

## Data Science Lab 4

Team nr: 2	<b>Student 1:</b> Åmund Grimstad	<b>IST nr:</b> 1116675
	<b>Student 2:</b> Arthur de Arruda Chau	<b>IST nr:</b> 1116090
	<b>Student 3:</b> Benjamin Raymond Kuhn	<b>IST nr:</b> 1115778
	<b>Student 4:</b> João Rafael Freitas Lourenço	<b>IST nr:</b> 425699

## CLASSIFICATION - MODELS' EVALUATION

### Traffic Accident Data

#### *Naïve Bayes*

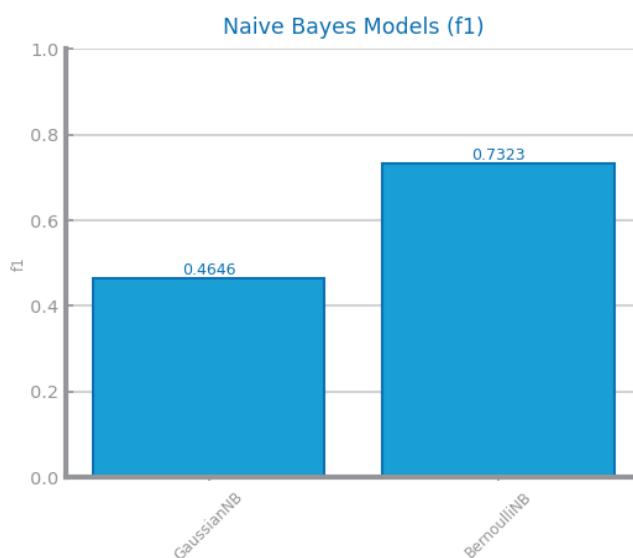


Figure 1: Naïve Bayes alternative comparison for Traffic Accident Data

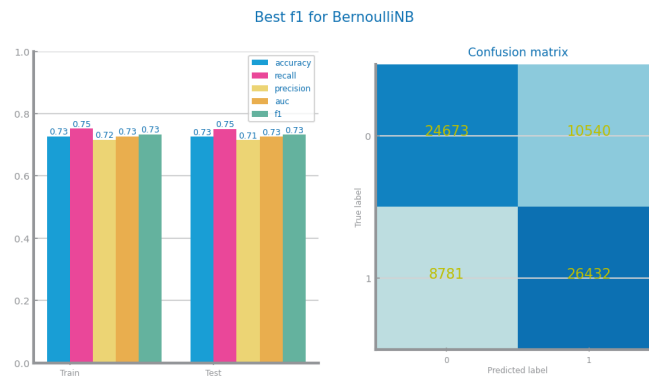


Figure 2: Naïve Bayes best model results for Traffic Accident Data

## KNN

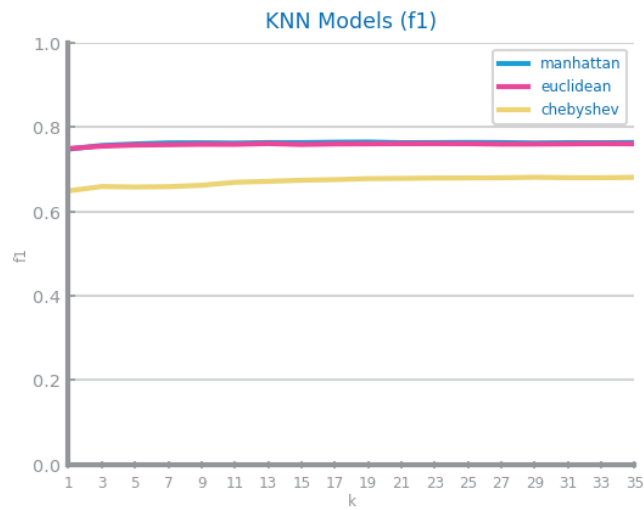


Figure 3: KNN different parameterisations comparison for Traffic Accident Data

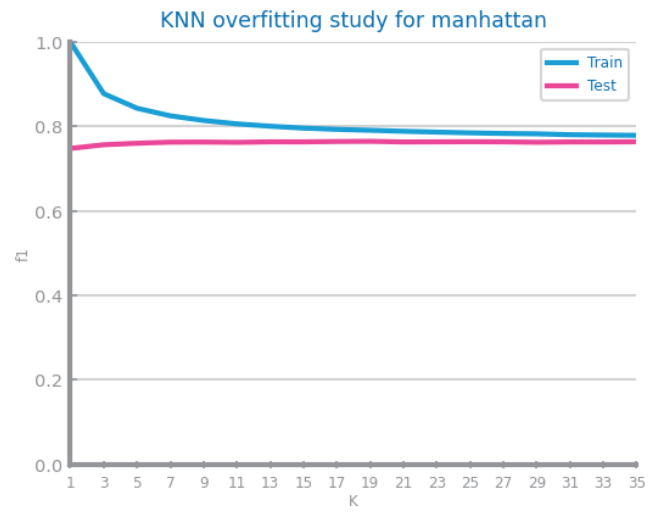


Figure 4: KNN overfitting analysis for Traffic Accident Data

