Data Science Project

Team nr: 16	Student 1:	Antero Morgado	IST nr:	1119213
	Student 2:	David Ferreira	IST nr:	1107077
	Student 3:	José Fernandes	IST nr:	1103727
	Student 4:	Olha Buts	IST nr:	1116276

CLASSIFICATION

1 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.**
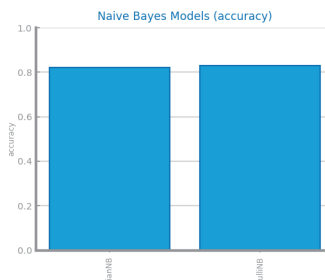
Naïve Bayes

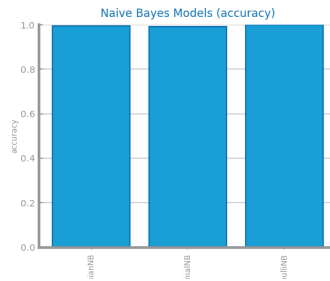Figure 1: Naïve Bayes alternatives comparison for dataset 1

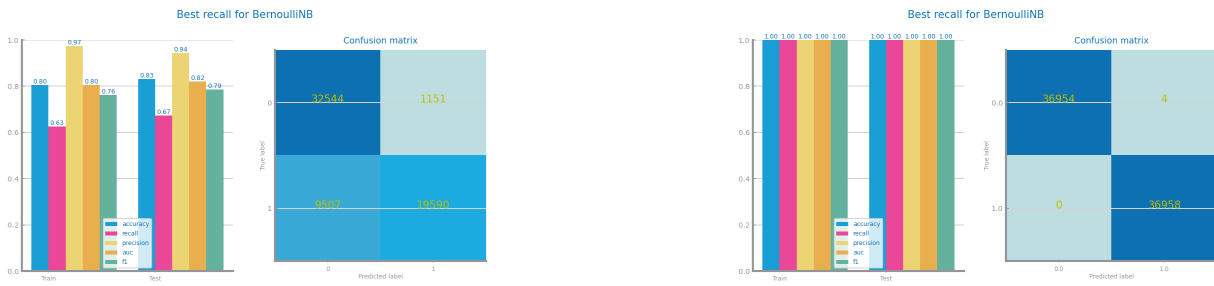Figure 2: Naïve Bayes alternative comparison for dataset 2

Figure 3: Naïve Bayes best model results for dataset 1 (left) and dataset 2 (right)
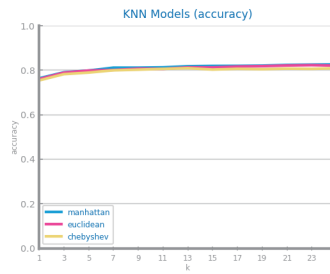
KNN

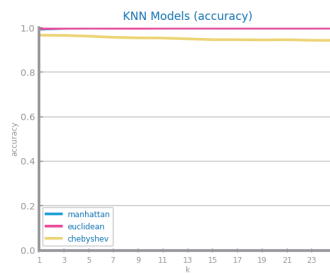Figure 4: KNN different parameterisations comparison for dataset 1

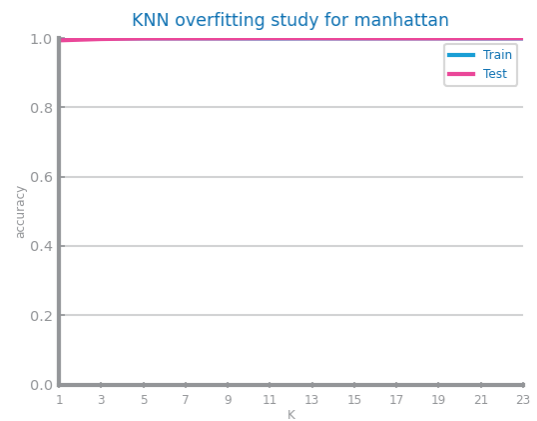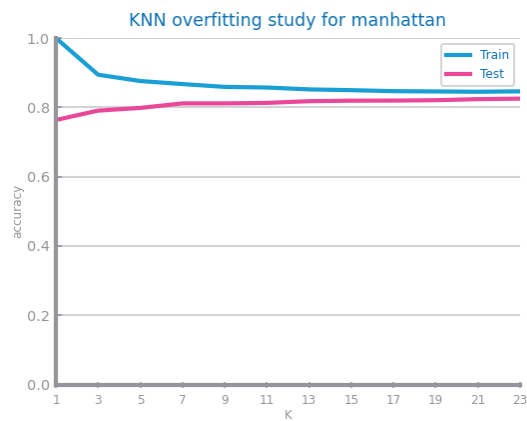Figure 5: KNN different parameterisations comparison for dataset 2

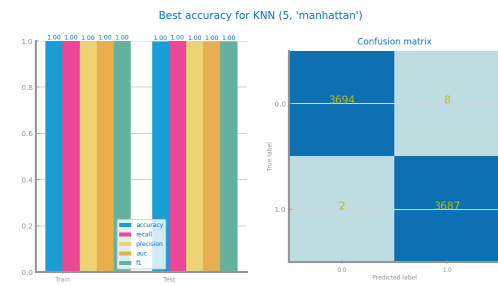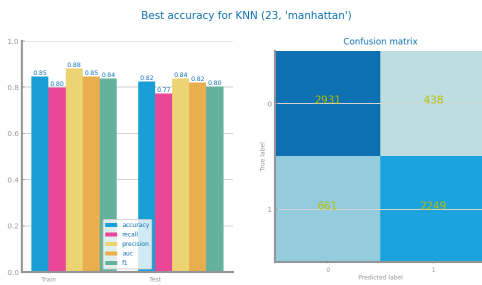Figure 6: KNN overfitting analysis for dataset 1 (left) and dataset 2 (right)

