, studying the conditions under which models face it.
Shall be used to present the evaluation of the best model achieved. **Shall not exceed 500 characters.**

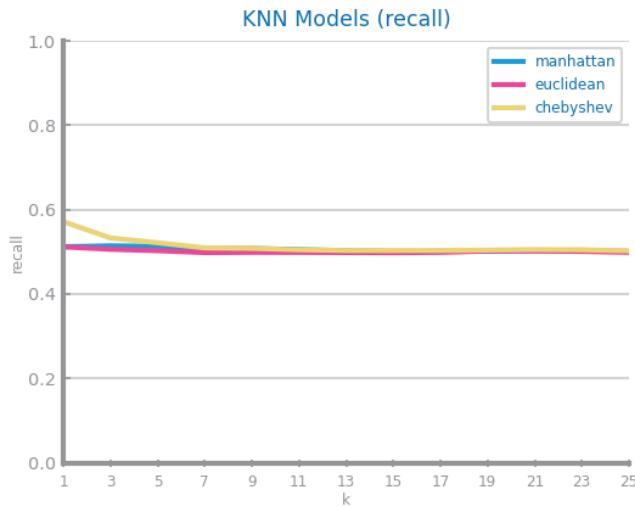


Figure 38: KNN different parameterisations comparison for dataset 1

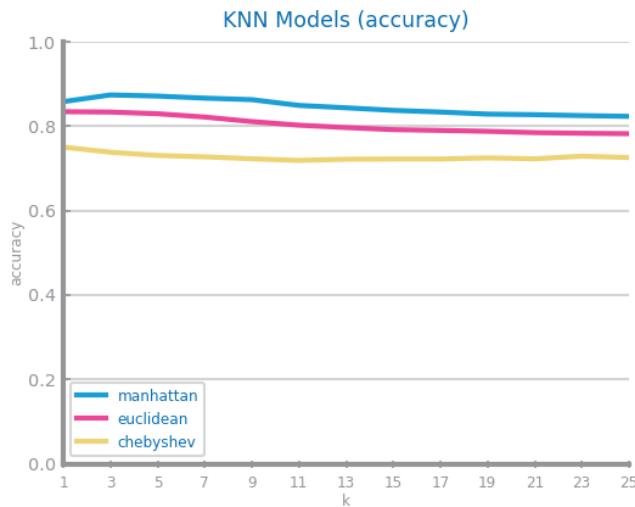


Figure 39: KNN different parameterisations comparison for dataset 2

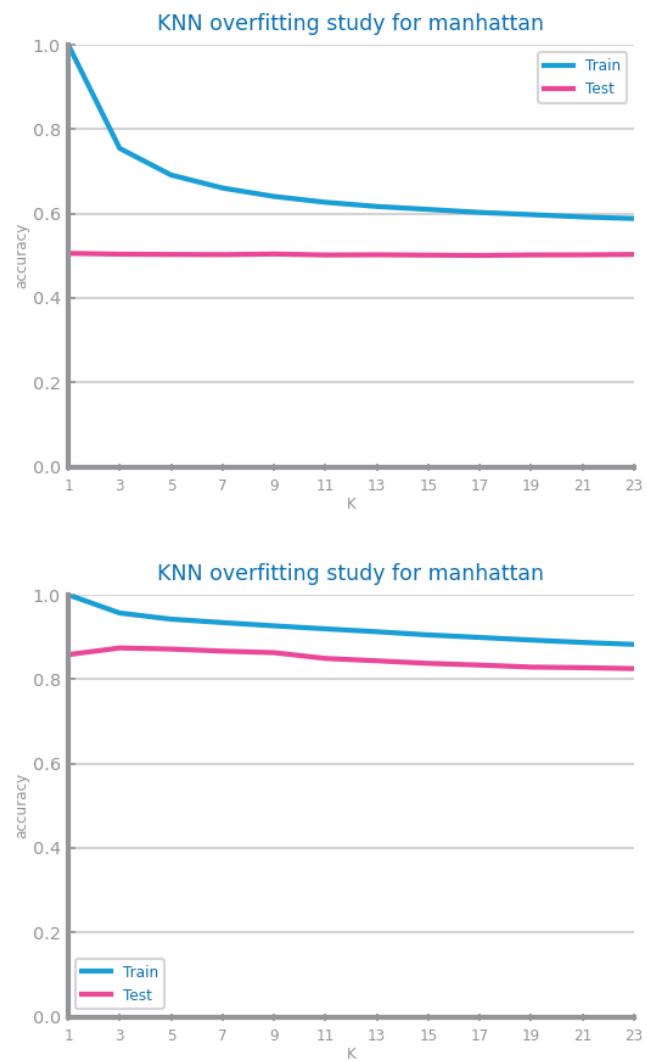


Figure 40: KNN overfitting analysis for dataset 1 (left) and dataset 2 (right)



Figure 41: KNN best model results for dataset 1 (left) and dataset 2 (right)

Decision Trees

Shall be used to present the results achieved through different parameterisations for the train of decision trees. The results shall be compared and explanations for them shall be presented. Shall be used to address the overfitting phenomenon,

studying the conditions under which models face it. Shall be used to present the evaluation of the best model achieved.
Shall be used to present the best tree achieved and its succinct description. **Shall not exceed 500 characters.**

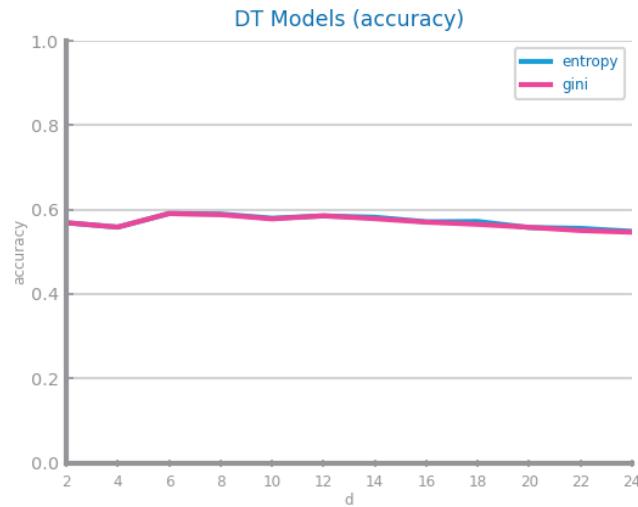


Figure 42: Decision Trees different parameterisations comparison for dataset 1

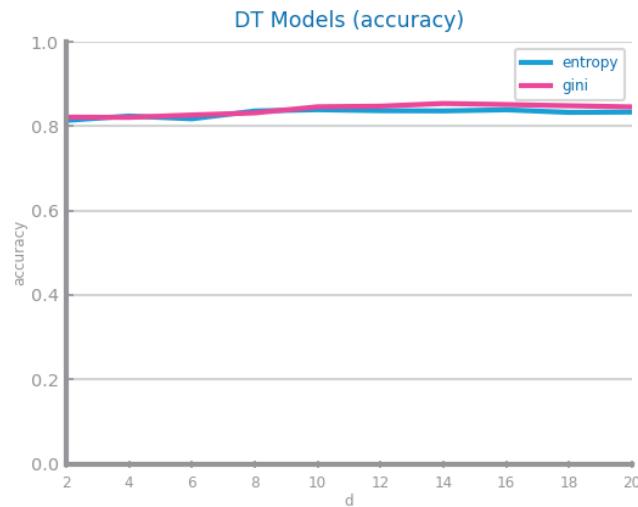


Figure 43: Decision Trees different parameterisations comparison for dataset 2

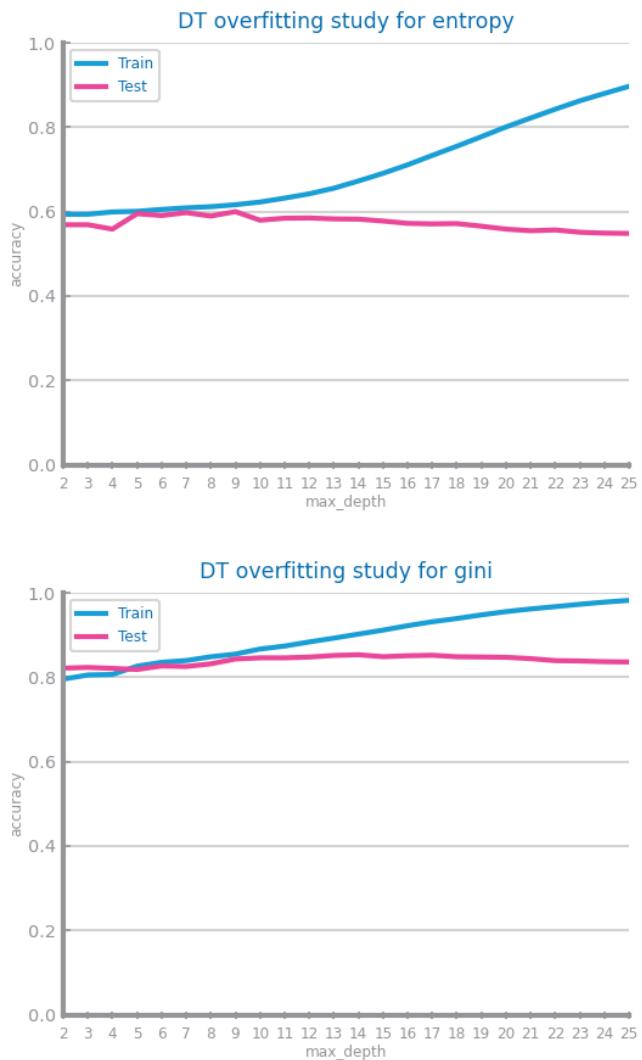
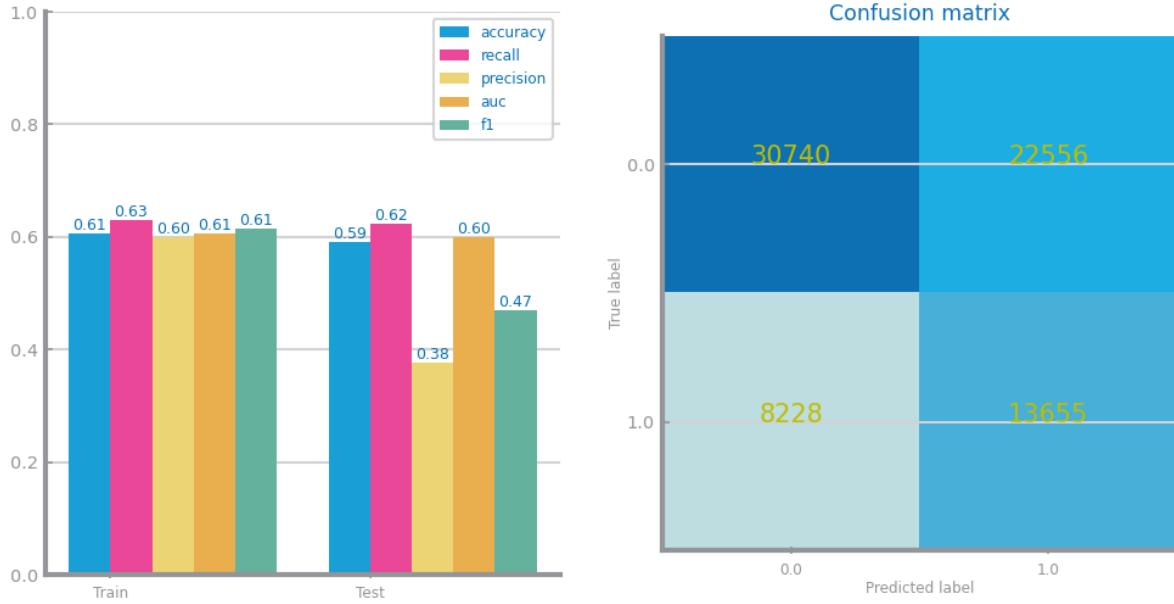


Figure 44: Decision Trees overfitting analysis for dataset 1 (left) and dataset 2 (right)

Best accuracy for DT ('entropy', 6)



Best accuracy for DT ('gini', 14)

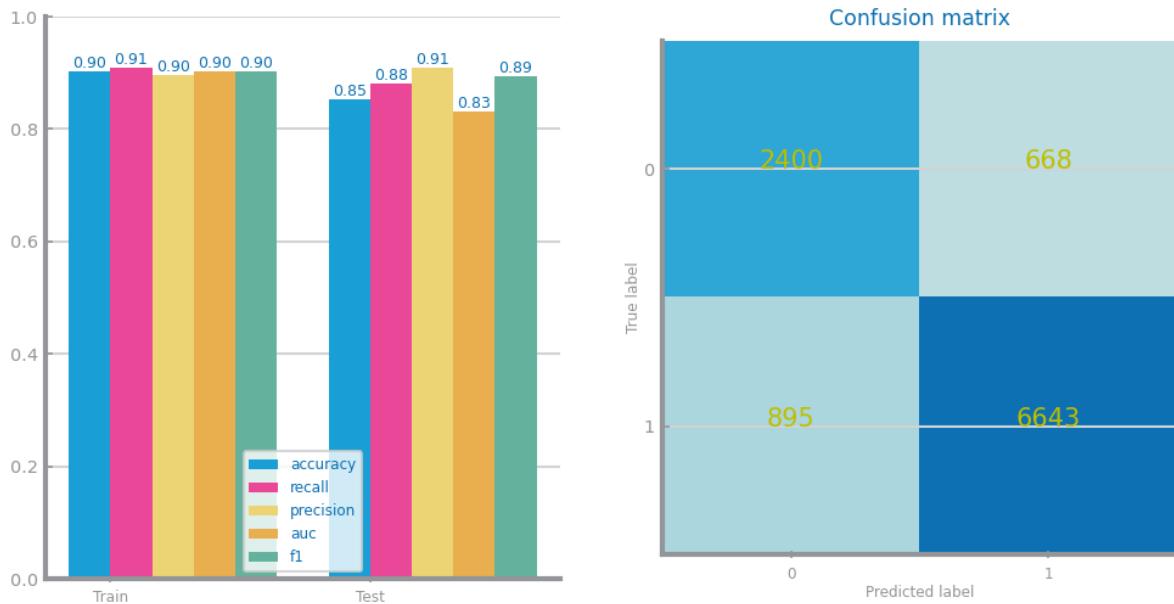


Figure 45: Decision trees best model results for dataset 1 (left) and dataset 2 (right)