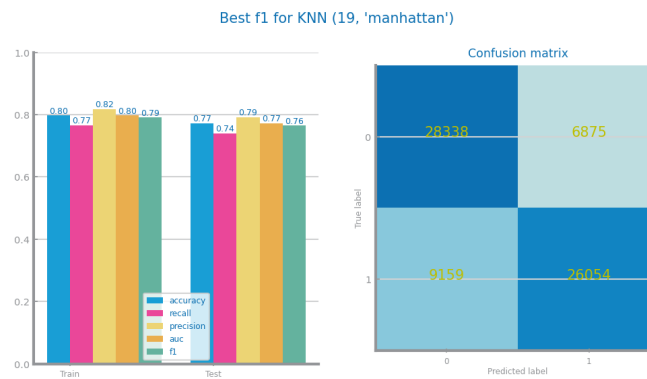


Figure 5: KNN best model results for Traffic Accident Data

## Decision Trees

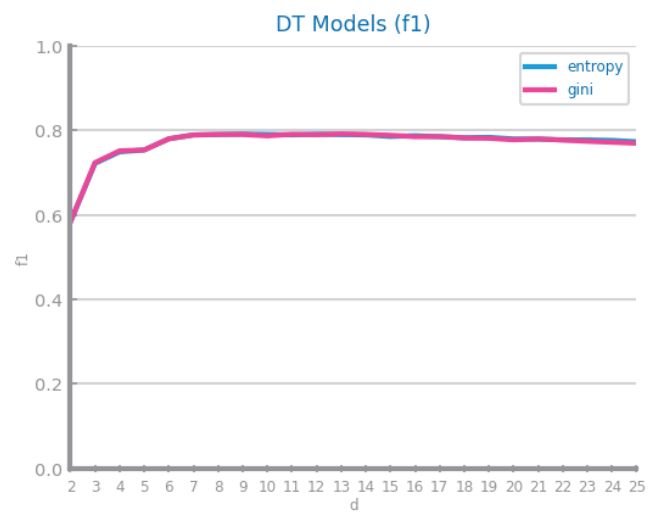


Figure 6: Decision Trees different parameterisations comparison for Traffic Accident Data

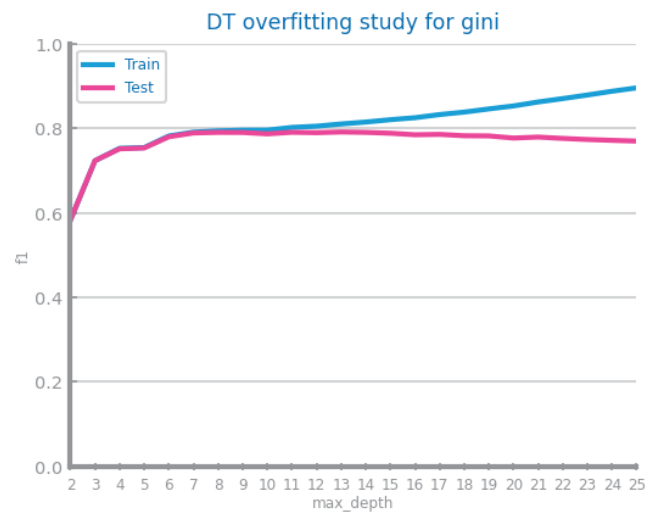


Figure 7: Decision Trees overfitting analysis for Traffic Accident Data

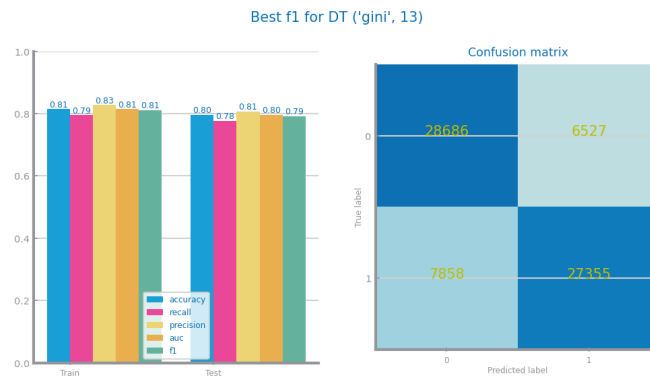


Figure 8: Decision trees best model results for Traffic Accident Data

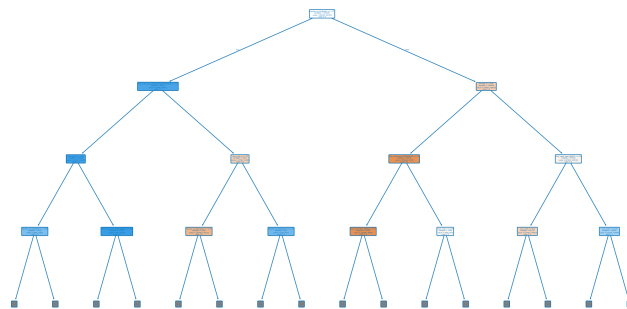


Figure 9: Best tree for Traffic Accident Data



Figure 10: Decision trees variables importance for Traffic Accident Data

## Random Forests

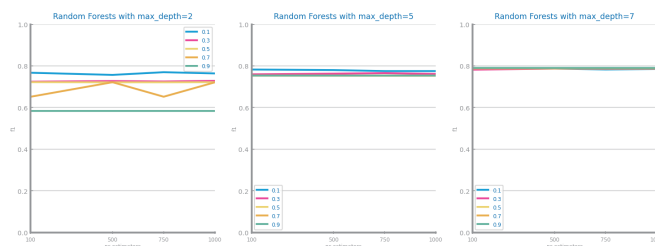


Figure 11: Random Forests different parameterisations comparison for Traffic Accident Data