Figure 7: KNN best model results for dataset 1 (left) and dataset 2 (right)
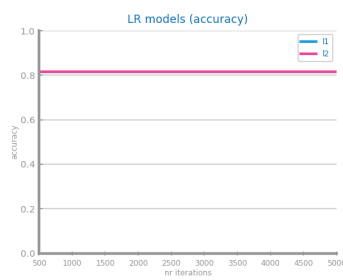
Logistic Regression

Figure 8: Logistic Regression different parameterisations comparison for dataset 1
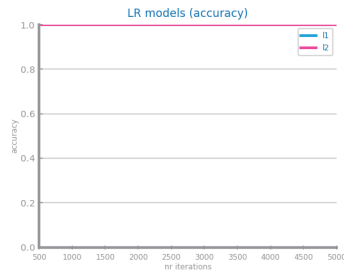
Figure 9: Logistic Regression different parameterisations comparison for dataset 2
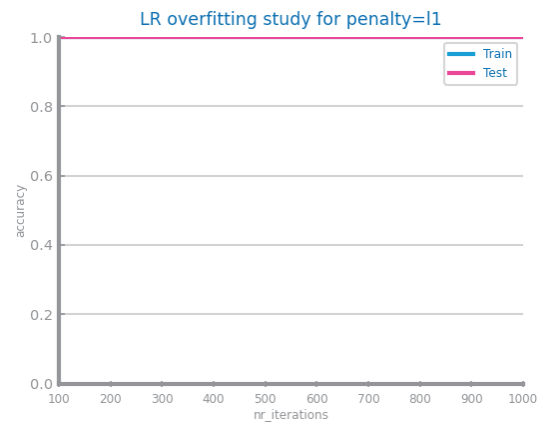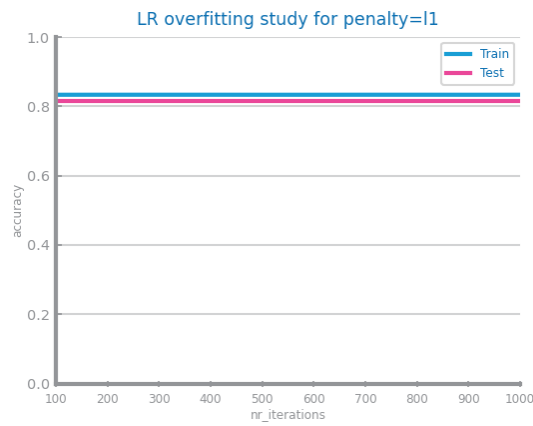

Figure 10: Logistic Regression overfitting analysis for dataset 1 (left) and dataset 2 (right)
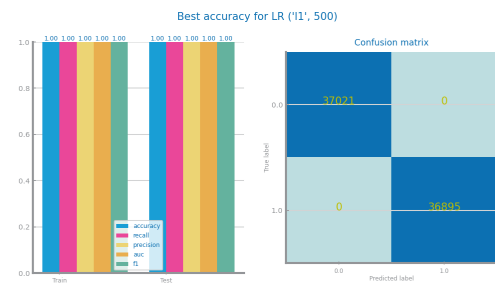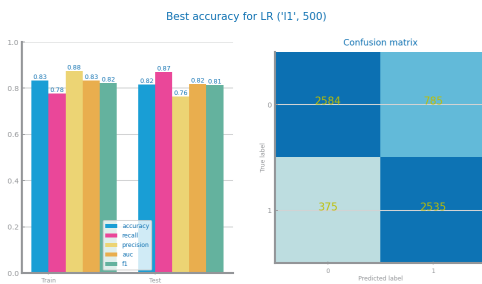