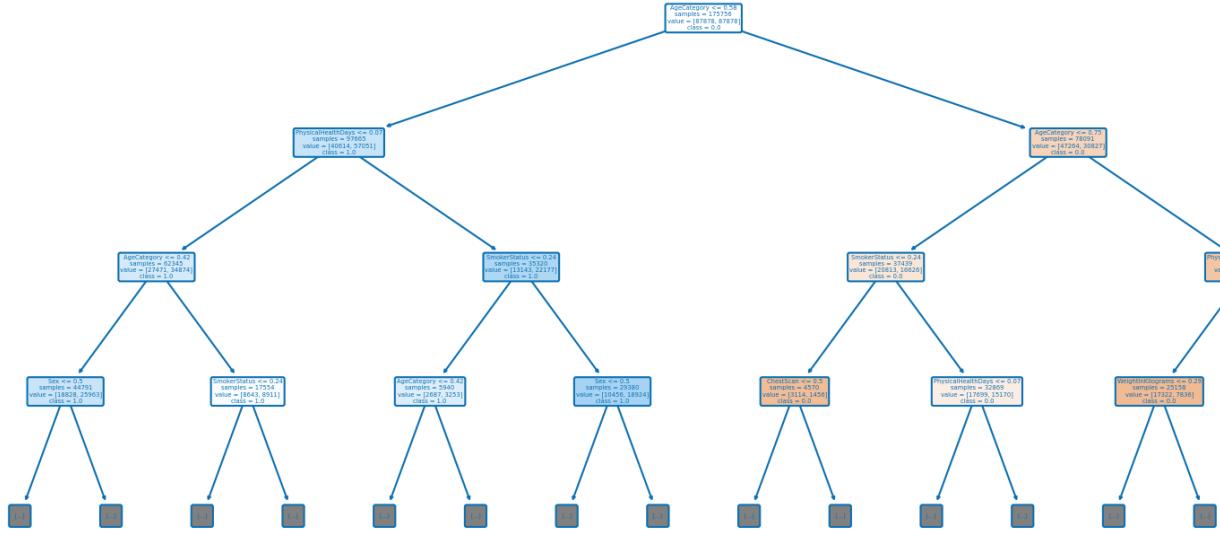


Figure 46: Best tree for dataset 1

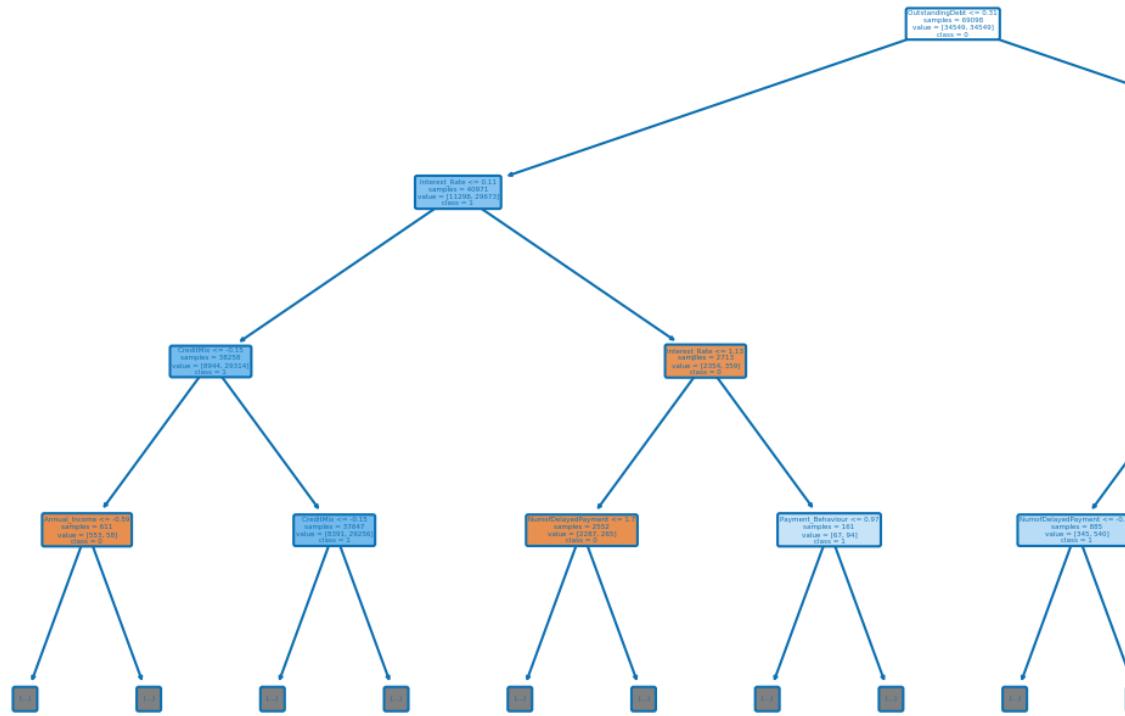
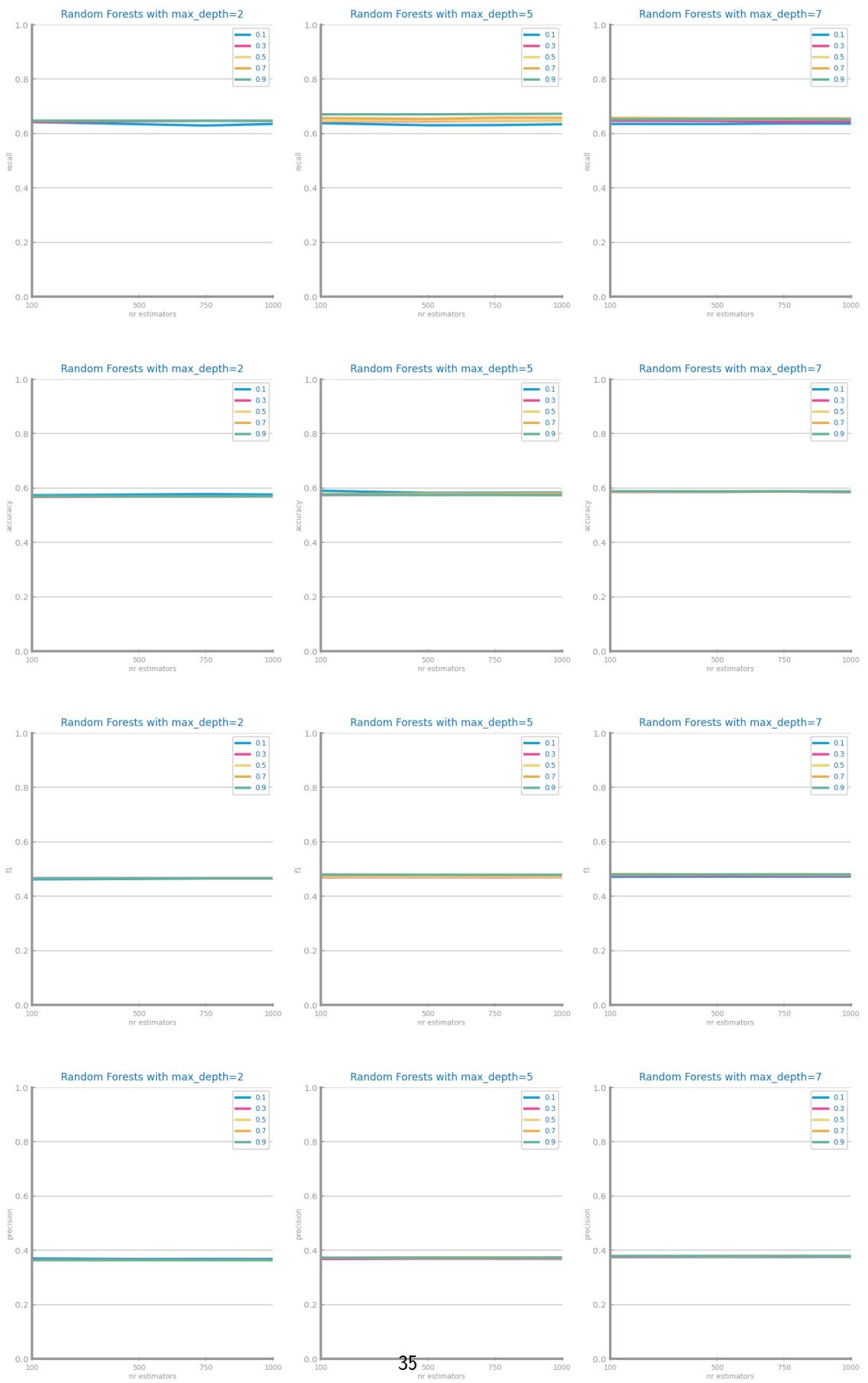
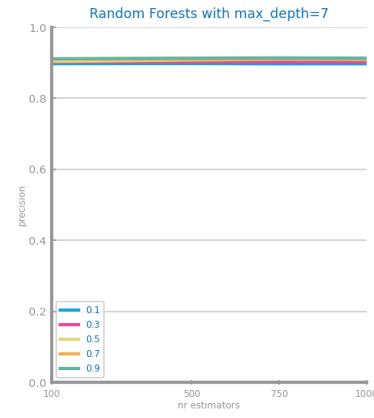
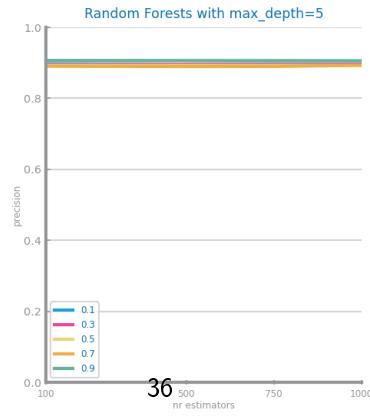
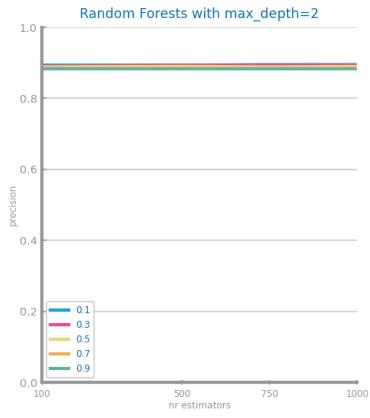
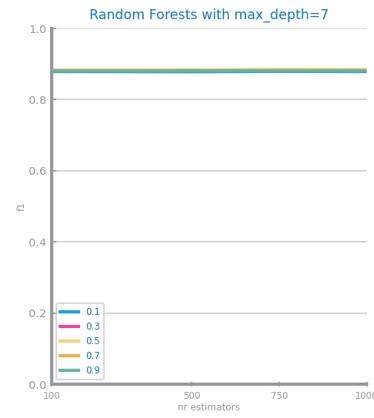
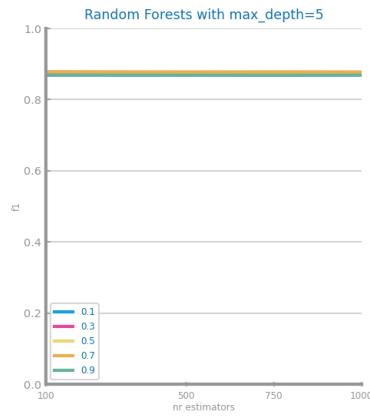
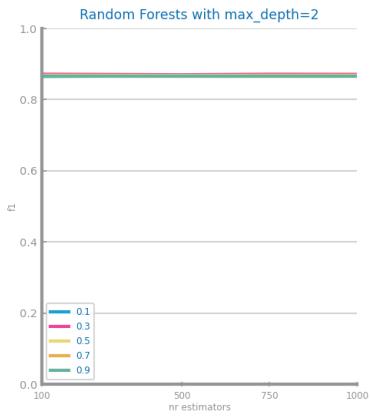
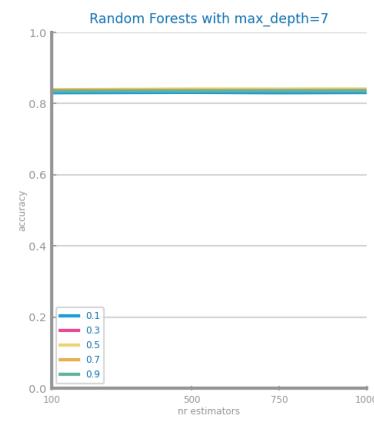
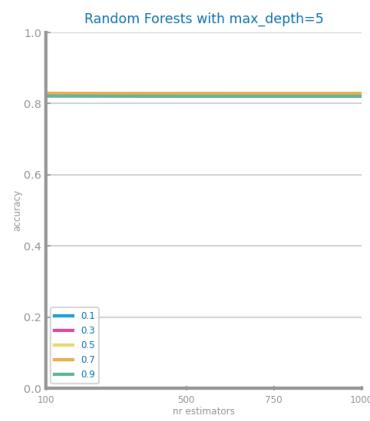
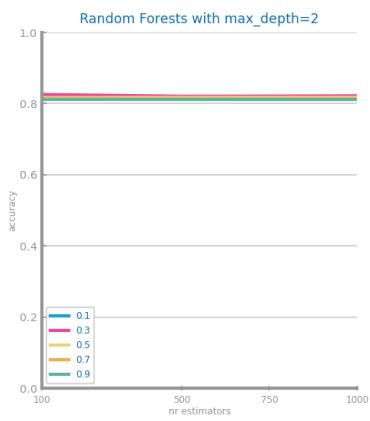
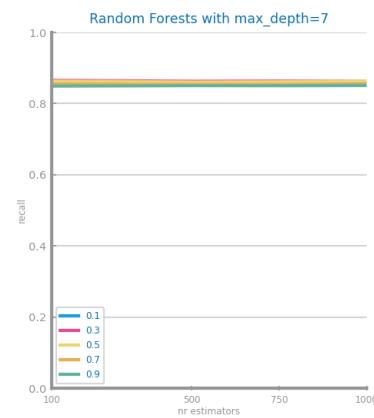
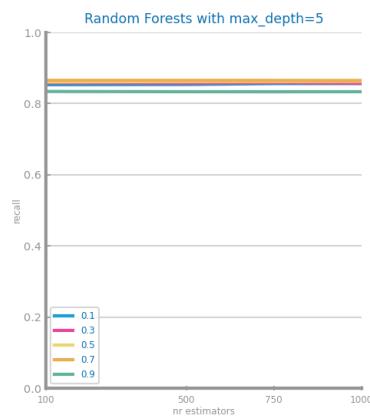
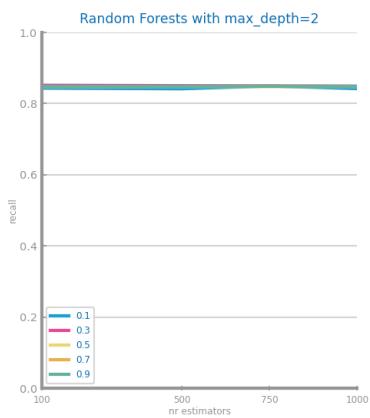


Figure 47: Best tree for dataset 2

Random Forests

Shall be used to present the results achieved through different parameterisations for the train of random forests. The results shall be compared and explanations for them shall be presented. Shall be used to address the overfitting phenomenon, studying the conditions under which models face it. Shall be used to present the evaluation of the best model achieved. May be used to present the most important variables in the model. **Shall not exceed 500 characters.**





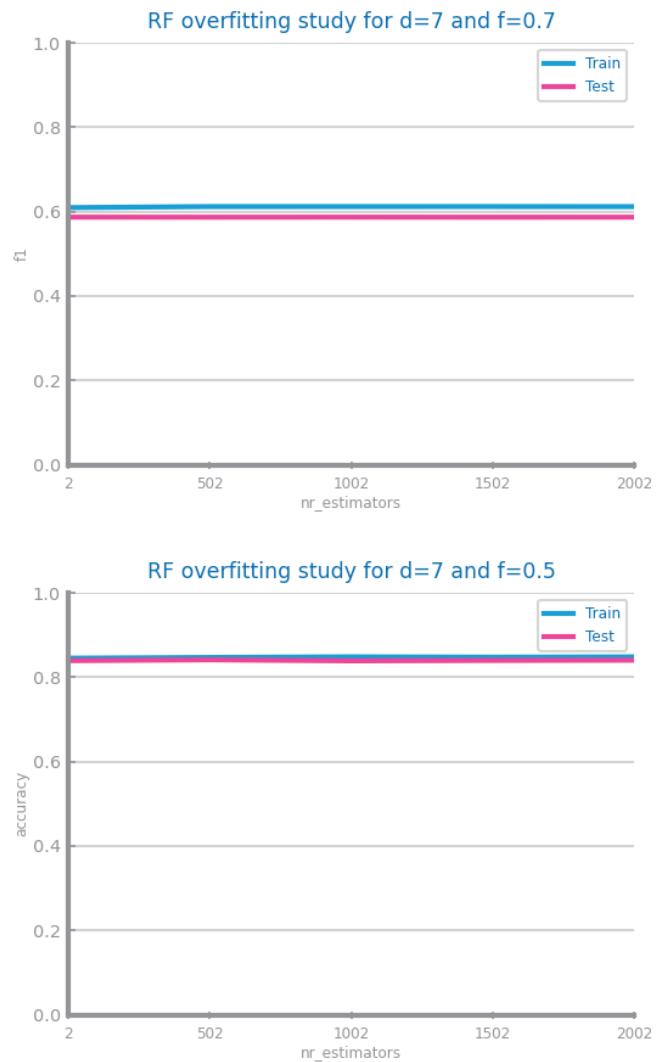
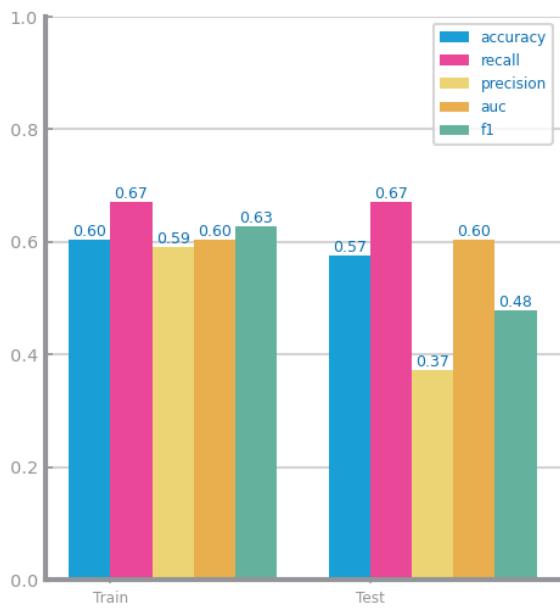
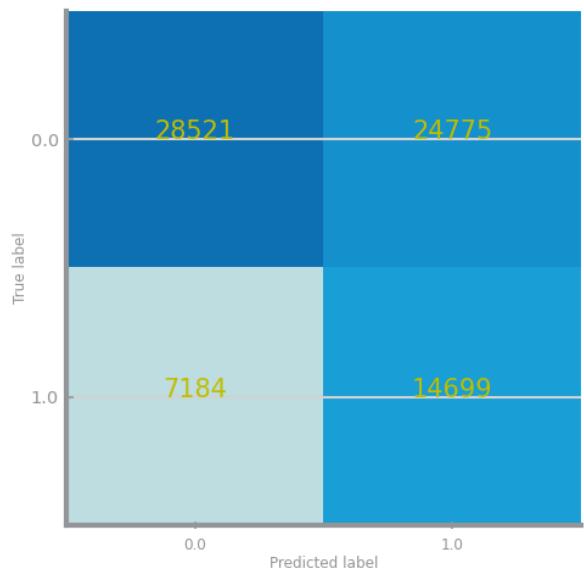