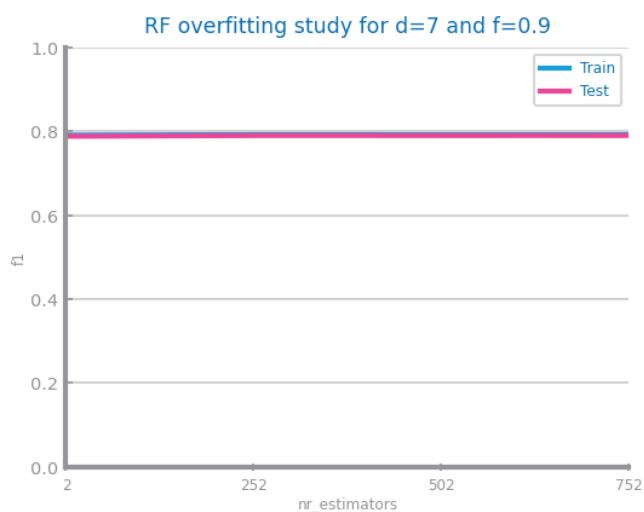


Figure 12: Random Forests overfitting analysis for Traffic Accident Data

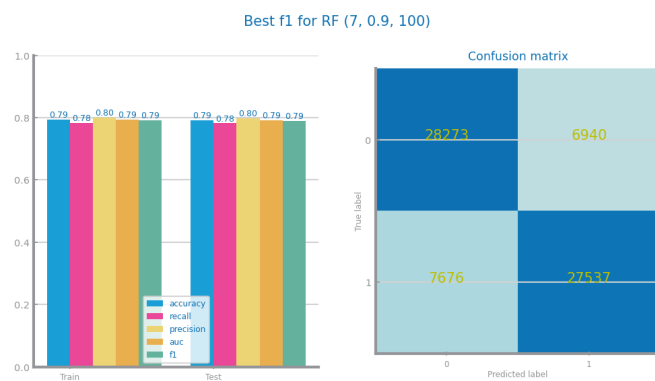


Figure 13: Random Forests best model results for Traffic Accident Data



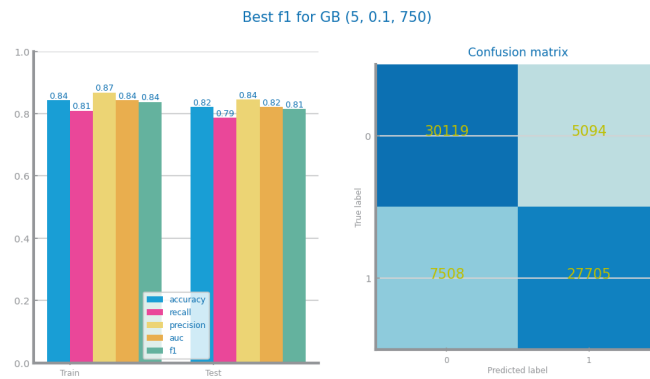


Figure 17: Gradient boosting best model results for Traffic Accident Data



Figure 18: Gradient boosting variables importance for Traffic Accident Data

## Logistical Regression

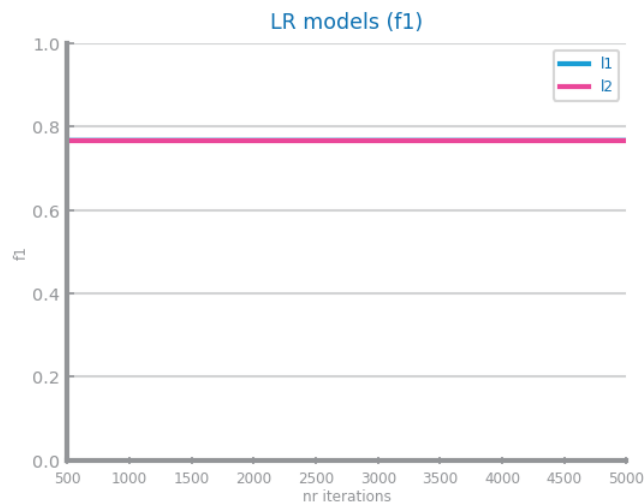


Figure 19: Logistical regression different parameterisations comparison for Traffic Accident Data