Figure 11: Logistic Regression best model results for dataset 1 (left) and dataset 2 (right)
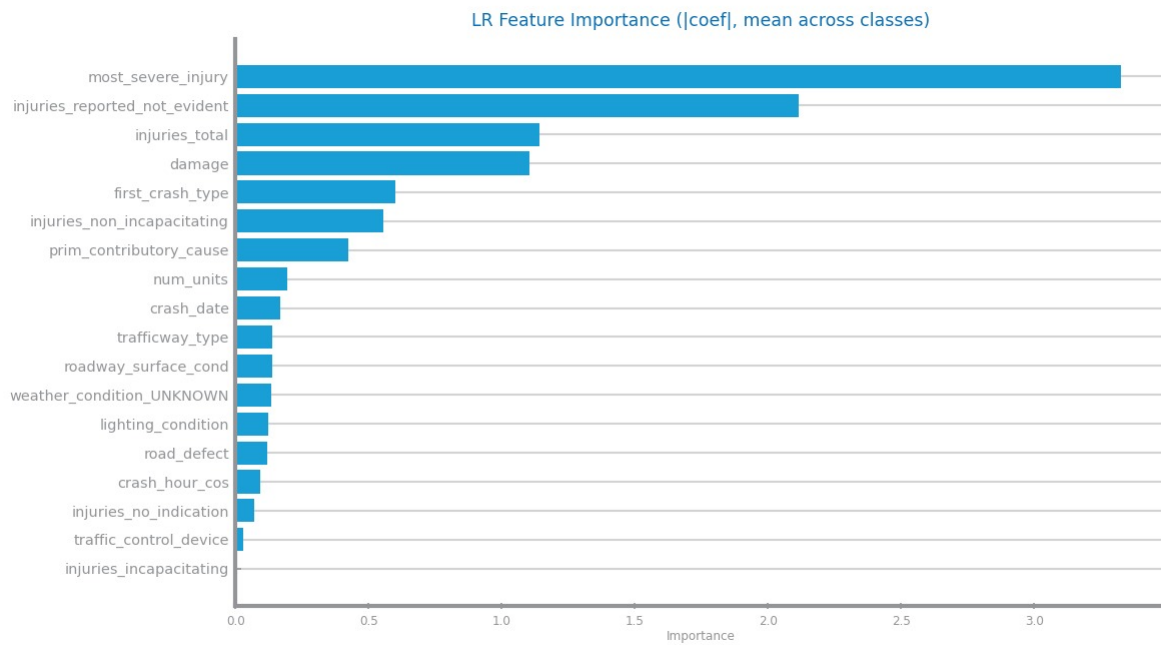
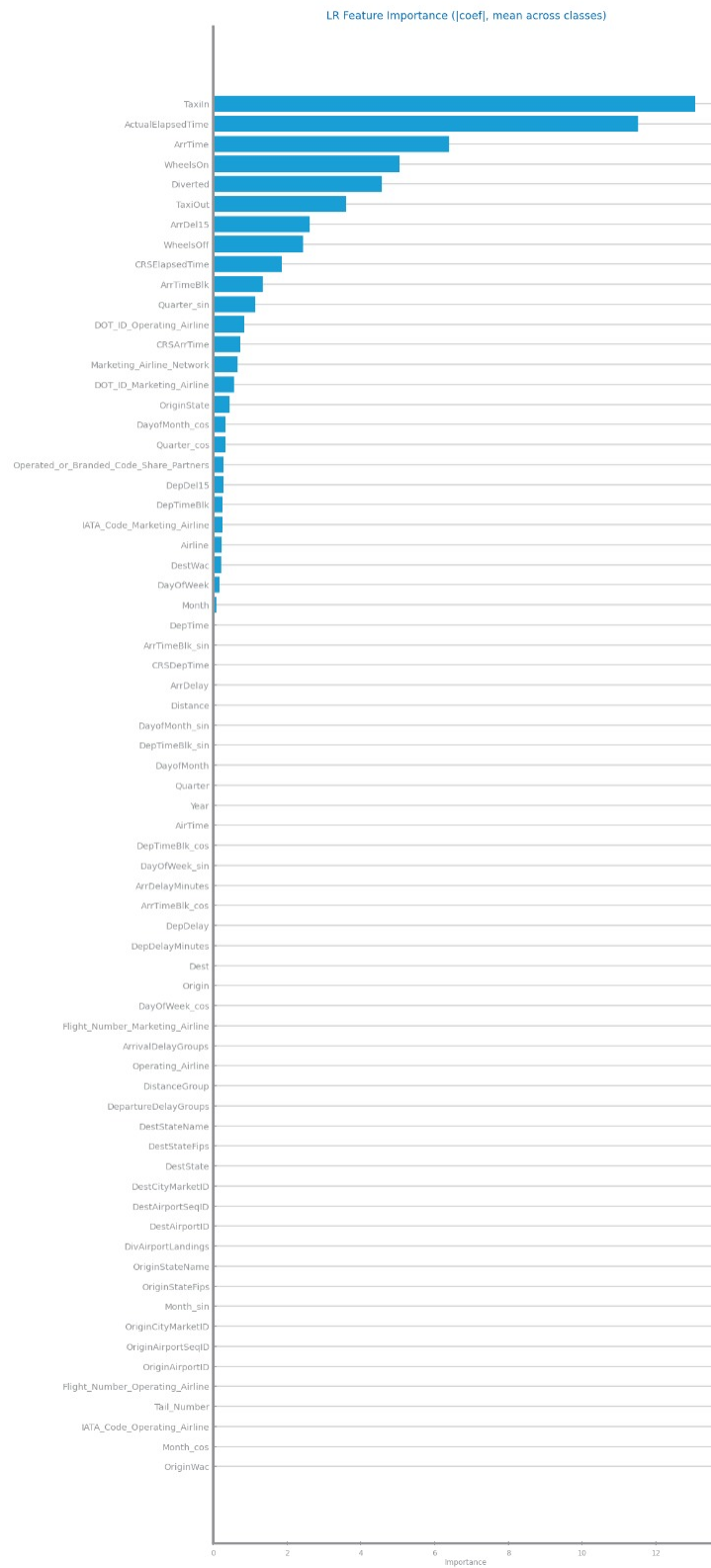Figure 12: Logistic Regression feature importance for dataset 1

Figure 13: Logistic Regression feature importance for dataset 2
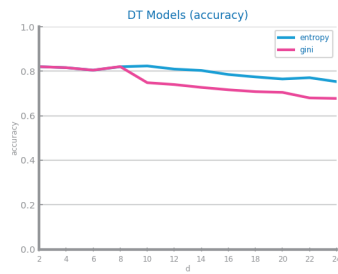
Decision Trees

Figure 14: Decision Trees different parameterisations comparison for dataset 1
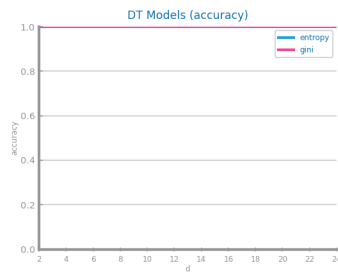