Figure 50: Random Forests overfitting analysis for dataset 1 (left) and dataset 2 (right)

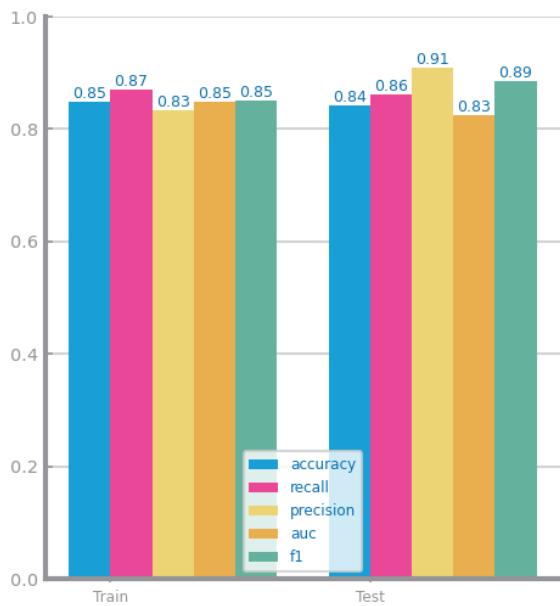
Best recall for RF (5, 0.9, 750)



Confusion matrix



Best accuracy for RF (7, 0.5, 1000)



Confusion matrix

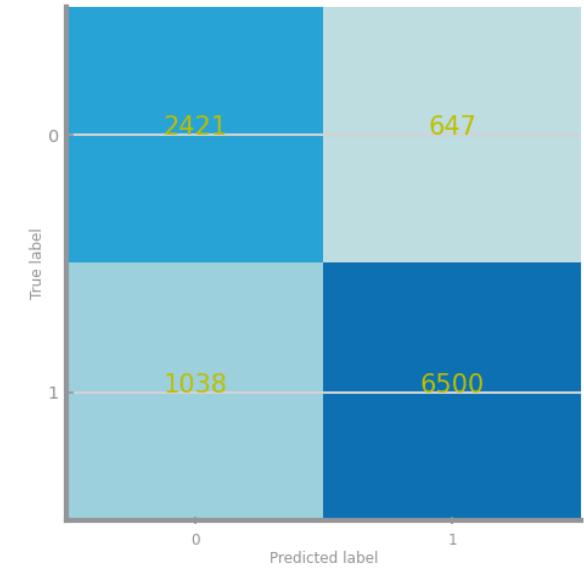


Figure 51: Random Forests best model results for dataset 1 (left) and dataset 2 (right)

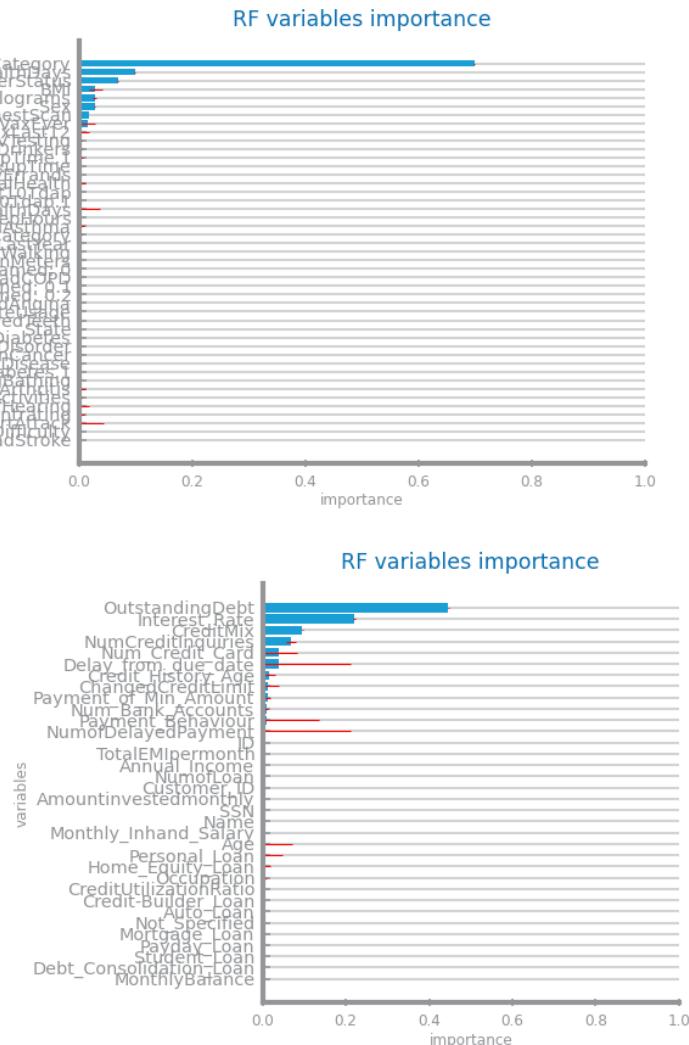
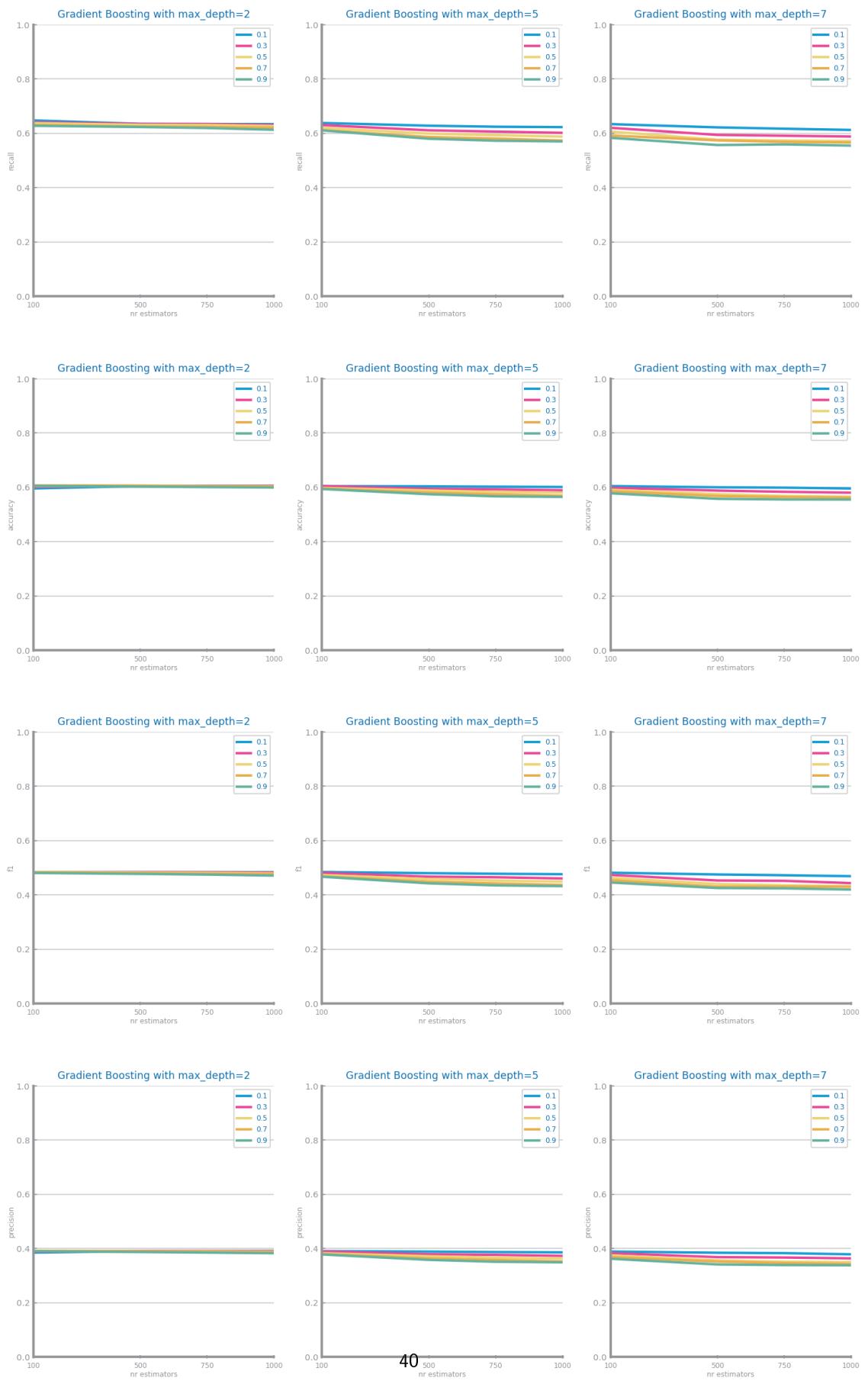


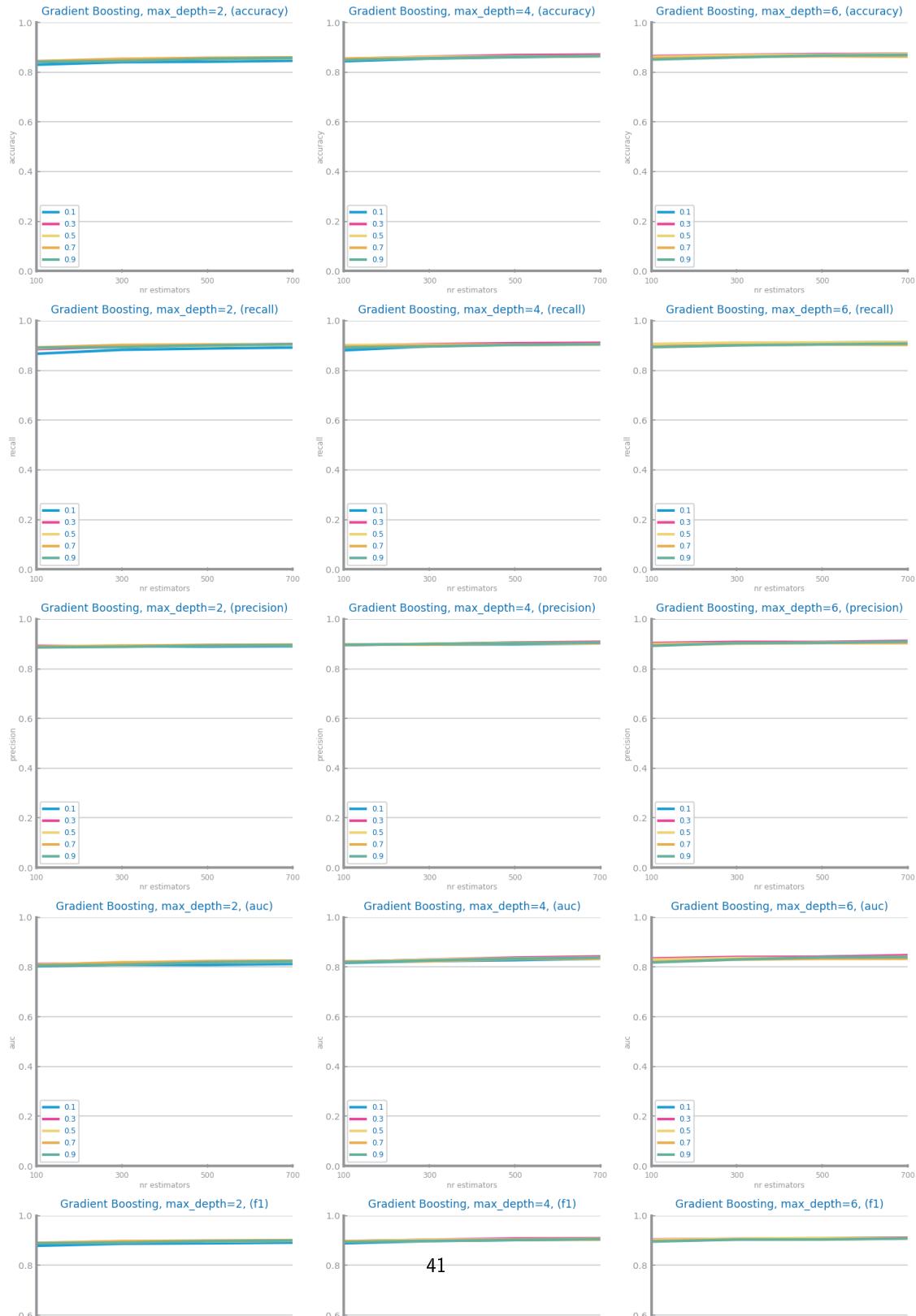
Figure 52: Random Forests variables importance for dataset 1 (left) and dataset 2 (right)

Gradient Boosting

Shall be used to present the results achieved through different parameterisations for the train of gradient boosting. The results shall be compared and explanations for them shall be presented. Shall be used to address the overfitting phenomenon, studying the conditions under which models face it. Shall be used to present the evaluation of the best model achieved. May be used to present the most important variables in the model. **Shall not exceed 500 characters.**



Gradient Boosting study for different parameters



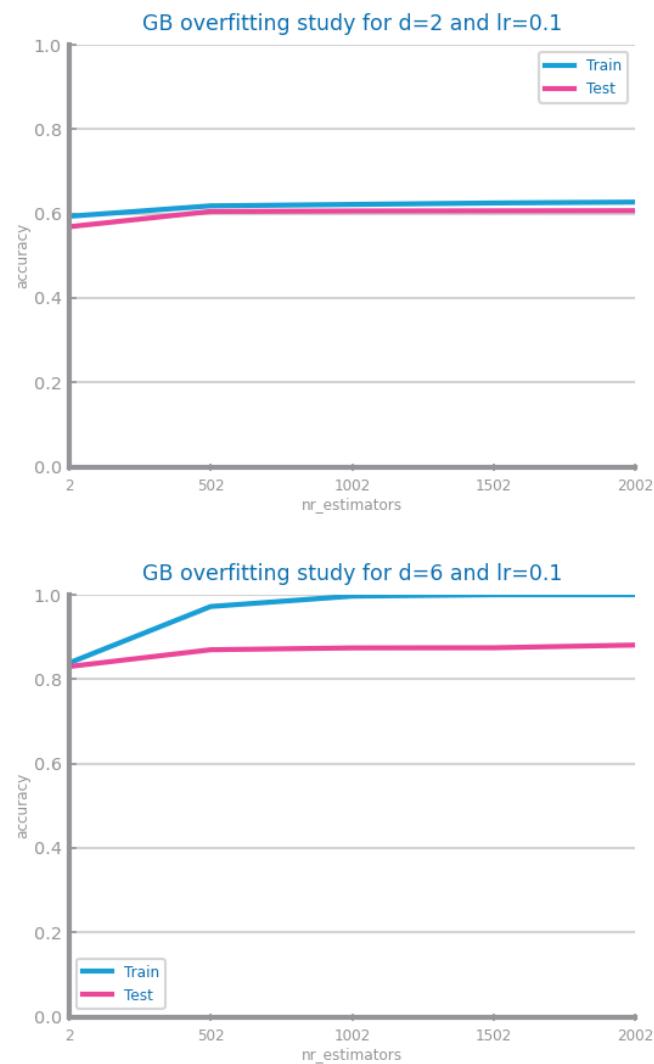
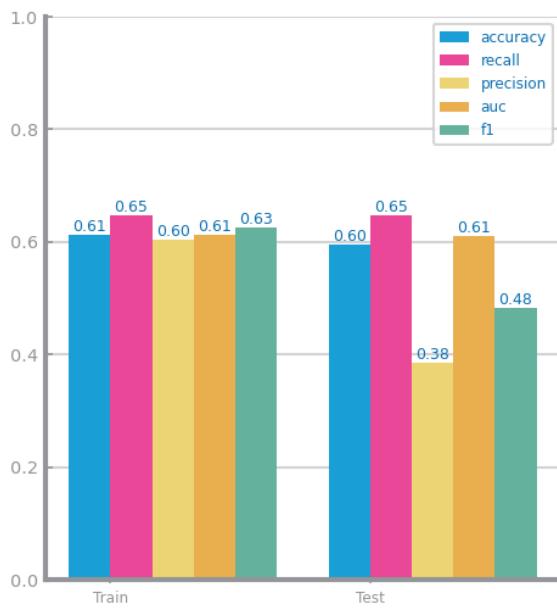
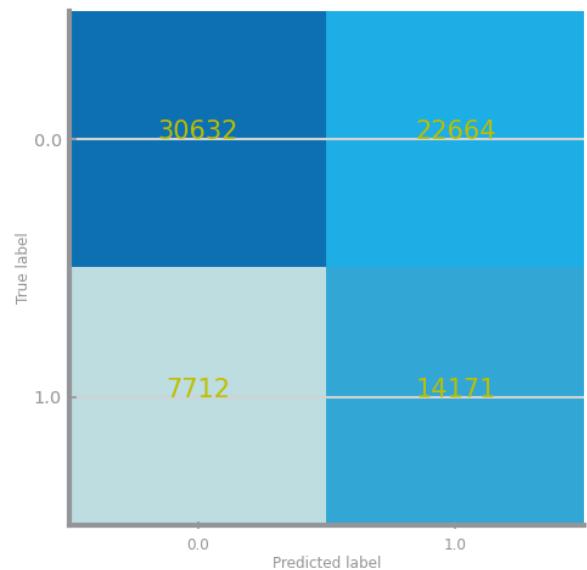