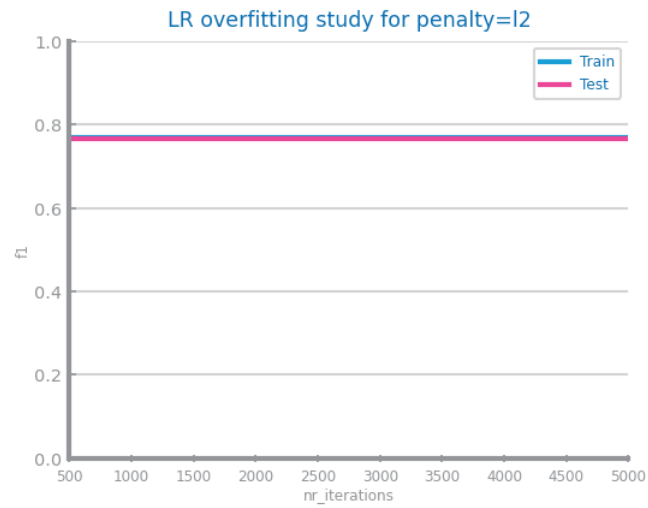


Figure 20: Logistical regression overfitting analysis for Traffic Accident Data

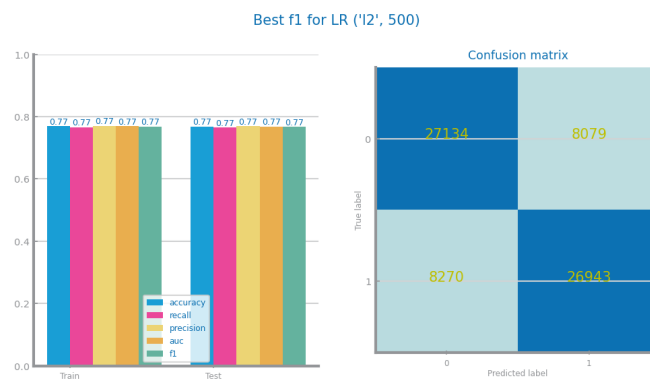


Figure 21: Logistical regression best model results for Traffic Accident Data

## Multi-Layer Perceptrons

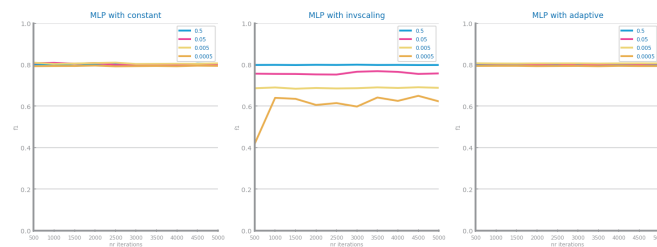


Figure 22: MLP different parameterisations comparison for Traffic Accident Data

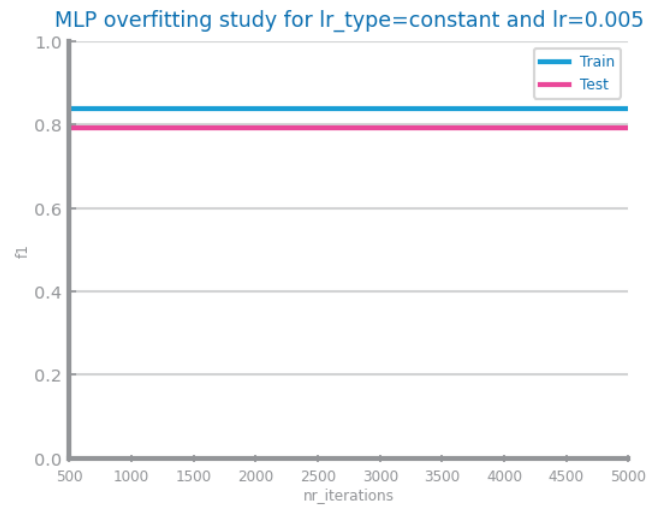


Figure 23: MLP overfitting analysis for Traffic Accident Data

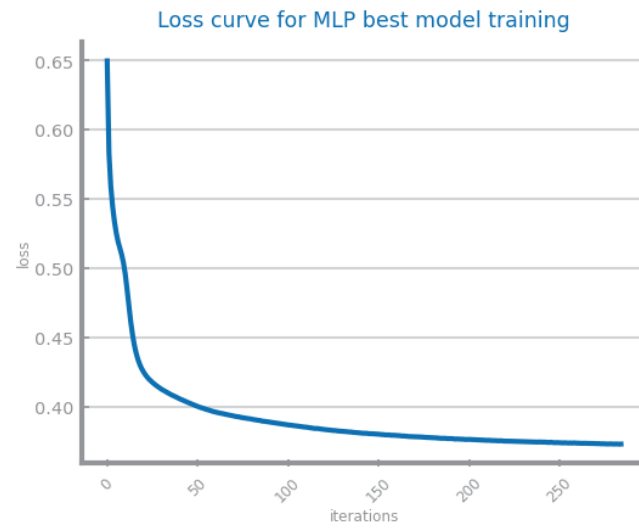


Figure 24: Loss curve analysis for Traffic Accident Data

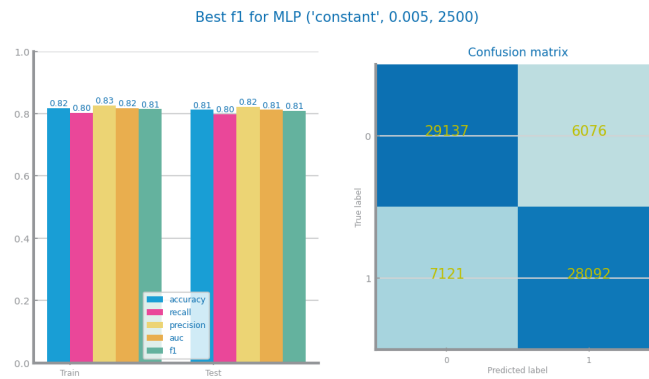


Figure 25: MLP best model results for Traffic Accident Data

## Flight Cancellation Data

### *Naïve Bayes*

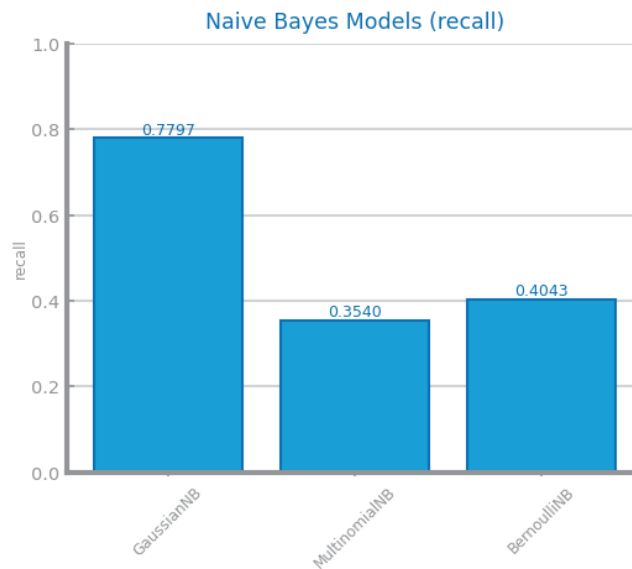


Figure 26: Naïve Bayes alternative comparison for Flight Cancellation Data

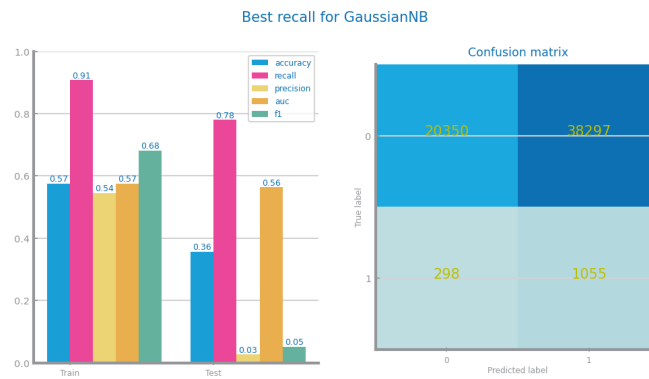


Figure 27: Naïve Bayes best model results for Flight Cancellation Data

## KNN

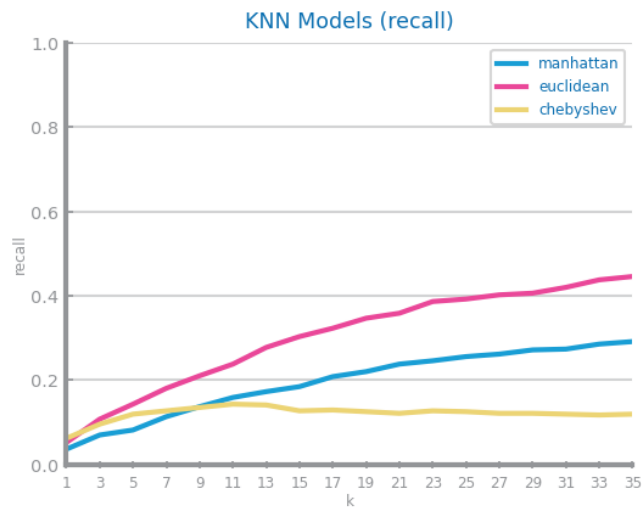


Figure 28: KNN different parameterisations comparison for Flight Cancellation Data