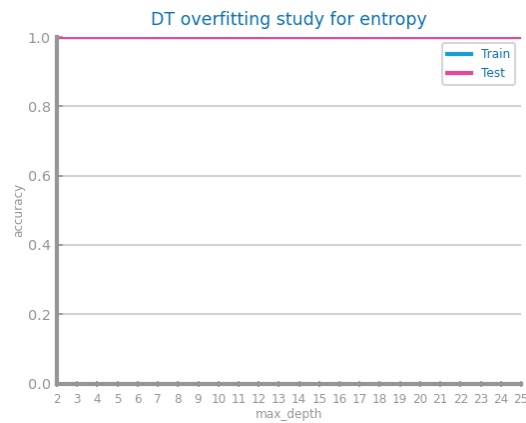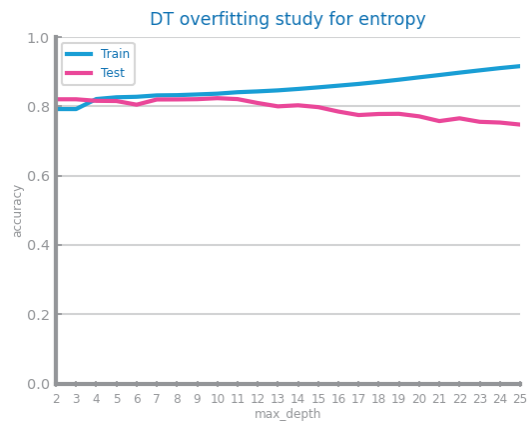Figure 15: Decision Trees different parameterisations comparison for dataset 2

Figure 16: Decision Trees overfitting analysis for dataset 1 (left) and dataset 2 (right)
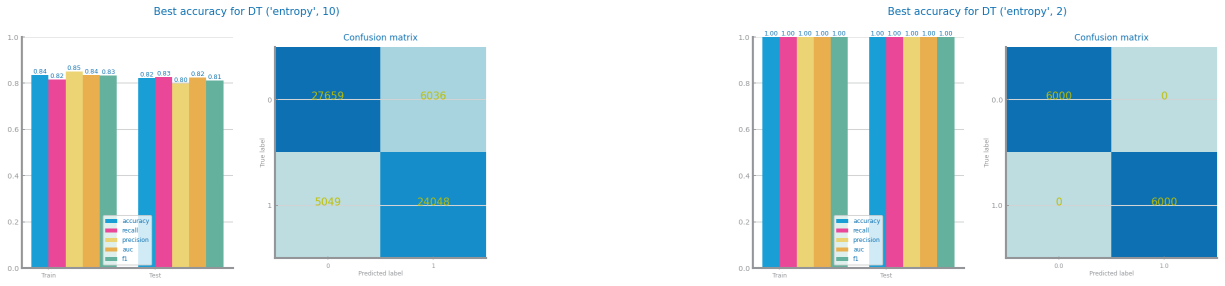
Figure 17: Decision trees best model results for dataset 1 (left) and dataset 2 (right)
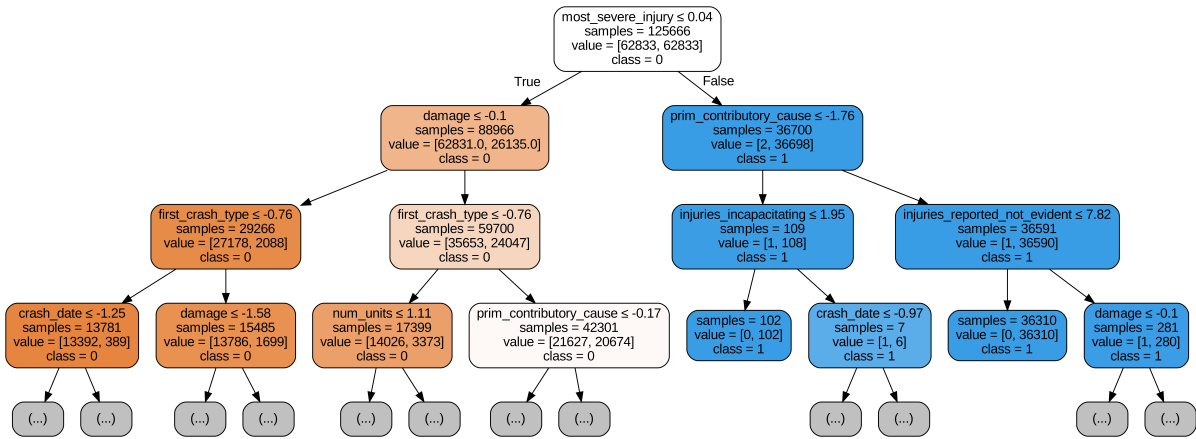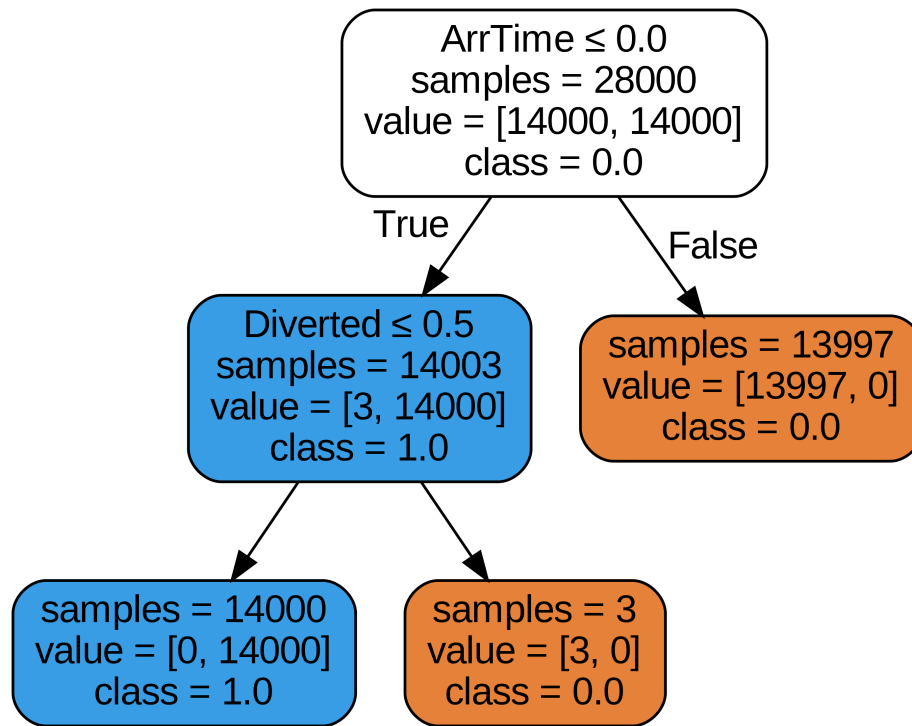