Figure 55: Gradient boosting overfitting analysis for dataset 1 (left) and dataset 2 (right)

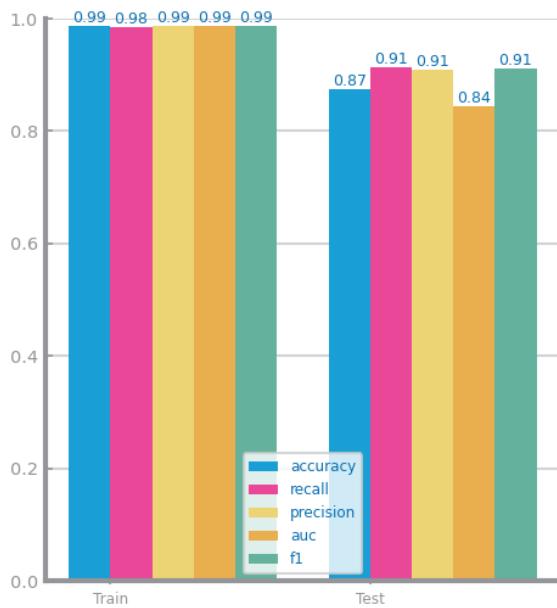
Best recall for GB (2, 0.1, 100)



Confusion matrix



Best accuracy for GB (6, 0.1, 700)



Confusion matrix

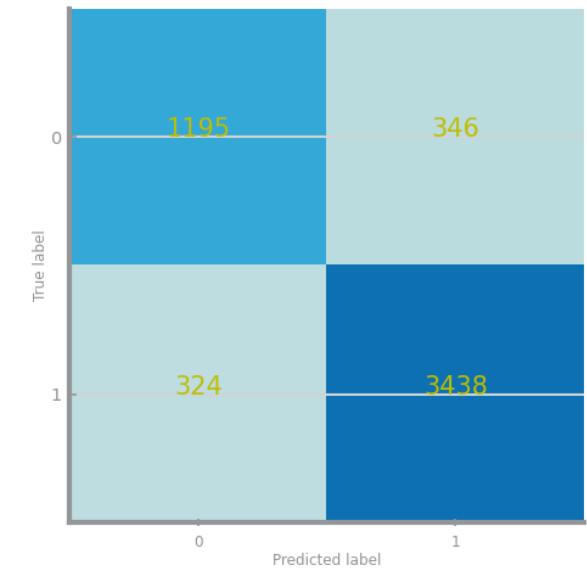


Figure 56: Gradient boosting best model results for dataset 1 (left) and dataset 2 (right)

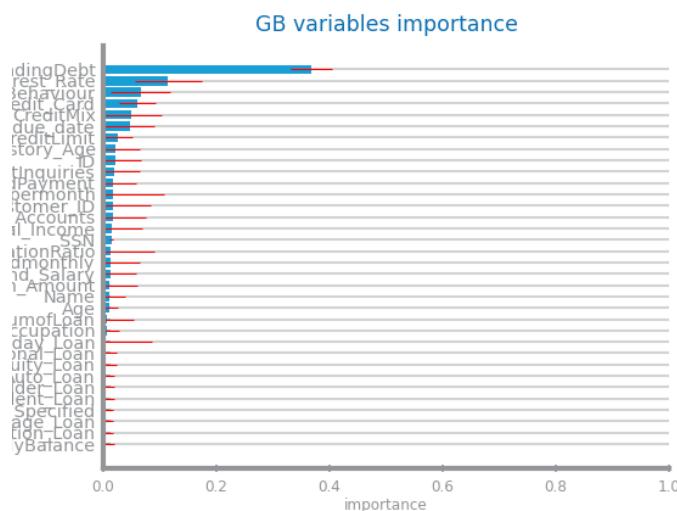
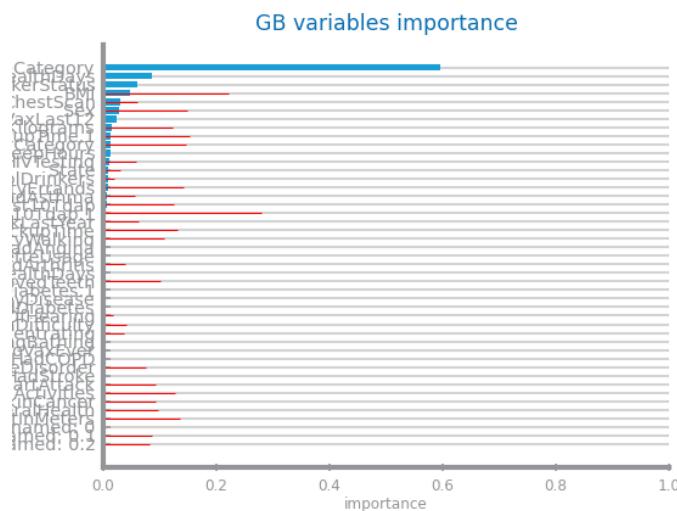
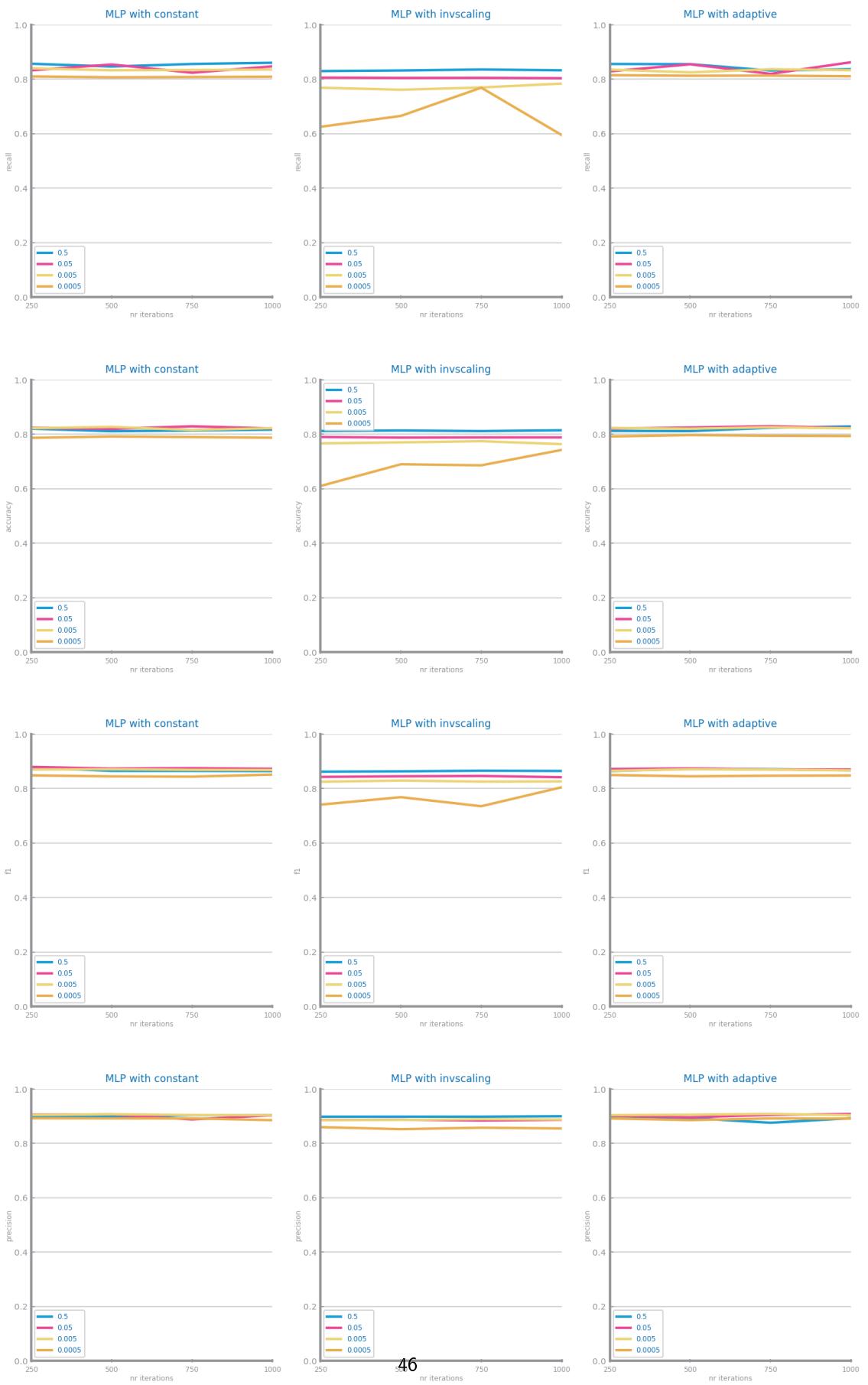


Figure 57: Gradient boosting variables importance for dataset 1 (left) and dataset 2 (right)

Multi-Layer Perceptrons

Shall be used to present the results achieved through different parameterisations for the train of MLPs. The results shall be compared and explanations for them shall be presented. Shall be used to address the overfitting phenomenon, studying the conditions under which models face it. Shall be used to present the evaluation of the best model achieved. **Shall not exceed 500 characters.**





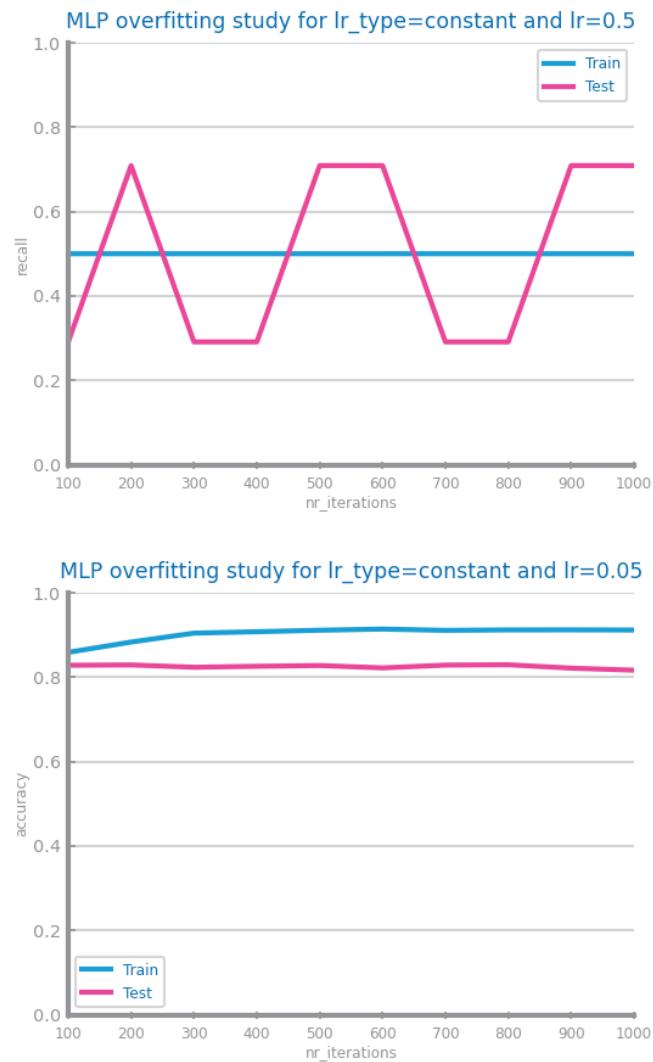


Figure 60: MLP overfitting analysis for dataset 1 (left) and dataset 2 (right)

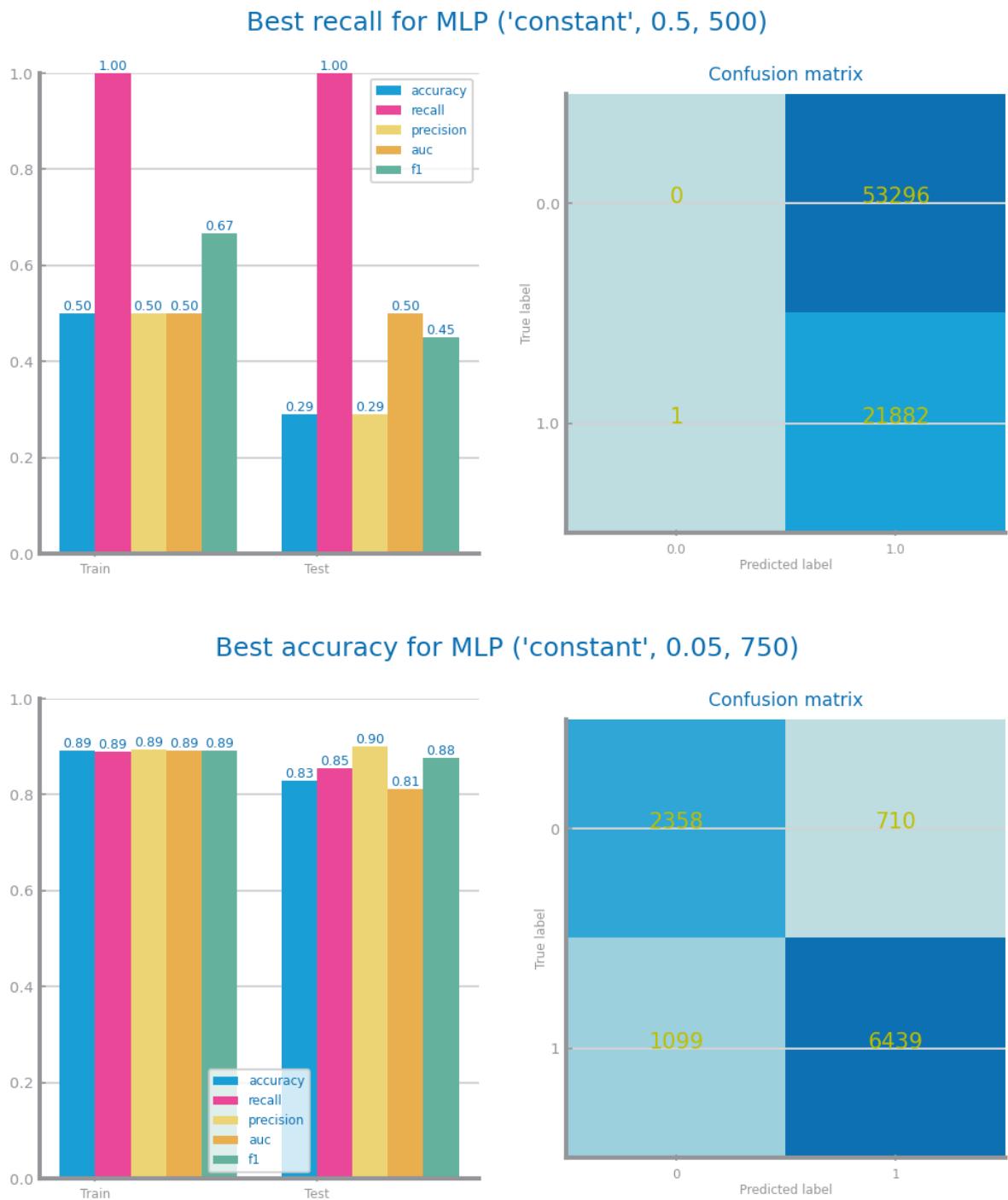


Figure 61: MLP best model results for dataset 1 (left) and dataset 2 (right)

4 CRITICAL ANALYSIS

Shall be used to present a summary of the results achieved with the different modelling techniques, and the impact of

the different preparation tasks on their performance. A cross-analysis of the different models may also be presented, identifying the most relevant variables common to all of them (when possible) and the relation among the patterns identified within the different classifiers. A critical assessment of the best models shall be presented, clearly stating if the models seem to be good enough for the problem at hand. **Additional charts may be presented here. Shall not exceed 2000 characters.**

TIME SERIES ANALYSIS

5 DATA PROFILING

Data Dimensionality and Granularity

May be used to identify the most atomic granularity and two other different granularities to consider. For the first dataset we used the "sum" function as "agg_fun" and studied the granularity at three different levels, weekly, monthly and quarterly. As you can also see, there is no cyclical behavior. **Shall not exceed 500 characters.**