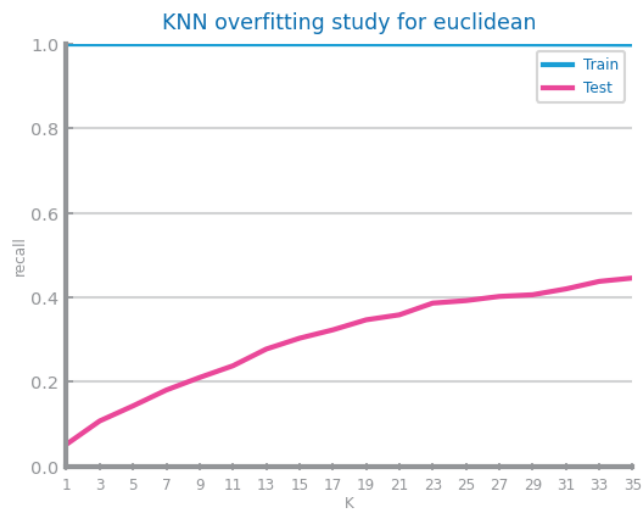


Figure 29: KNN overfitting analysis for Flight Cancellation Data

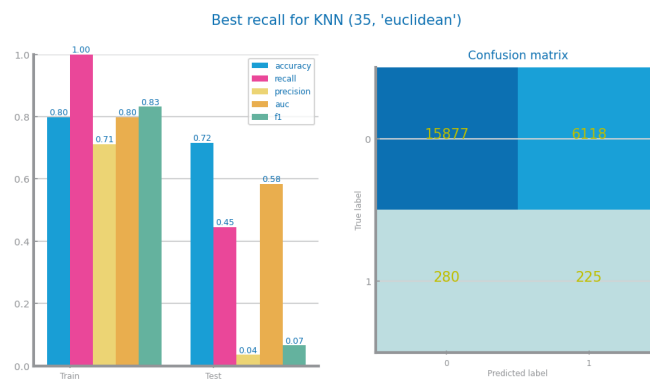


Figure 30: KNN best model results for Flight Cancellation Data

## Decision Trees

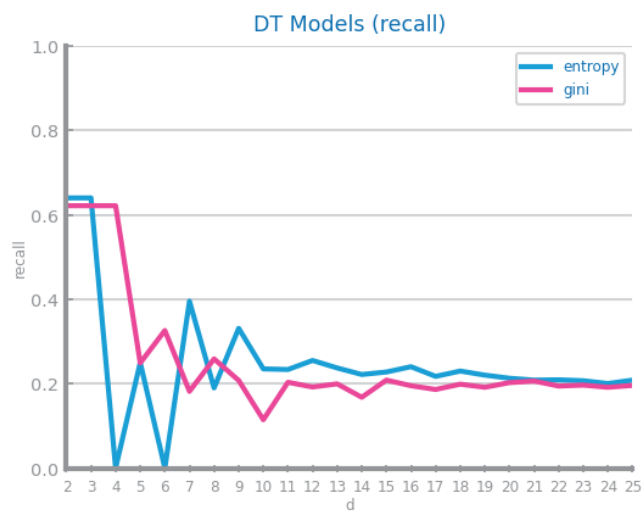


Figure 31: Decision Trees different parameterisations comparison for Flight Cancellation Data