Figure 18: Best tree for dataset 1

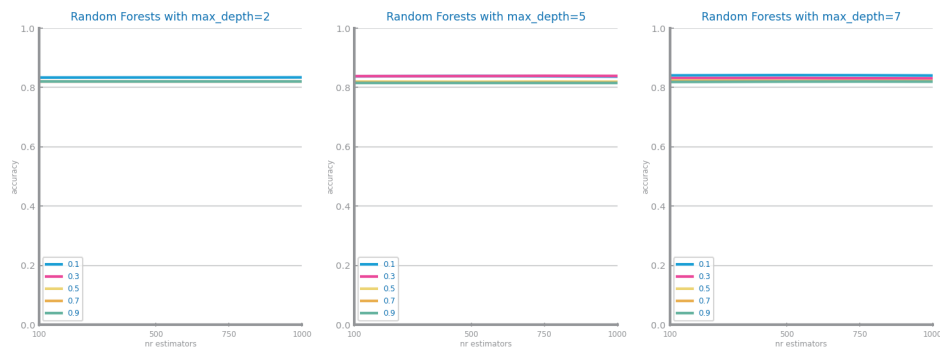Figure 19: Best tree for dataset 2

Random Forests

Figure 20: Random Forests different parameterisations comparison for dataset 1
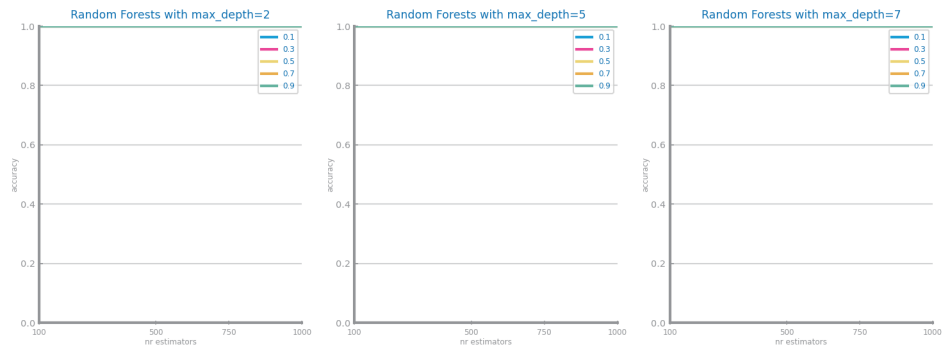
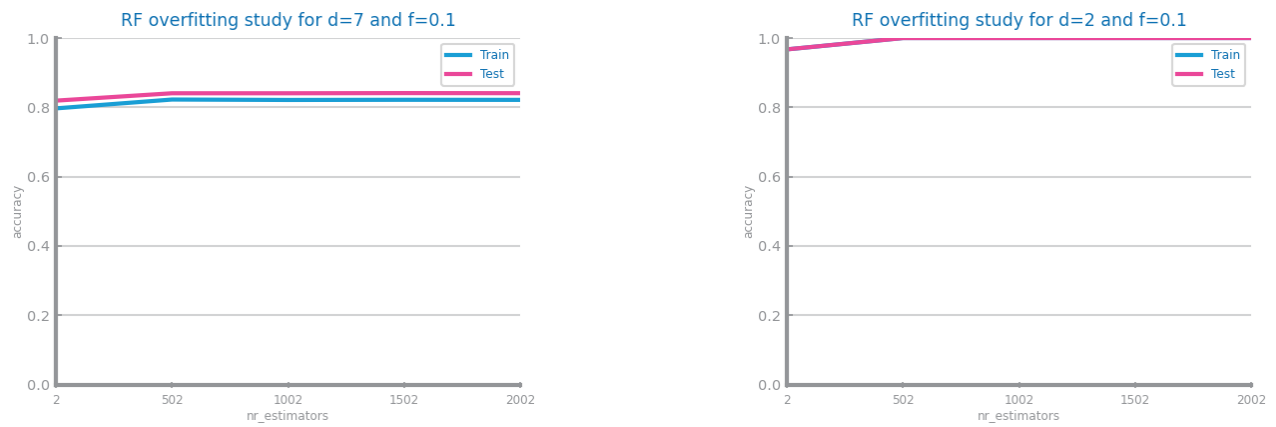Figure 21: Random Forests different parameterisations comparison for dataset 2

Figure 22: Random Forests overfitting analysis for dataset 1 (left) and dataset 2 (right)
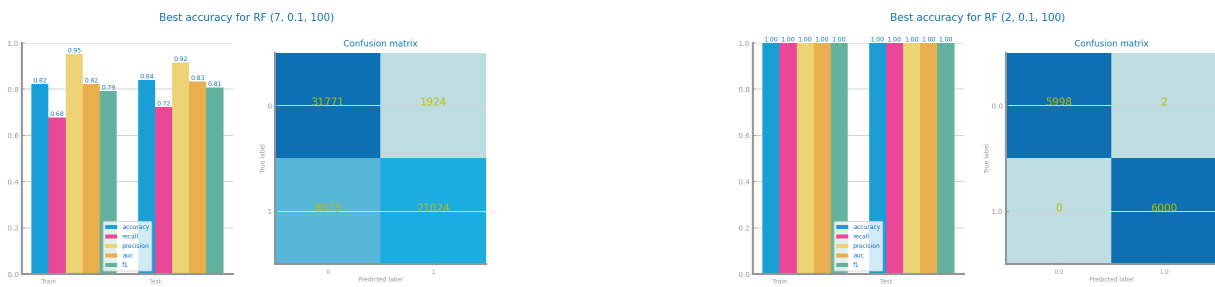