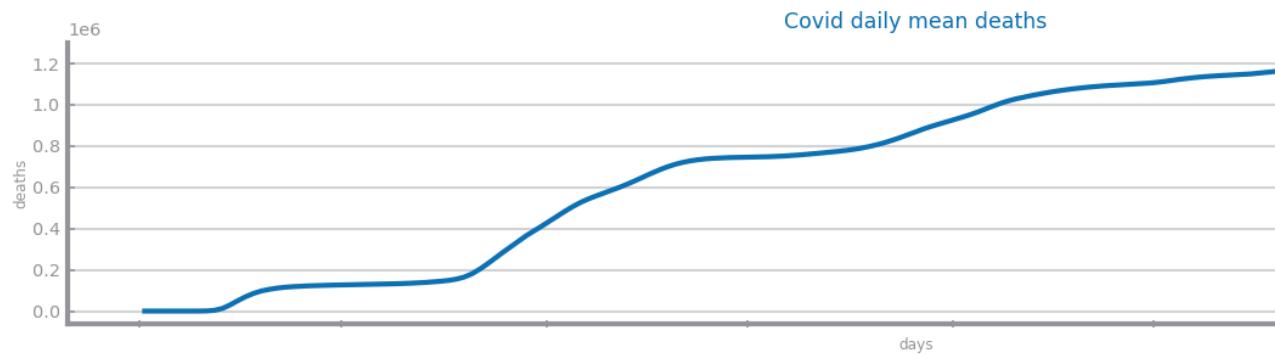


Figure 62: Time series 1 at the most granular detail

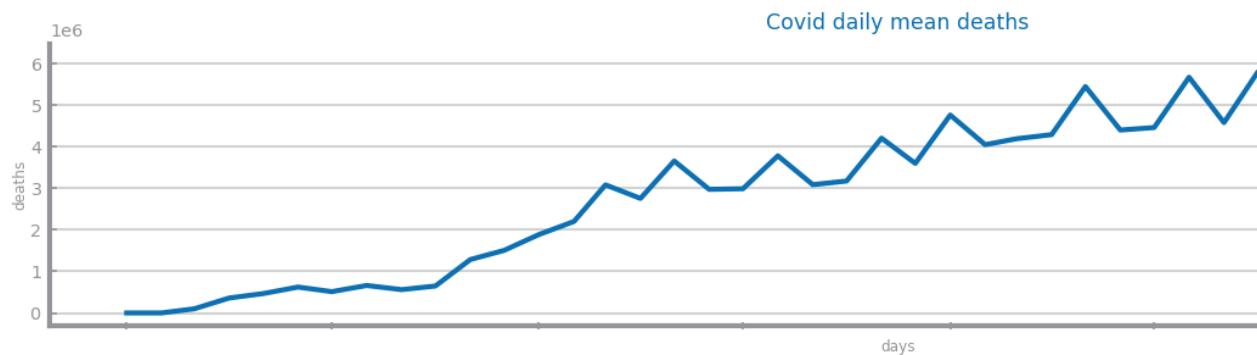


Figure 63: Time series 1 at the second chosen granularity

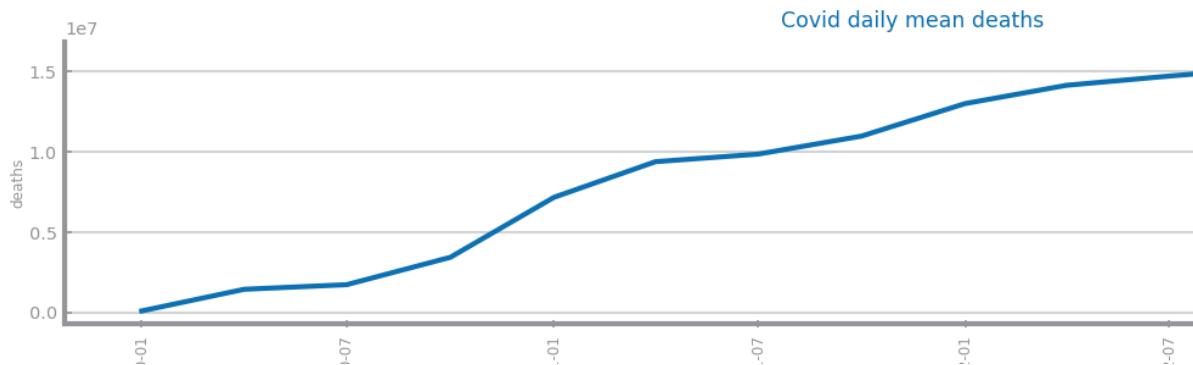


Figure 64: Time series 1 at the third chosen granularity

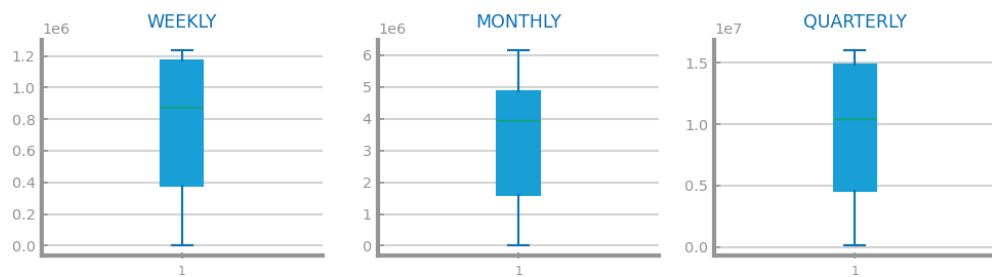
Figure 65: Time series 2 at the most granular detail

Figure 66: Time series 2 at the second chosen granularity

Figure 67: Time series 2 at the third chosen granularity

Data Distribution

Shall be used to perform the data analysis at those three different granularities, concerning the series distribution. **Shall not exceed 500 characters.**



```
count    1.990000e+02
mean     7.743876e+05
std      4.356312e+05
min      0.000000e+00
25%     3.775975e+05
50%     8.691990e+05
75%     1.172049e+06
max     1.238650e+06
Name: deaths, dtype: float64
```

```
count    1.990000e+02
mean     7.743876e+05
std      4.356312e+05
min      0.000000e+00
25%     3.775975e+05
50%     8.691990e+05
75%     1.172049e+06
max     1.238650e+06
Name: deaths, dtype: float64
```

```
count    4.600000e+01
mean     3.350068e+06
std      1.937462e+06
min      1.000000e+00
25%     1.602803e+06
50%     3.919796e+06
75%     4.874868e+06
max     6.172614e+06
Name: deaths, dtype: float64
```

Figure 68: Boxplot(s) for time series 1

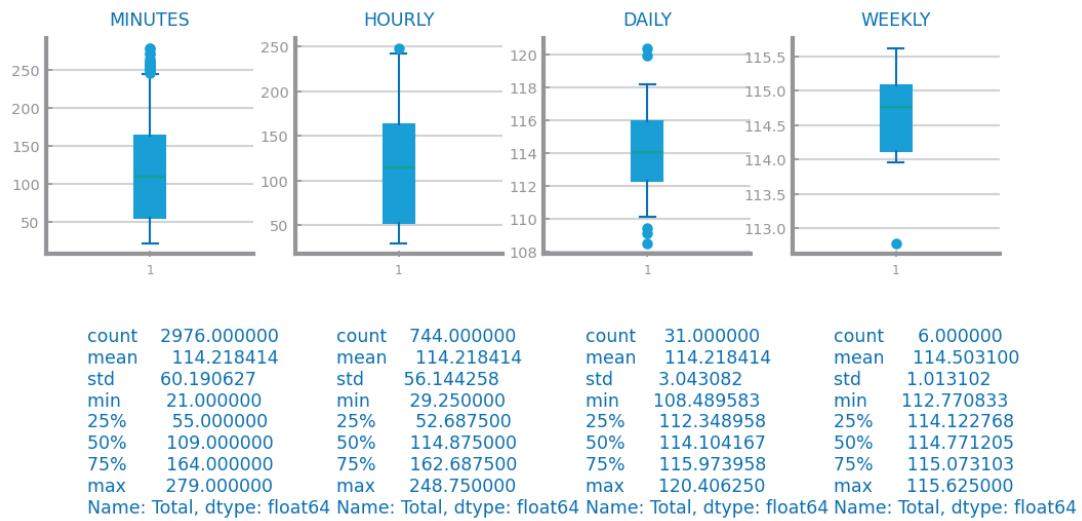


Figure 69: Boxplot(s) for time series 2

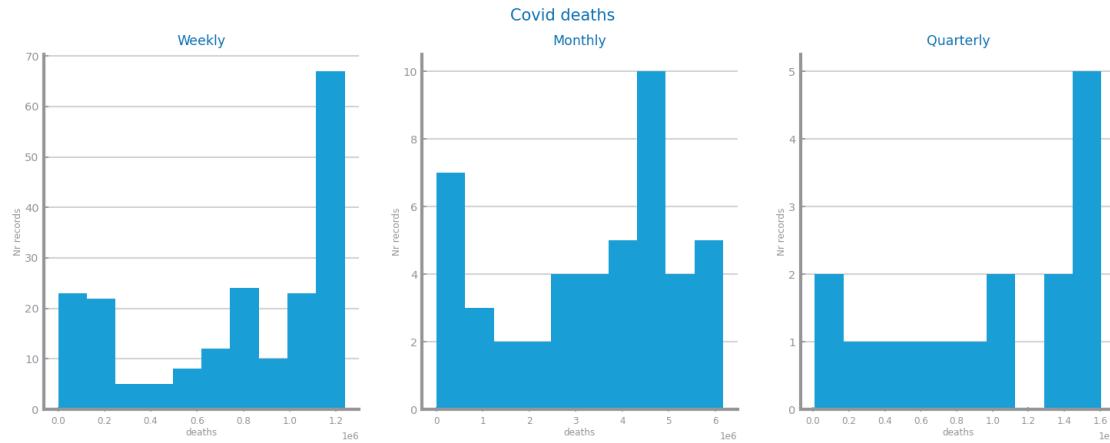


Figure 70: Histogram(s) for time series 1



Figure 71: Histogram(s) for time series 2

Data Stationarity

Shall be used to perform the data analysis at those three different granularities, concerning the series stationarity. For the 1st dataset we obtained a p-value of 0.223. Looking at the graphs, we can see that there is a trend in the first 2 and that there is evidence of a seasonal trend in the 3rd. For the 2nd dataset we obtained a p-value of 0. and that there is no trend but there is some seasonality. **Shall not exceed 300 characters.**