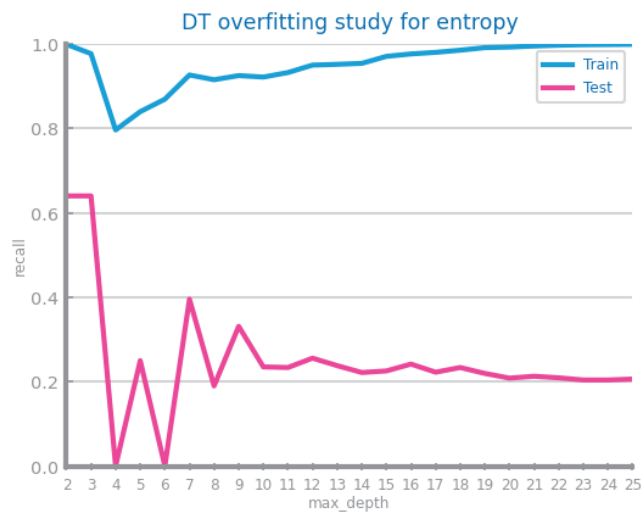


Figure 32: Decision Trees overfitting analysis for Flight Cancellation Data

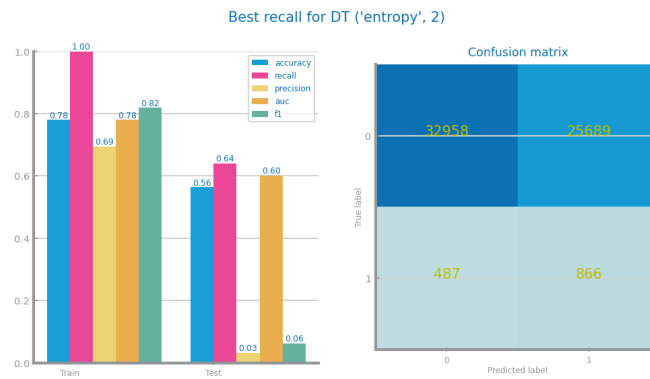


Figure 33: Decision trees best model results for Flight Cancellation Data

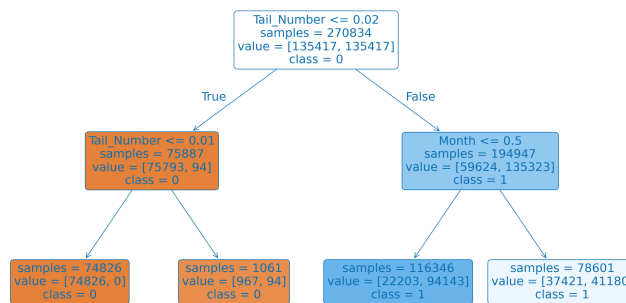


Figure 34: Best tree for Flight Cancellation Data

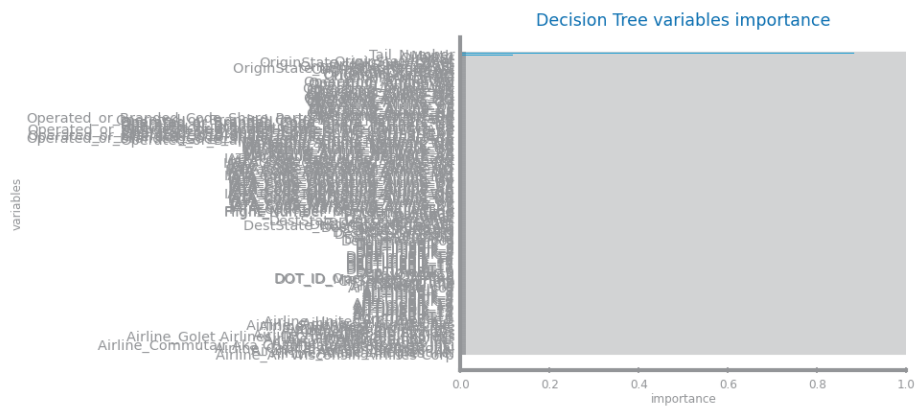


Figure 35: Decision trees variables importance for Flight Cancellation Data

## Random Forests

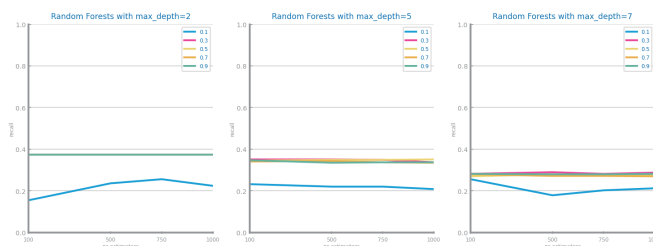


Figure 36: Random Forests different parameterisations comparison for Flight Cancellation Data

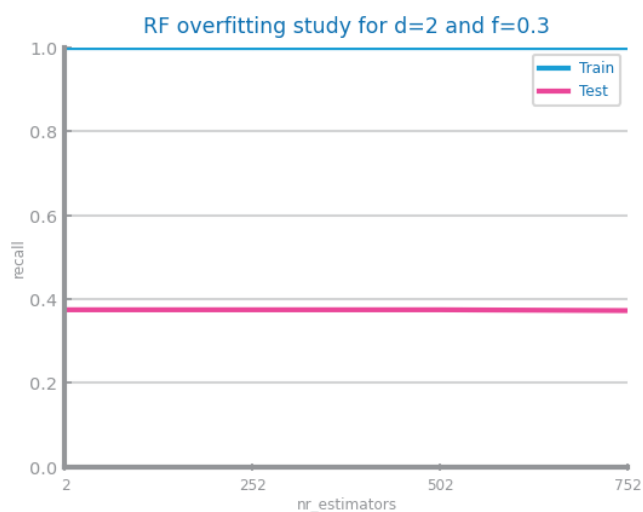


Figure 37: Random Forests overfitting analysis for Flight Cancellation Data

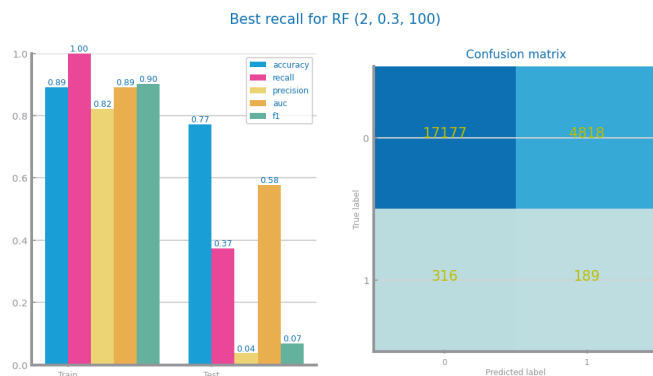


Figure 38: Random Forests best model results for Flight Cancellation Data



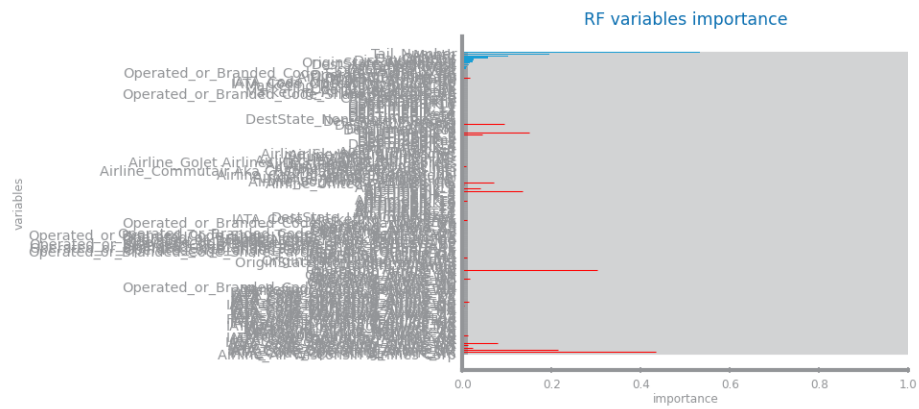


Figure 39: Random Forests variables importance for Flight Cancellation Data

## Gradient Boosting

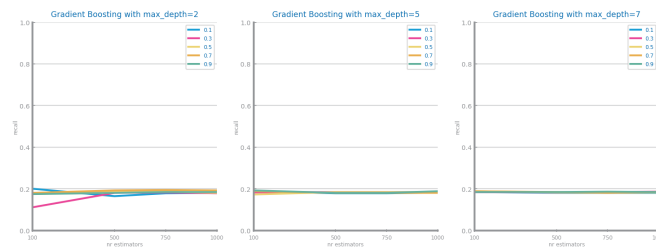


Figure 40: Gradient boosting different parameterisations comparison for Flight Cancellation Data