Figure 23: Random Forests best model results for dataset 1 (left) and dataset 2 (right)
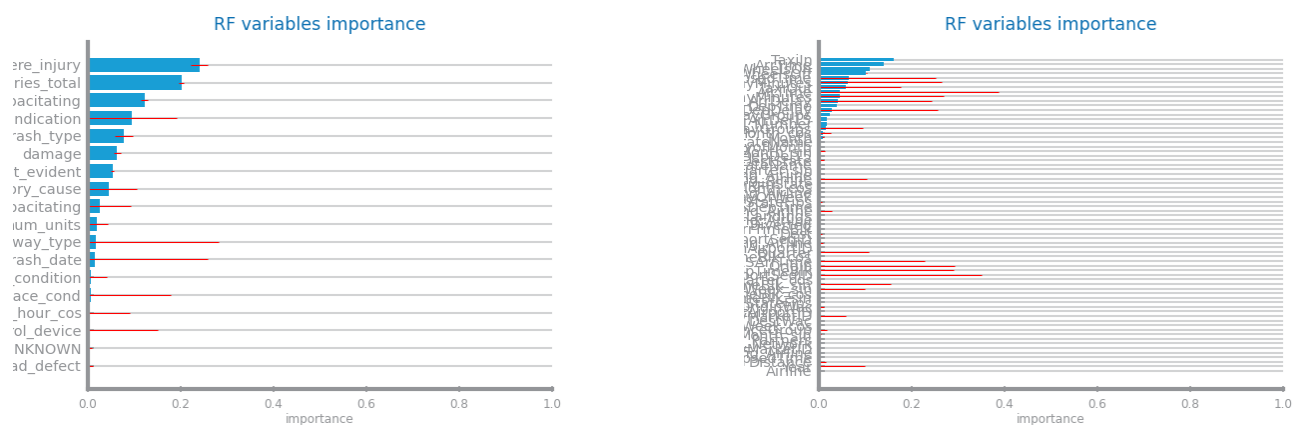
Figure 24: Random Forests variables importance for dataset 1 (left) and dataset 2 (right)
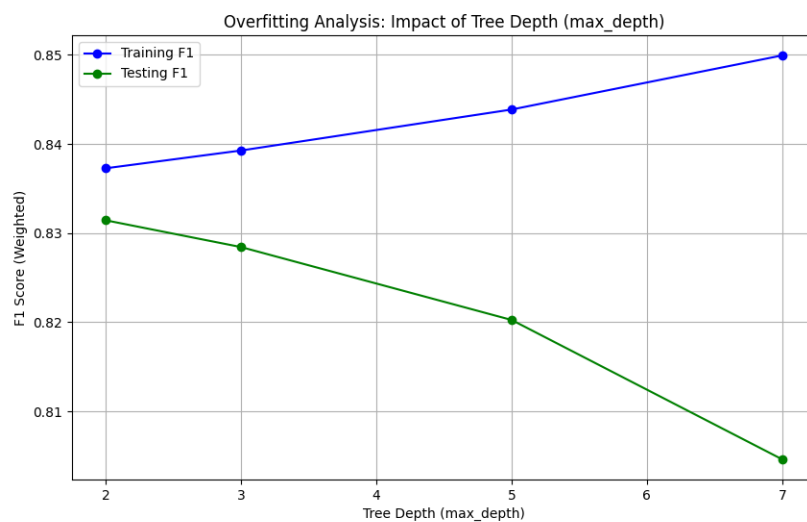
Gradient Boosting

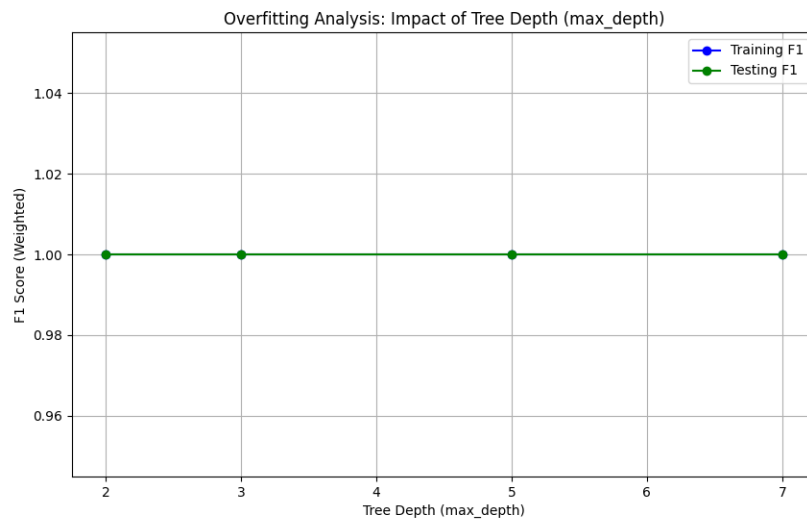Figure 25: Gradient boosting different parameterisations comparison for dataset 1

