Appendix D: DMAIC and CRISP-DM integration

*) Indicator for the framework to which the selected tool belongs. D = DMAIC, C = CRISP-DM

DMAIC	CRISP-DM	*	Activity/Tool	Result
Define	Business Understanding	D	Project Charter	Project (goal) definition and communication: reduction of overtime for repairs by > 40%
		D	Business Case	Commitment by process owners
		D	VOC-CTS	Customer requirements and output variables; Refined project goals
		D	SIPOC / RACI	Gain a common understanding of the process
		D	In- / out of scope	Identification of relevant processes and deliverables
		D	Stakeholder Analysis	Identification of project participants, decision-makers and individuals impacted by the project
Measure	Data Understanding	D	Operational Definition	Definition of parameters and how they are measured
		D	Data Collection Plan	Plan to collect the data
		C	Cluster Analysis, unsupervised (k-means)	Confirmation that process parameters measured by different machines are similar. Therefore, the team continued with the data collection.
		D + C	Graphical Analysis	Used both Six Sigma and CRISP-DM graphical tools: Control chart: Process stability and capability Seaborn plots: Visual detection of anomalies. Correlation matrix: Identification of the 2 most relevant input parameters (welding current and voltage)
		C	Measurement System Analysis	Six Sigma tool enhanced by big data analytics: Confirmation of data quality (measured input parameters)
	Data Preparation	C	Scitkit Learn RobustScaler	Preprocessed, scaled data for data evaluation and modelling
		C	Defect Annotation / Data labelling	Joined view on time (welding data) and location data (error location from quality protocol)
		C	Feature Selection	Identification of the parameter settings for machine learning models

Appendix D: DMAIC and CRISP-DM integration

	Modeling	C	Machine Learning Algorithms/ Model selection	Models that effectively predict defects based on parameters set. Selected best models.
Analyse		C	Data Augmentation	More versatile training data based on the data collected
		C	Stratified k-fold Splitting	Even distribution of train and validation datasets
	Evaluation	D	Process diagram	Visualisation of process steps and roles involved
		D	Ishikawa	Identified root causes of organisational nature
		D	5-times Why	Identified root causes that can be influenced
		D	FMEA	Identified risks and decided on mitigation strategies
		C	10-fold cross validation	Data Augmentation improved the performance of the algorithms KNN, NN, catboost, but not for MPCNN. Selected NN for baseline-testing and MPCNN for model deployment.
Improve		C	Model Training	Collect new data to train the MPCNN model further and optimise it.
	Deployment	D	PICK Matrix	Prioritised improvement actions, as not all can be implemented.
		D	Implementation Plan	Plan by whom and by when improvement actions should be implemented.
Control		D	Control Plan	Instructions for monitoring and counteractions for the operations team.
		C	MLops framework	For recurring training and deployment