Covid weekly deaths

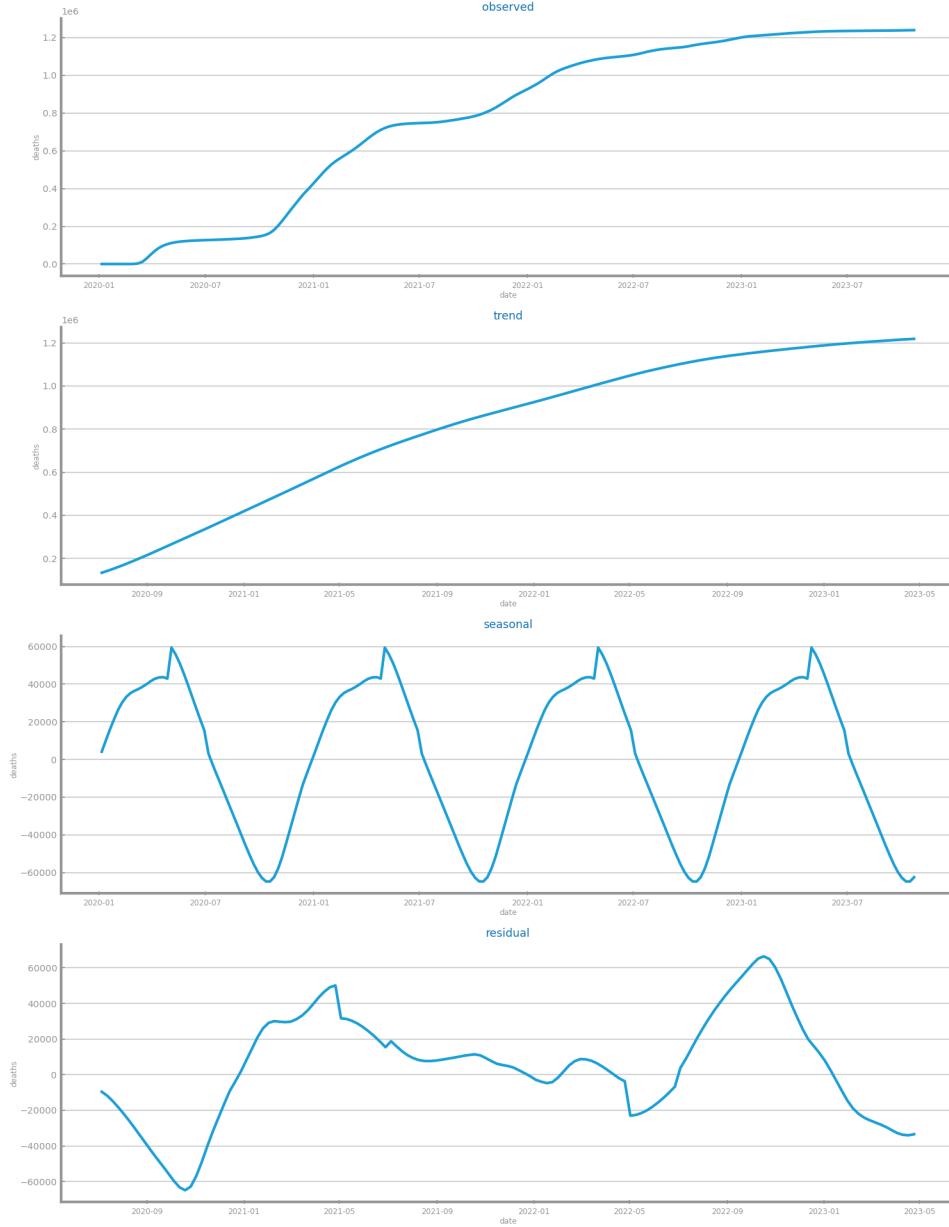


Figure 72: Components study for time series 1

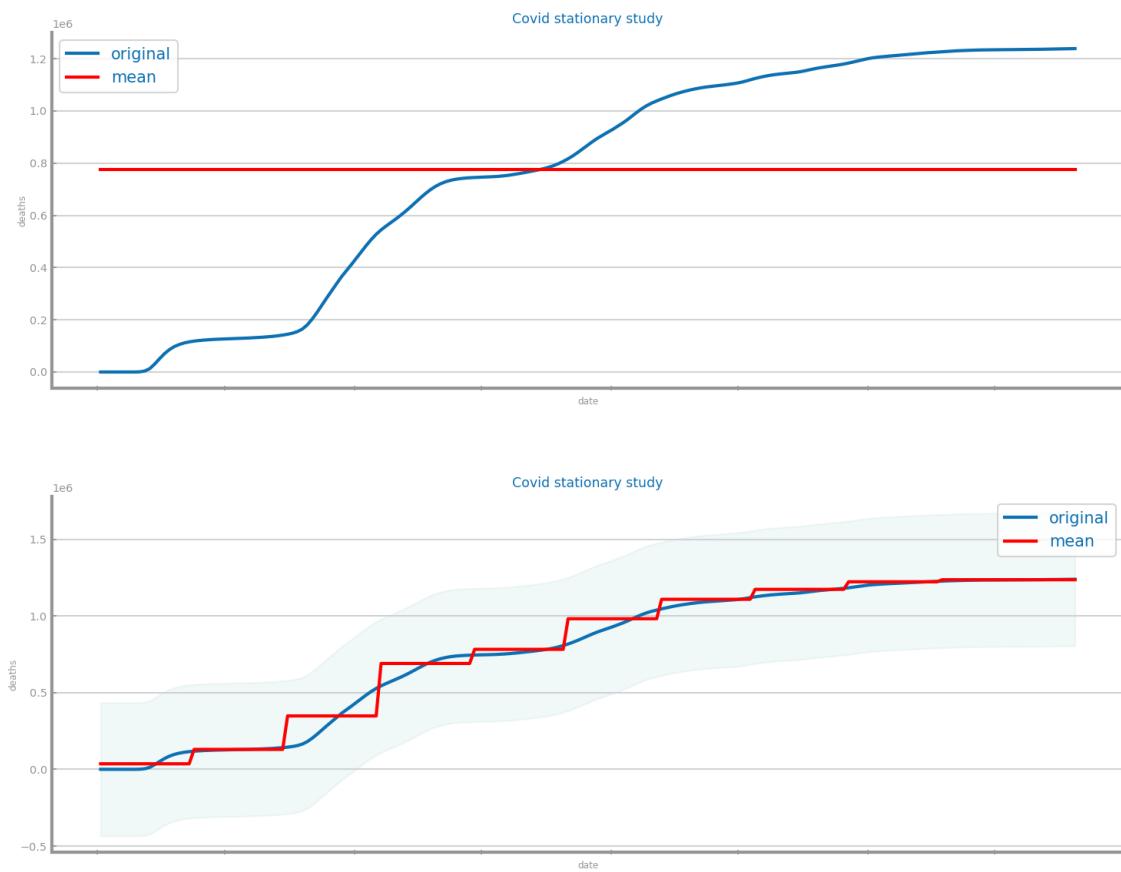


Figure 73: Stationarity study for time series 1

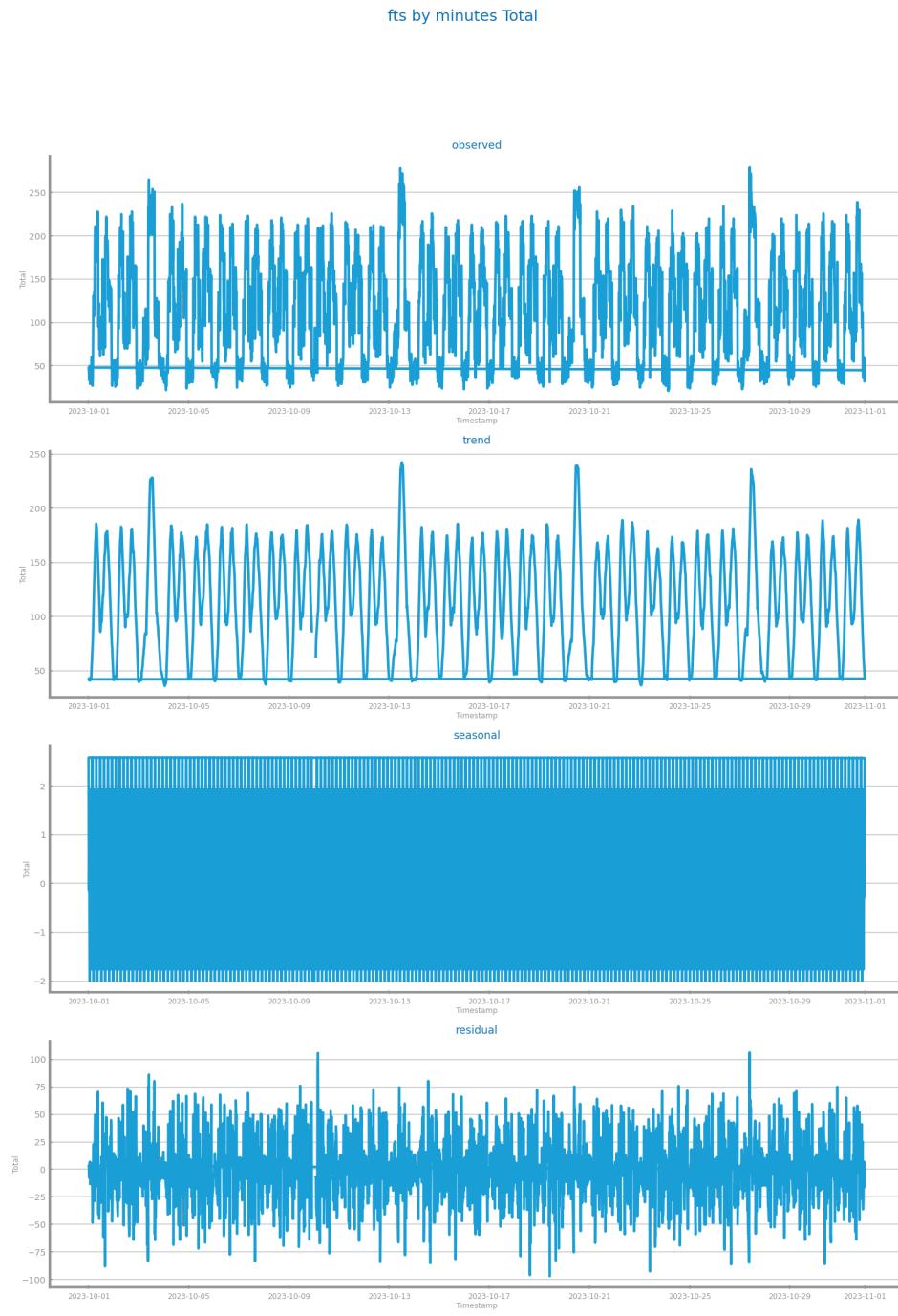


Figure 74: Components study for time series 2

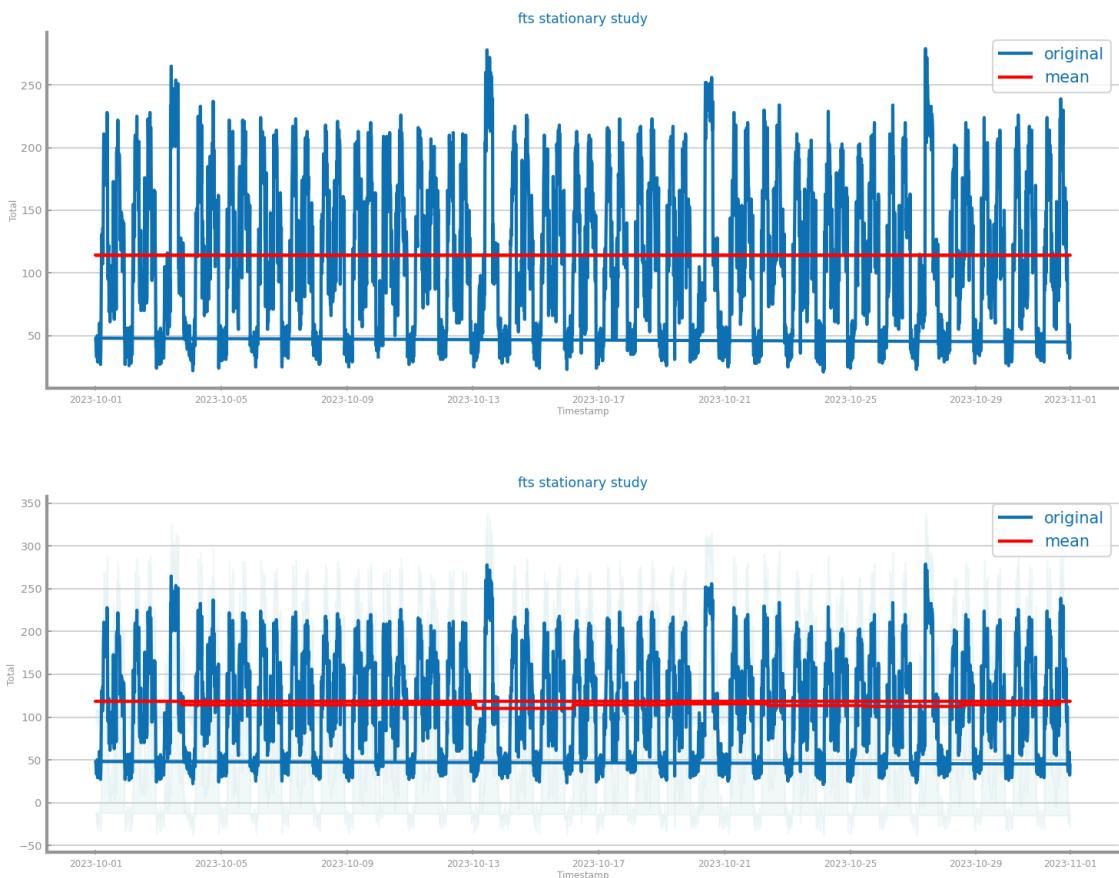


Figure 75: Stationarity study for time series 2

6 DATA TRANSFORMATION

Aggregation

Shall describe the results of applying three different aggregations over both datasets, and identifying the granularity chosen to proceed. To study the best aggregation, we applied the model to the 3 different levels studied in Data Profiling. In the first dataset we chose the weekly aggregation and in the second the minute aggregation because that was where we obtained the lowest values for the different errors. **Shall not exceed 300 characters.**

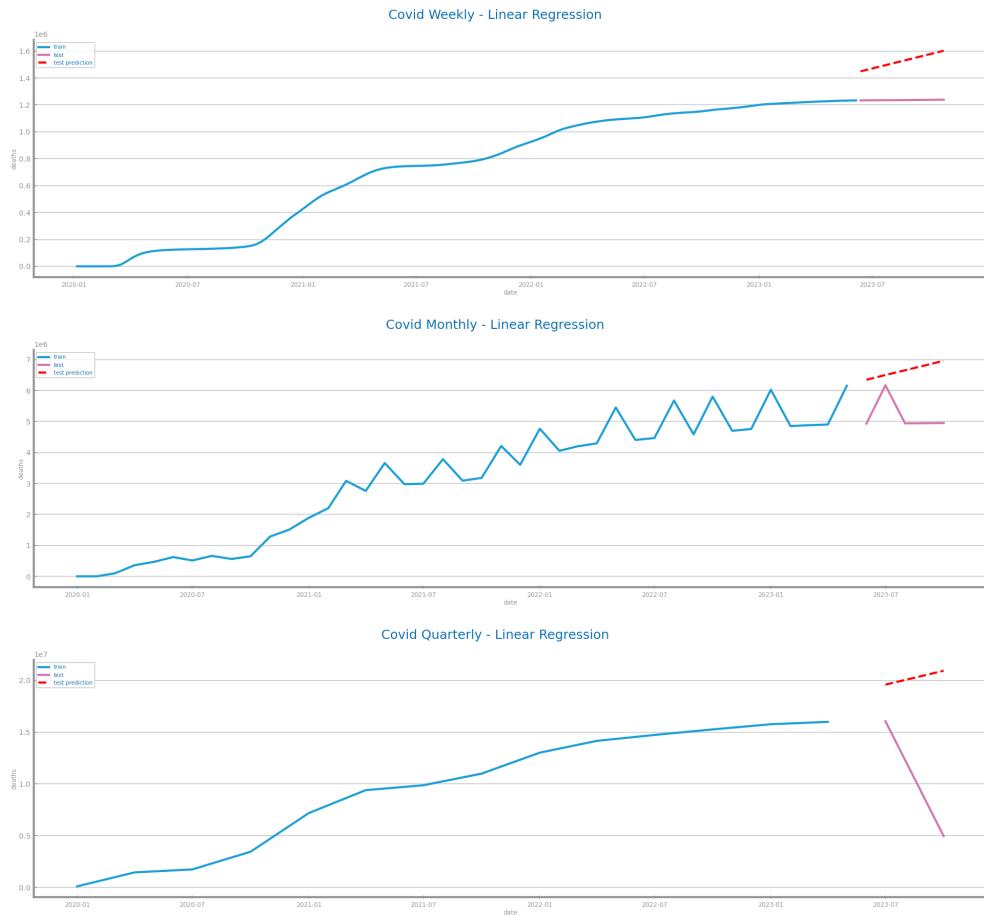


Figure 76: Forecasting plots after different aggregations on time series 1



Figure 77: Forecasting results after different aggregations on time series 1

fts - Linear R

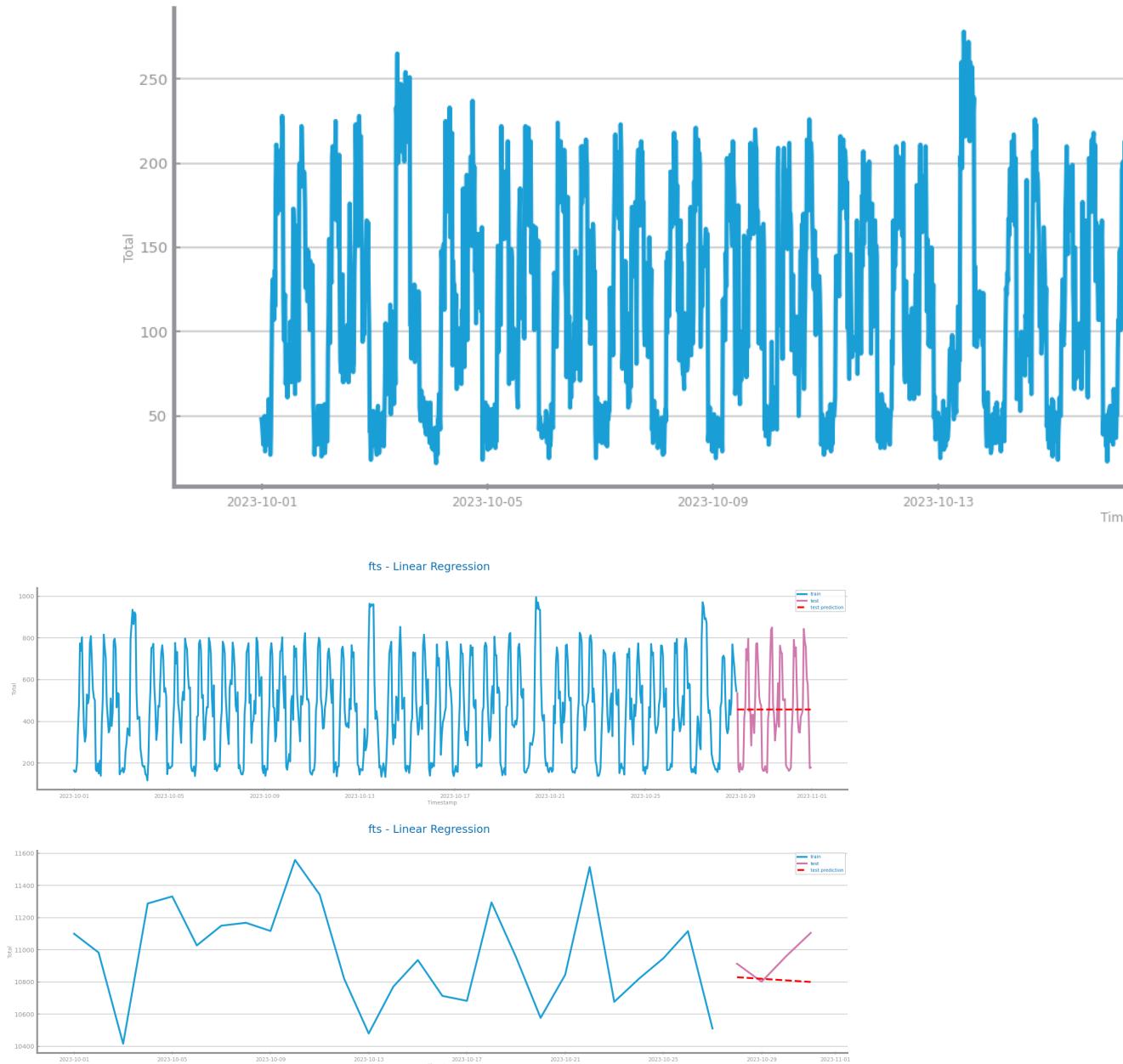


Figure 78: Forecasting plots after different aggregations on time series 2

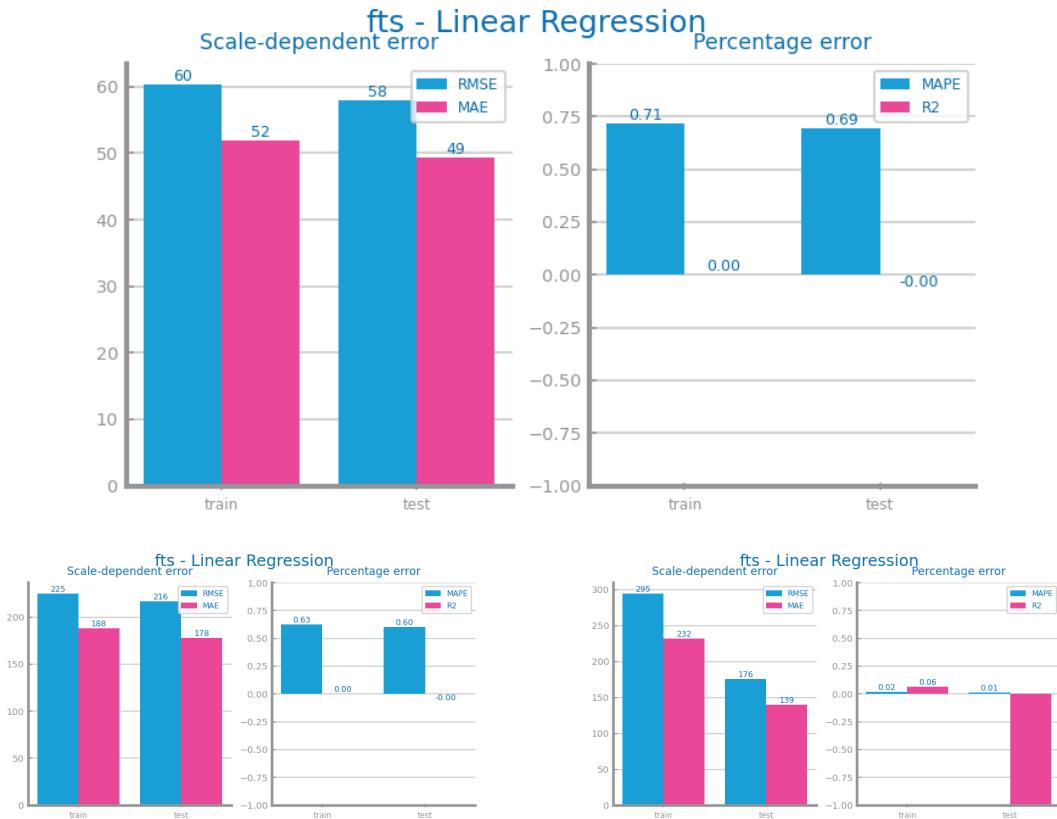


Figure 79: Forecasting results after different aggregations on time series 2

Smoothing

Shall describe the results of applying smoothing transformations over both datasets, and identifying the best result to proceed. To study the best Window Size, we applied the model to 4 different values (25, 50, 75 and 100). In this case we chose 100 for both datasets since it was where we obtained the lowest values for the different errors. **Shall not exceed 300 characters.**