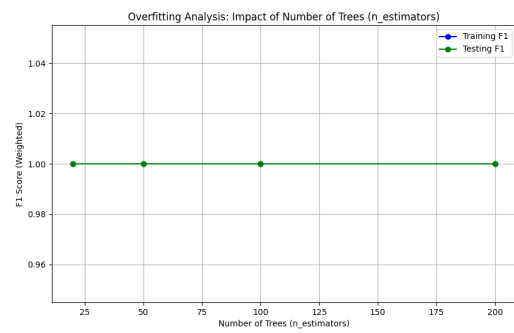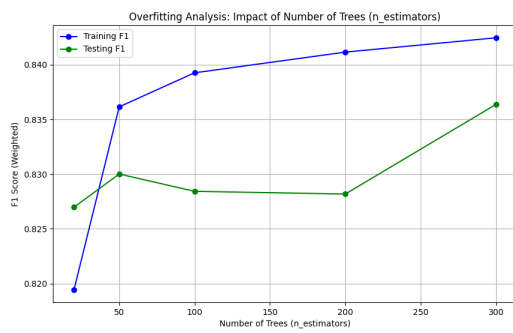Figure 26: Gradient boosting different parameterisations comparison for dataset 2

Figure 27: Gradient boosting overfitting analysis for dataset 1 (left) and dataset 2 (right)
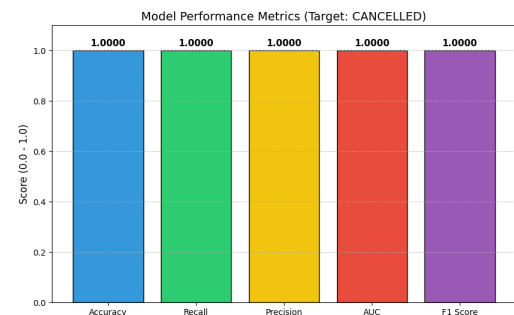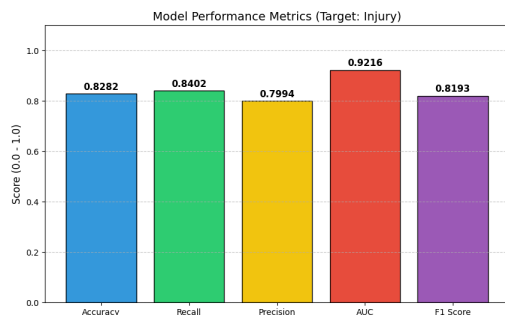

Figure 28: Gradient boosting best model results for dataset 1 (left) and dataset 2 (right)
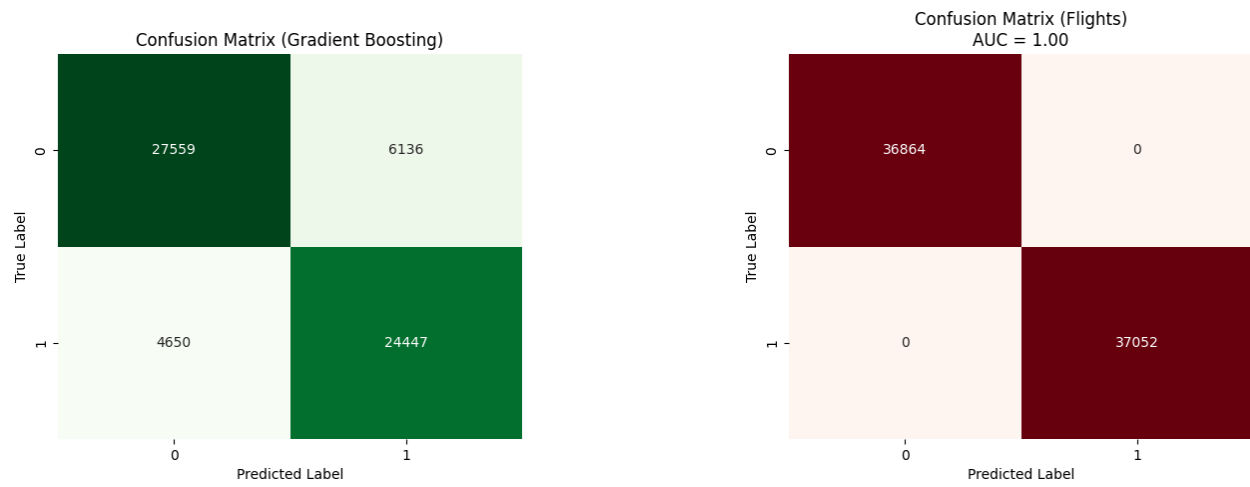
Figure 29: Gradient boosting variables importance for dataset 1 (left) and dataset 2 (right)
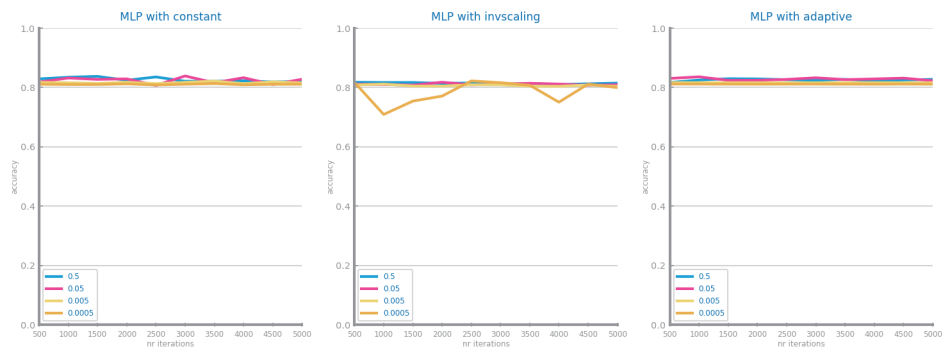
Multi-Layer Perceptrons