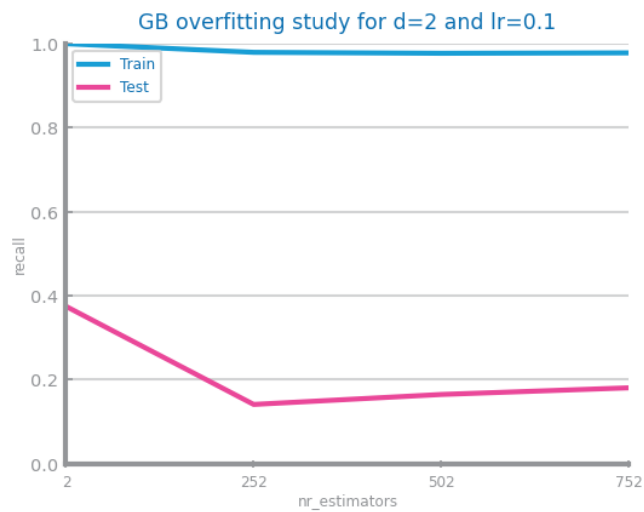


Figure 41: Gradient boosting overfitting analysis for Flight Cancellation Data

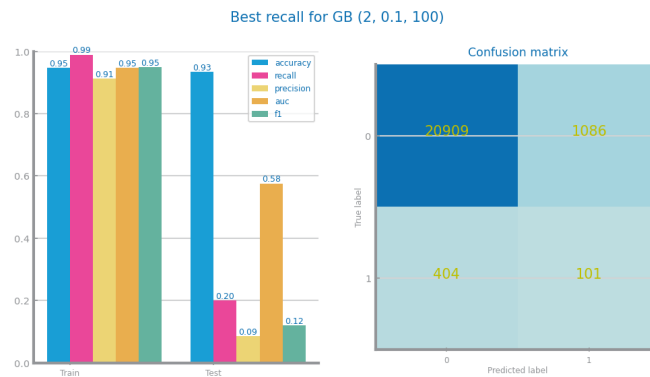


Figure 42: Gradient boosting best model results for Flight Cancellation Data

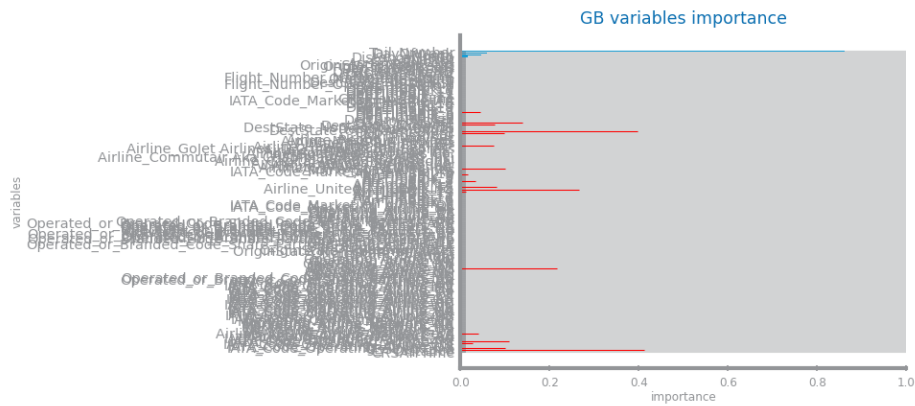


Figure 43: Gradient boosting variables importance for Flight Cancellation Data

## Logistical Regression

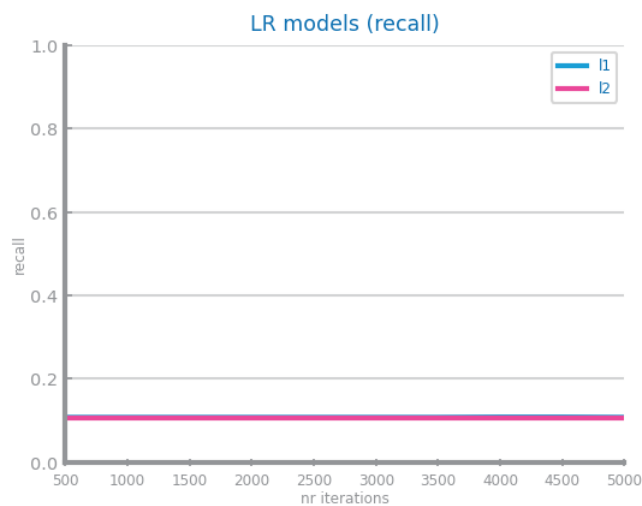


Figure 44: Logistical regression different parameterisations comparison for Flight Cancellation Data

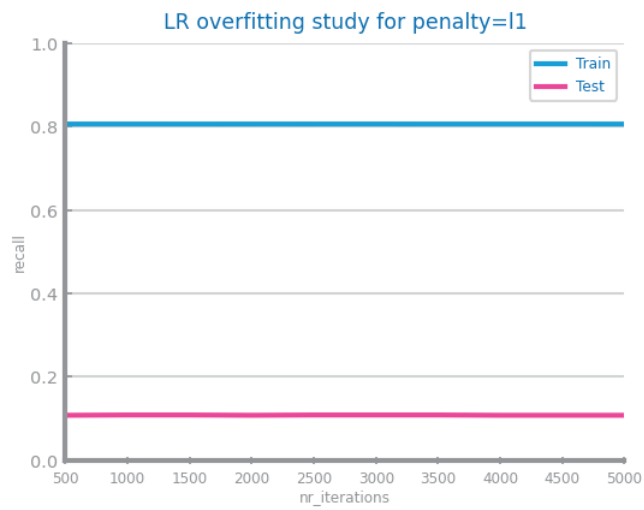


Figure 45: Logistical regression overfitting analysis for Flight Cancellation Data

