DS 2025

Data Science Lab 2

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

CLASSIFICATION – DATA PREPARATION

1 Traffic Accident Data

Variables Encoding

Missing Value Imputation

Figure 1: Missing values imputation results with different approaches for Traffic Accident Data

Outliers Treatment

Figure 2: Outliers imputation results with different approaches for Traffic Accident Data

Scaling

Figure 3: Scaling results with different approaches for Traffic Accident Data

Balancing

Figure 4: Balancing results with different approaches for Traffic Accident Data

Feature Selection

Figure 5: Feature selection of redundant variables results with different parameters for Traffic Accident Data

Figure 6: Feature selection of relevant variables results with different parameters for Traffic Accident Data (variance study)

Feature Extraction (optional)

Figure 7: Principal components analysis and feature extraction results for Traffic Accident Data

Additional Feature Generation (if done)

Figure 8: Feature generation results for Traffic Accident Data

Best approach

Figure 9: Performance and confusion matrix for best approach for Traffic Accident Data

2 Flight Cancellation Data

Variables Encoding

Missing Value Imputation

Figure 10: Missing values imputation results with different approaches for Flight Cancellation Data

Outliers Treatment

Figure 11: Outliers imputation results with different approaches for Flight Cancellation Data

Scaling

Figure 12: Scaling results with different approaches for Flight Cancellation Data

Balancing

Figure 13: Balancing results with different approaches for Flight Cancellation Data

Feature Selection

Figure 14: Feature selection of redundant variables results with different parameters for Flight Cancellation Data

Figure 15: Feature selection of relevant variables results with different parameters for Flight Cancellation Data (variance study)

Feature Extraction (optional)

Figure 16: Principal components analysis and feature extraction results for Flight Cancellation Data

Additional Feature Generation (if done)

Figure 17: Feature generation results for Flight Cancellation Data

Best approach

Figure 18: Performance and confusion matrix for best approach for Flight Cancellation Data