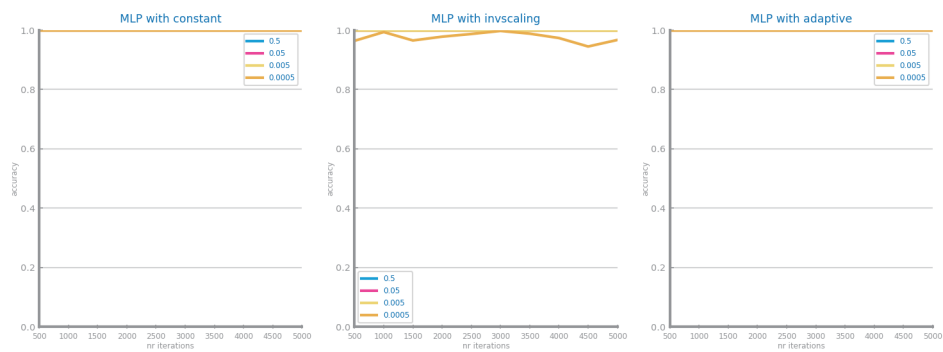Figure 30: MLP different parameterisations comparison for dataset 1

Figure 31: MLP different parameterisations comparison for dataset 2
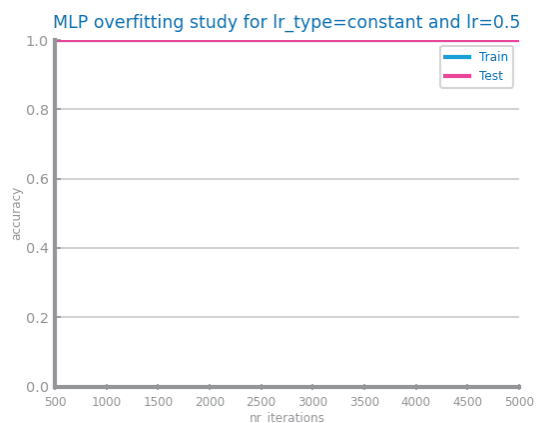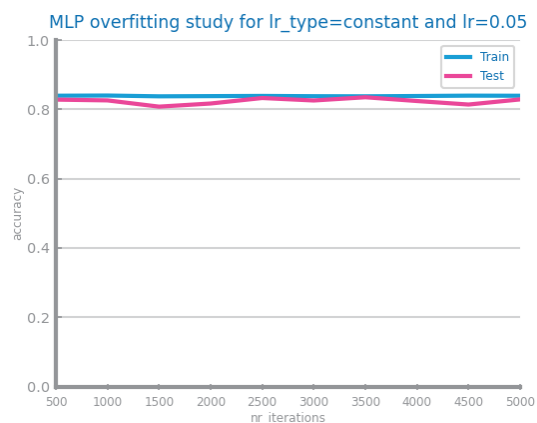
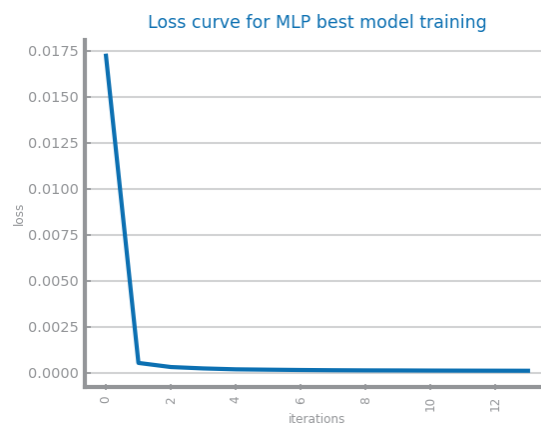Figure 32: MLP overfitting analysis for dataset 1 (left) and dataset 2 (right)
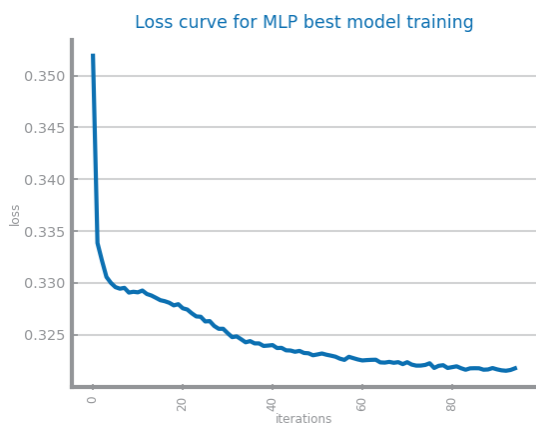

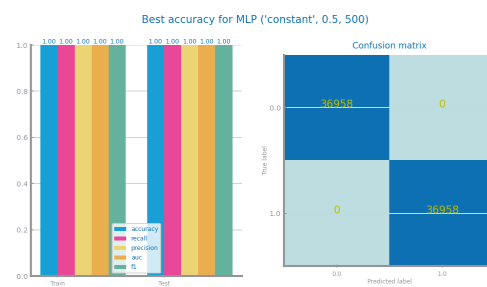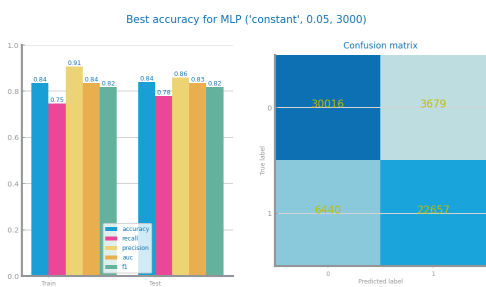Figure 33: Loss curve analysis for dataset 1 (left) and dataset 2 (right)

Figure 34: MLP best model results for dataset 1 (left) and dataset 2 (right)