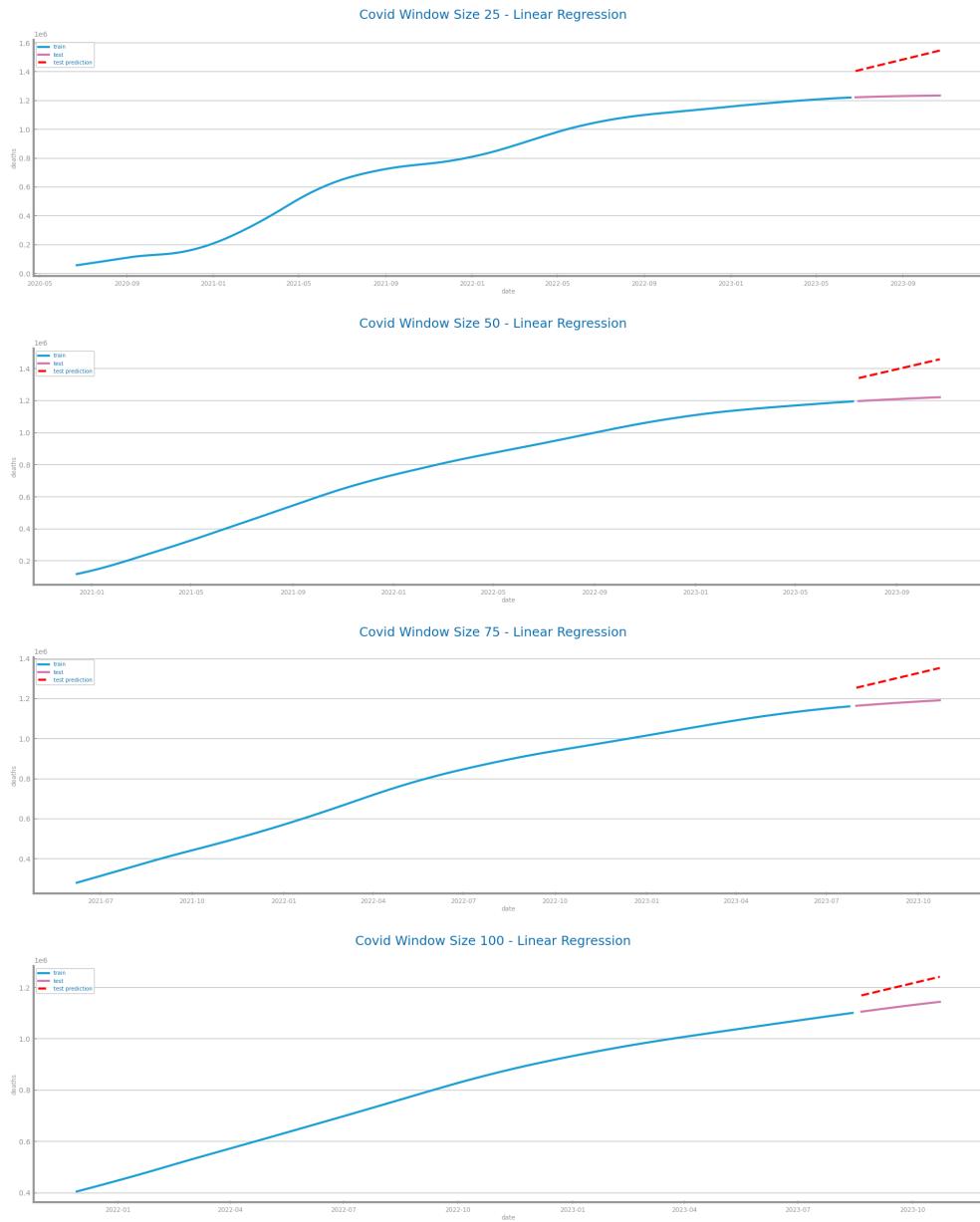
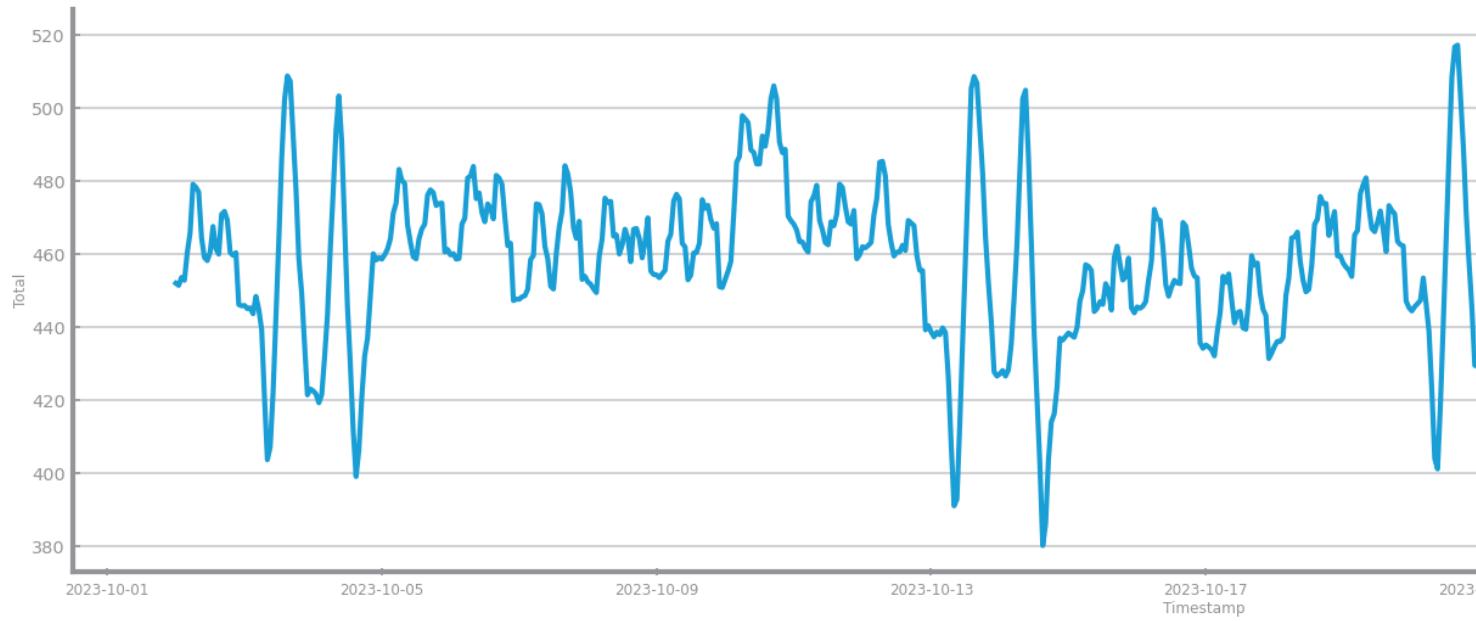


Figure 80: Forecasting plots after different smoothing parameterisations on time series 1

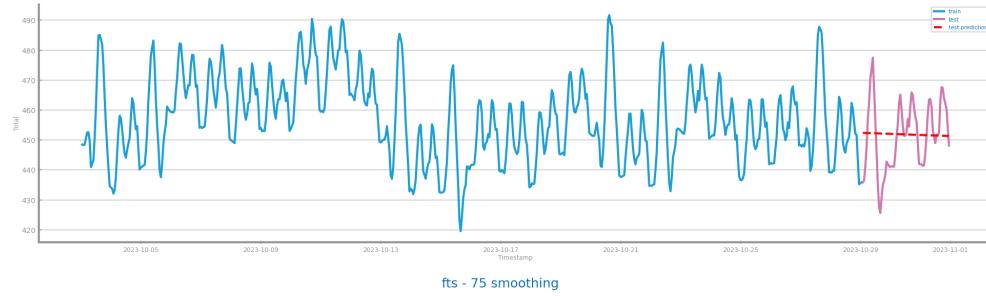


Figure 81: Forecasting results after different smoothing parameterisations on time series 1

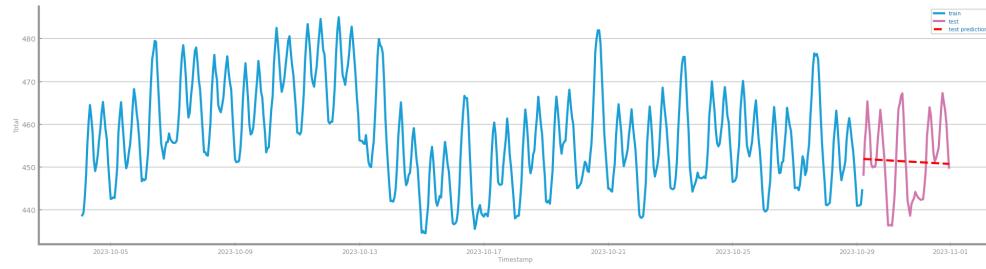
fts - 25 smoothing



fts - 50 smoothing



fts - 75 smoothing



fts - 100 smoothing

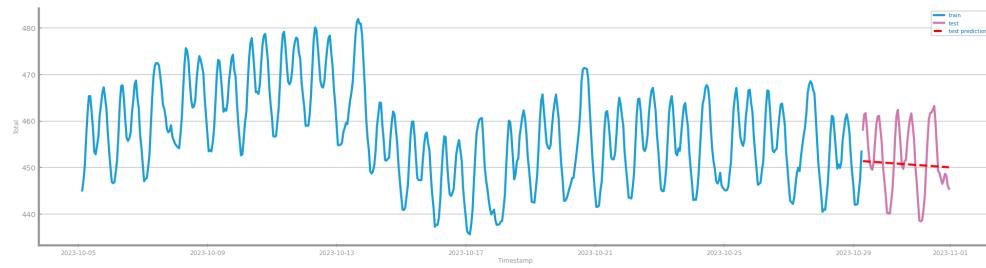


Figure 82: Forecasting plots after different smoothing parameterisations on time series 2

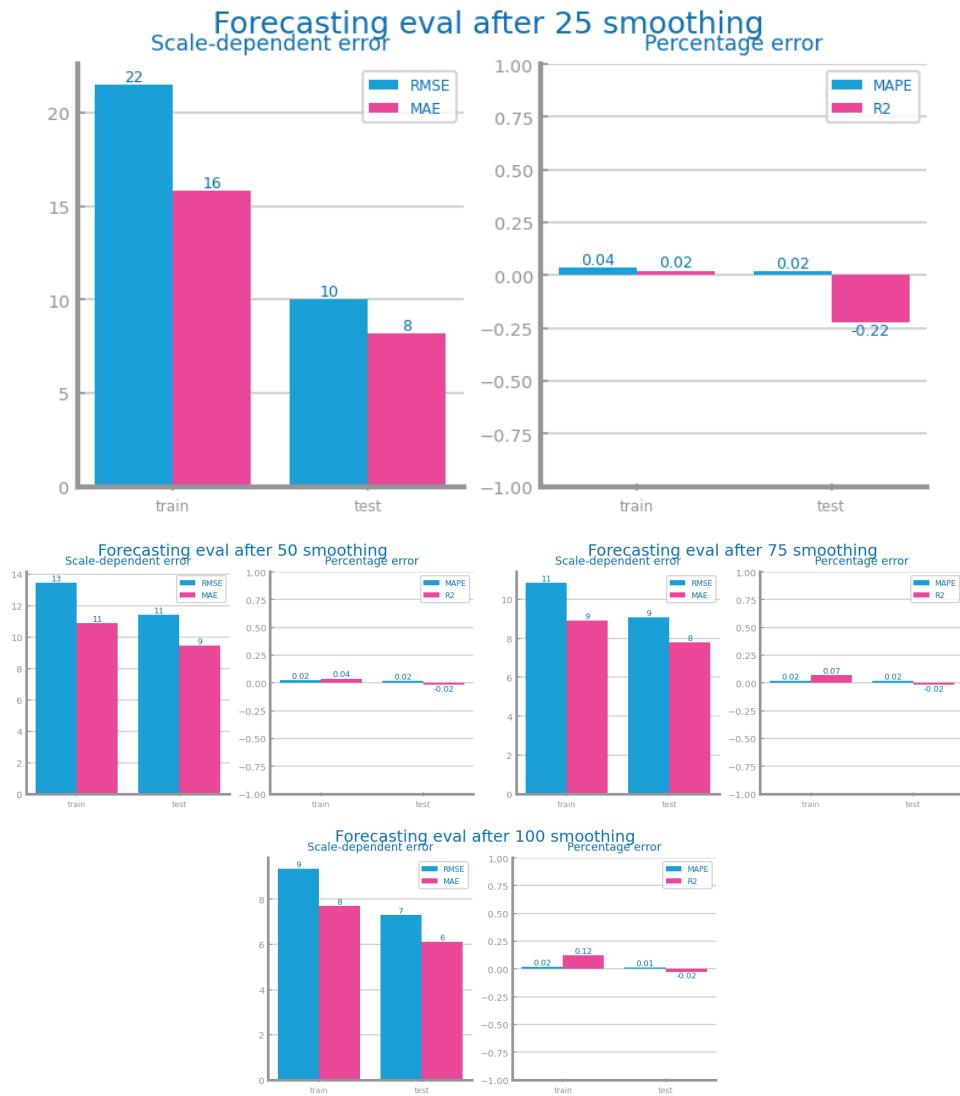


Figure 83: Forecasting results after different smoothing parameterisations on time series 2

Differentiation

Shall describe the results of applying two consecutive differentiation of both datasets, and identifying the best result to proceed. To study the best derivative, we applied the model to the first two derivatives. Once again, we proceed with the second derivative for both datasets since it was where we obtained the lowest values for the different errors. **Shall not exceed 300 characters.**

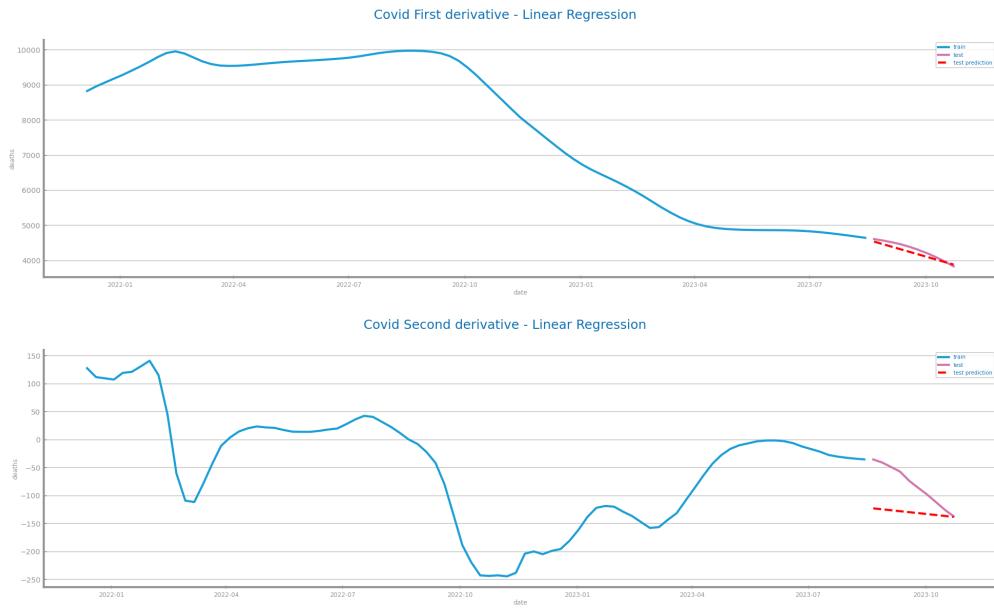


Figure 84: Forecasting plots after first and second differentiation of time series 1

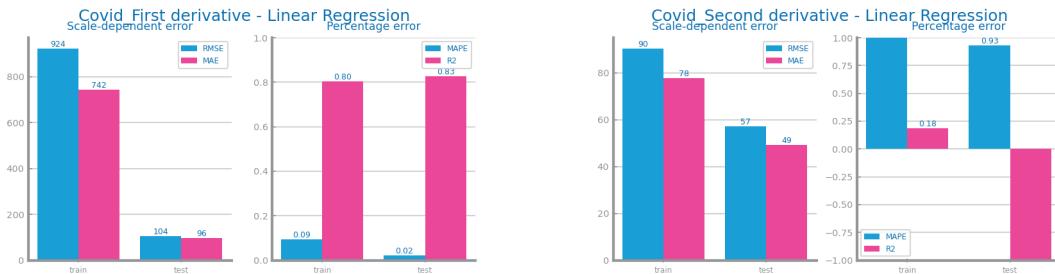


Figure 85: Forecasting results after first and second differentiation of time series 1