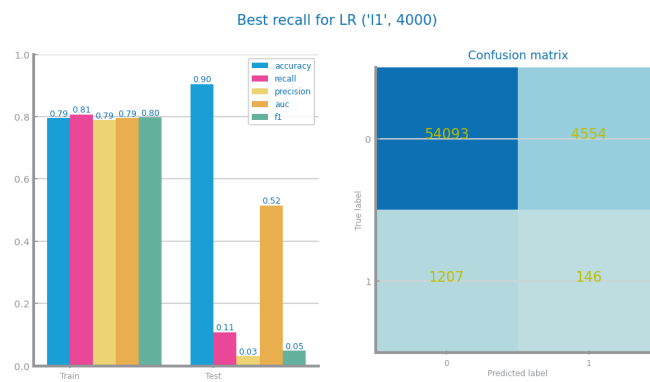


Figure 46: Logistical regression best model results for Flight Cancellation Data

Multi-Layer Perceptrons

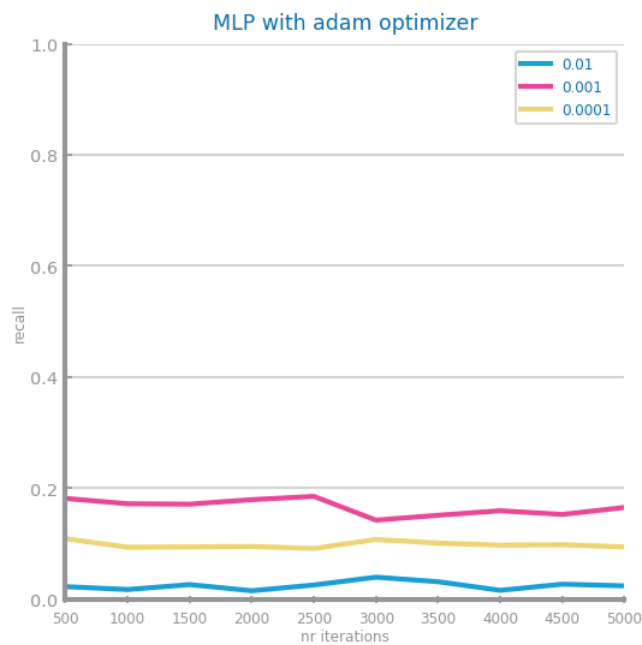


Figure 47: MLP different parameterisations comparison for Flight Cancellation Data

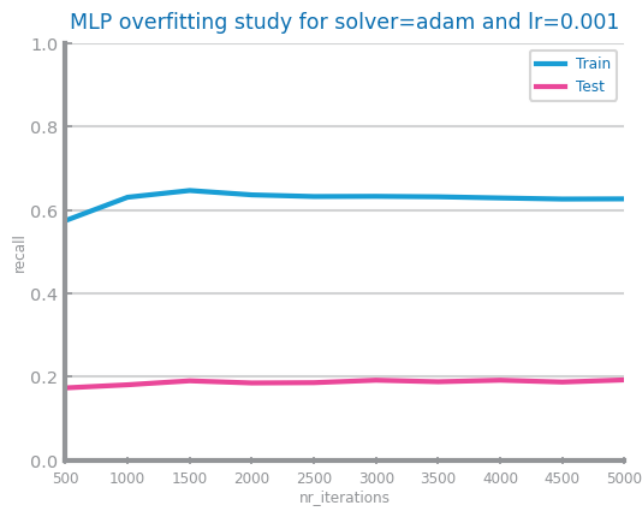


Figure 48: MLP overfitting analysis for Flight Cancellation Data

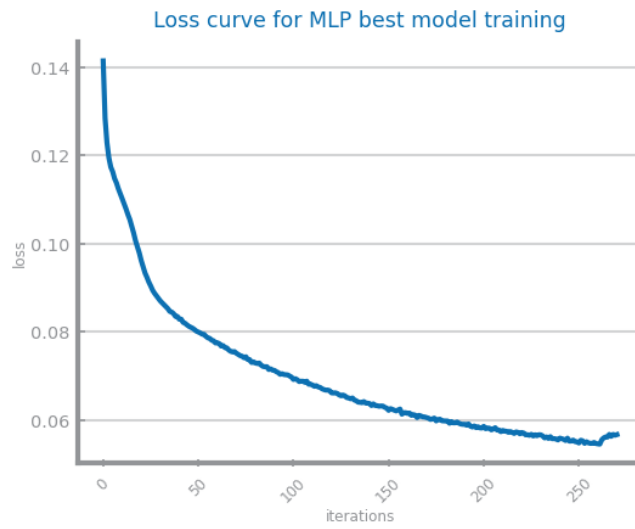


Figure 49: Loss curve analysis for Flight Cancellation Data

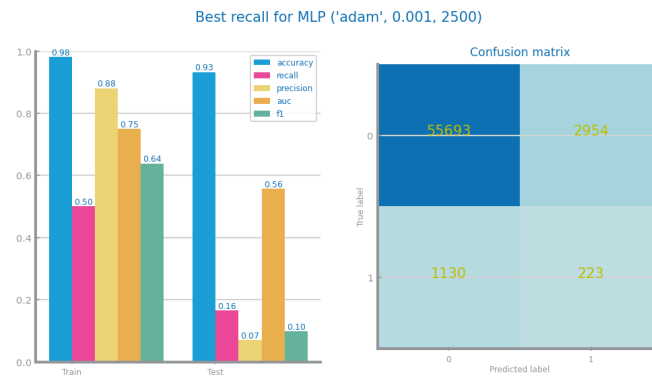


Figure 50: MLP best model results for Flight Cancellation Data