Figure 86: Forecasting plots after first and second differentiation of time series 2

Figure 87: Forecasting results after first and second differentiation of time series 2

Other transformations (optional)

Shall describe the results of applying other transformations over both datasets, and identifying the best result to proceed. Finally, we applied scaling in both datasets in order to have best values to use in the models' evaluation specifically in the LSTM model. **Shall not exceed 500 characters.**

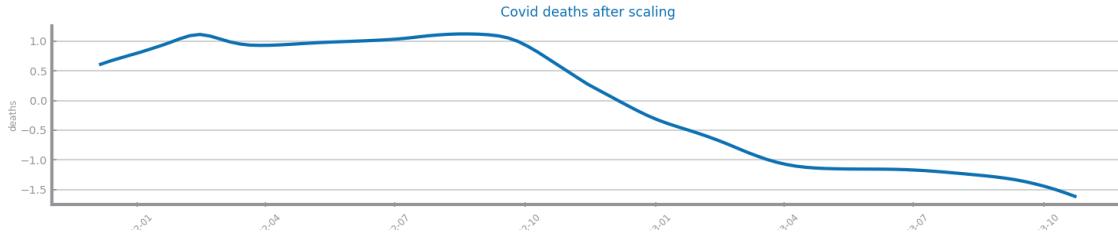


Figure 88: Forecasting plots after applying scaling over time series 1

Figure 89: Forecasting results after applying other transformations over time series 1

Figure 90: Forecasting plots after applying other transformations over time series 2

Figure 91: Forecasting results after applying other transformations over time series 2

7 MODELS' EVALUATION

Shall be used to summarise the transformations done over the original time series. **Shall not exceed 500 characters.**

Simple Average Model

Shall be used to present the results achieved through the simple average model. **Shall not exceed 200 characters.**

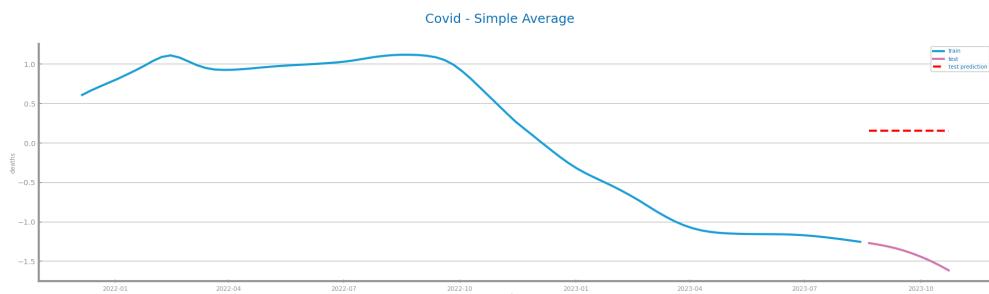


Figure 92: Forecasting plots obtained with Simple Average model over time series 1

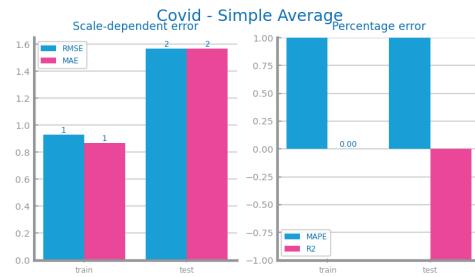


Figure 93: Forecasting results obtained with Simple Average model over time series 1

Figure 94: Forecasting plots obtained with Simple Average model over time series 2

Figure 95: Forecasting results obtained with Simple Average model over time series 2

Persistence Model

Shall be used to present the results achieved through the persistence model. **Shall not exceed 500 characters.**

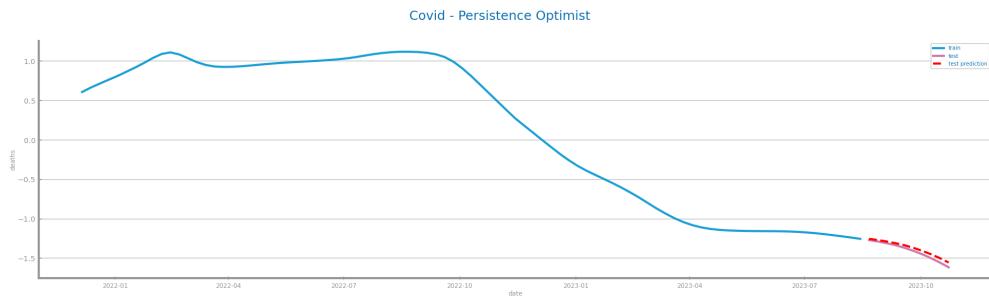


Figure 96: Forecasting plots obtained with Persistence model (long term) over time series 1



Figure 97: Forecasting plots obtained with Persistence model (one-set-behind) over time series 1

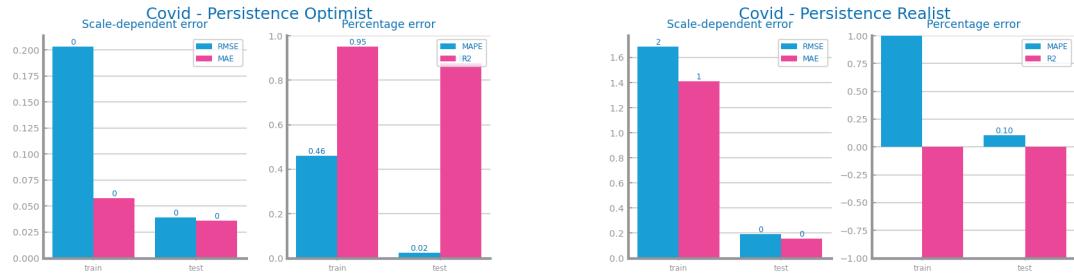


Figure 98: Forecasting results obtained with Persistence model in both situations over time series 1

Figure 99: Forecasting plots obtained with Persistence model (long term) over time series 2

Figure 100: Forecasting plots obtained with Persistence model (one-set-behind) over time series 2

Figure 101: Forecasting results obtained with Persistence model in both situations over time series 2

Rolling Mean Model

Shall be used to present the results achieved through the rolling mean forecasting algorithms. **Shall not exceed 500 characters.**

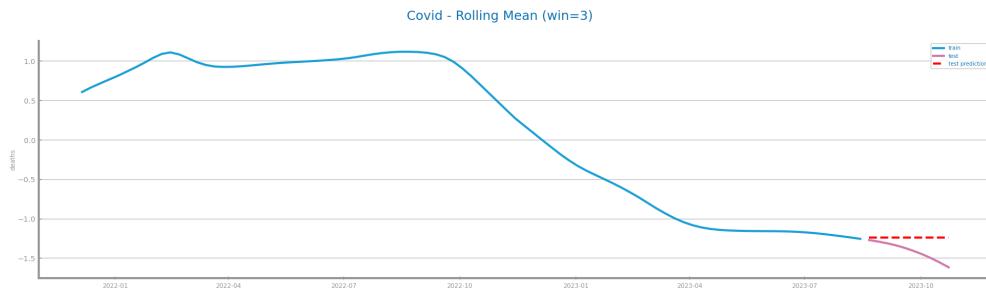


Figure 102: Forecasting study over different parameterisations of the rolling mean algorithm over time series 1

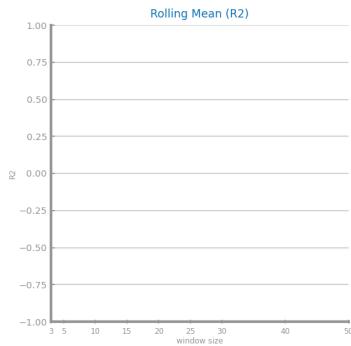


Figure 103: Forecasting plots obtained with the best parameterisation of rolling mean algorithm, over time series 1

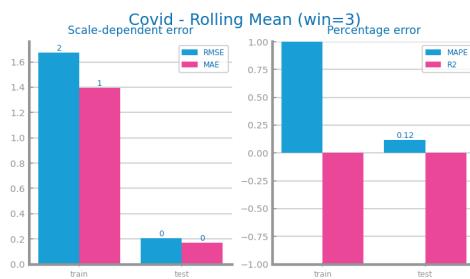


Figure 104: Forecasting results obtained with the best parameterisation of rolling mean algorithm, over time series 1

Figure 105: Forecasting study over different parameterisations of the rolling mean algorithm over time series 2

Figure 106: Forecasting plots obtained with the best parameterisation of rolling mean algorithm, over time series 2

ARIMA Model

Shall be used to present the results achieved through the ARIMA forecasting algorithms. **Shall not exceed 500 characters.**

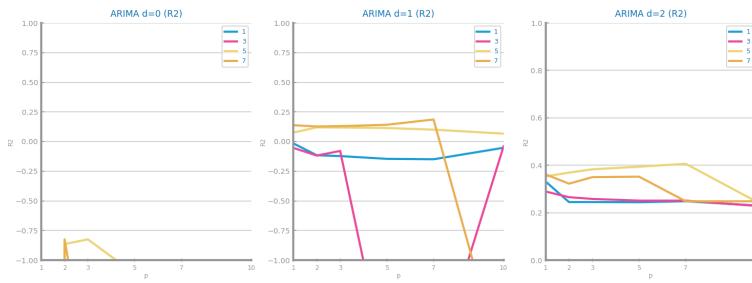


Figure 108: Forecasting study over different parameterisations of the ARIMA algorithm over time series 1

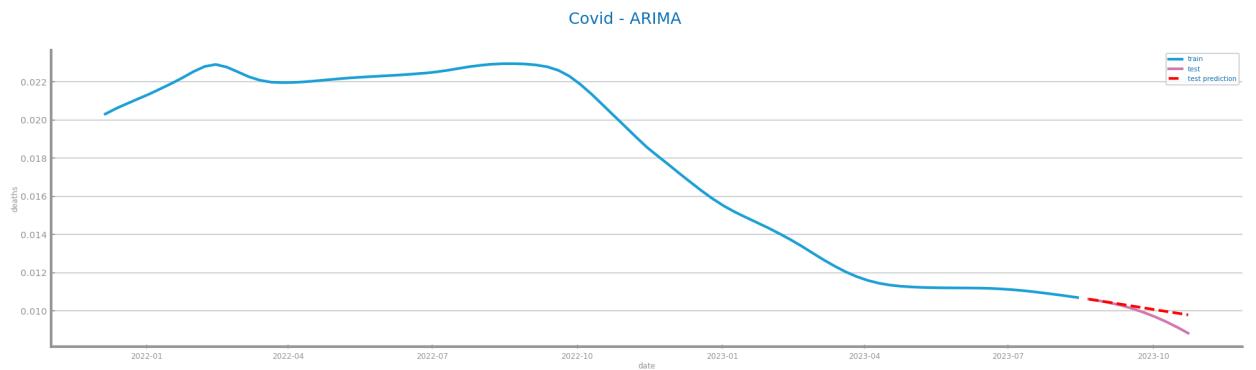


Figure 109: Forecasting plots obtained with the best parameterisation of ARIMA algorithm, over time series 1

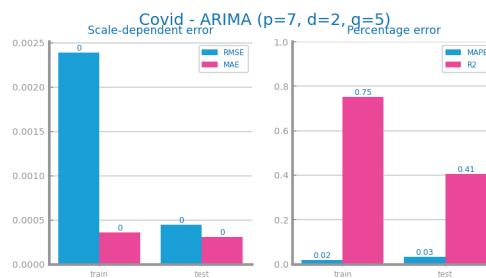


Figure 110: Forecasting results obtained with the best parameterisation of ARIMA algorithm, over time series 1

Figure 111: Forecasting study over different parameterisations of the ARIMA algorithm with external variables over time series 1

Figure 112: Forecasting plots obtained with the best parameterisation of ARIMA algorithm with external variables over time series 1

Figure 113: Forecasting results obtained with the best parameterisation of ARIMA algorithm with external variables over time series 1

Figure 114: Forecasting study over different parameterisations of the ARIMA algorithm over time series 2

Figure 115: Forecasting plots obtained with the best parameterisation of ARIMA algorithm, over time series 2

Figure 116: Forecasting results obtained with the best parameterisation of ARIMA algorithm, over time series 2

Figure 117: Forecasting study over different parameterisations of the ARIMA algorithm with external variables over time series 2

Figure 118: Forecasting plots obtained with the best parameterisation of ARIMA algorithm with external variables over time series 2

Figure 119: Forecasting results obtained with the best parameterisation of ARIMA algorithm with external variables over time series 2

LSTMs Model

Shall be used to present the results achieved through LSTMs. **Shall not exceed 500 characters.**

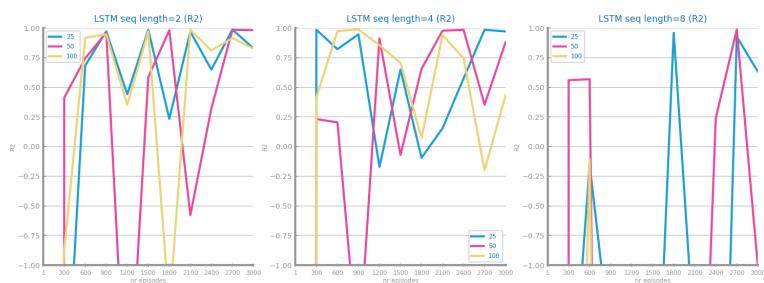


Figure 120: Forecasting study over different parameterisations of LSTMs over time series 1

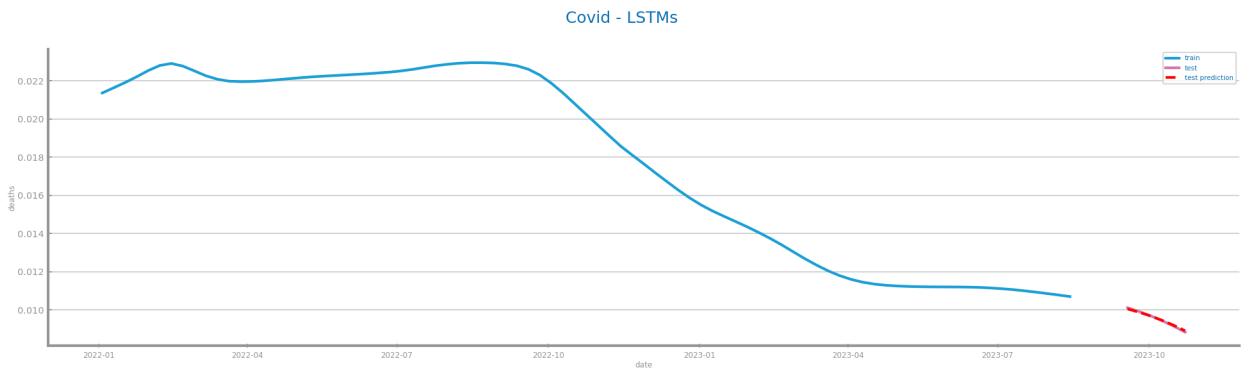


Figure 121: Forecasting plots obtained with the best parameterisation of LSTMs, over time series 1

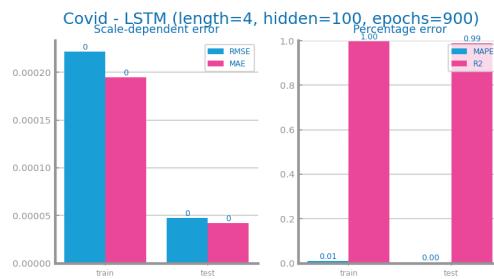


Figure 122: Forecasting results obtained with the best parameterisation of LSTMs, over time series 1

Figure 123: Forecasting study over different parameterisations of LSTMs with multiple variables over time series 1

Figure 124: Forecasting plots obtained with the best parameterisation of LSTMs with multiple variables over time series 1

Figure 125: Forecasting results obtained with the best parameterisation of LSTMs with multiple variables over time series 1

Figure 126: Forecasting study over different parameterisations of the LSTMs over time series 2

Figure 127: Forecasting plots obtained with the best parameterisation of LSTMs, over time series 2

Figure 128: Forecasting results obtained with the best parameterisation of LSTMs, over time series 2

Figure 129: Forecasting study over different parameterisations of LSTMs with multiple variables over time series 2

Figure 130: Forecasting plots obtained with the best parameterisation of LSTMs with multiple variables over time series 2

Figure 131: Forecasting results obtained with the best parameterisation of LSTMs with multiple variables over time series 2

8 CRITICAL ANALYSIS

Shall be used to present a summary of the results achieved with the different forecasting techniques, and the impact of the different preparation tasks on their performance. A critical assessment of the best models shall be presented, clearly stating if the models seem to be good enough for the problem at hand. Additional charts may be presented here. **Shall not exceed 2000 characters.**