# CRISP ML(Q):

1. **Business and Data Understanding**
   1. Scope of Application:   
      Business Needs
   2. Measurable Success Criteria
      1. Business Success Criteria:
         1. Purpose and Success Criteria from a business point of view
      2. ML Success Criteria:
         1. Define minimum acceptable level of performance to meet the business goals
      3. Economic Success Criteria
         1. Objective ML Success by Introduction of KPI(s)
   3. Feasibility
      1. Applicability of ML Technology
      2. Legal Constraints
      3. Requirements on the application
         1. Robustness
         2. Scalability
         3. Explainability
         4. Resource Demand
   4. Data Collection
      1. Data Collection Plan
         1. In terms of Cost and Time needed to collect enough consistent data
      2. Data Version Control
         1. Data is collected iteratively, hence (planed) modifications of the data set should be documented
   5. Data Quality Verification
      1. Data Description
         1. Expert Knowledge regarding data sets like expected value ranges of features, maximum number of missing values. Guide to identify non-plausible data
      2. Data Requirements
      3. Data Verification
         1. Initial data, added data & production data must be checked according to the requirements
   6. Review of Output Documents
2. **Data Preperation**
   1. Select Data
   2. Select Features:
      1. Can be separated in three categories:
         1. Filter Methods
         2. Wrapper Methods
         3. Embedded Methods
      2. Would be good to have someone with expert knowledge have a look at it again.
      3. Data Selection
         1. Discarding features/samples should be well documented and strictly be based on objective quality criteria’s
      4. Unbalanced Classes
   3. Clean Data
      1. Noise reduction
      2. Data imputation
   4. Construct Data
      1. Feature Engineering
      2. Data augmentation
   5. Standardized Data
      1. File Format
      2. Normalisation
3. **Modelling**
   1. Literature Research on similar Problems
   2. Define Quality Measures of the model
      1. Robustness
      2. Explainability
      3. Scalability
      4. Resource Demand
      5. Model Complexity
   3. Model Selection
   4. Incorporate domain knowledge
   5. Model Training
   6. Model Compression
   7. Ensemble methods
   8. Result reproducibility
   9. Experimental Documentation
4. **Evaluation:** 
   1. Validate Performance
      1. Come up with a plan to validate the performance
   2. Determine robustness
   3. Increase Explainability
   4. Compare results with defined success criteria
5. **Deployment:**
   1. Define inference hardware
   2. Model evaluation under production condition
   3. Assure User Acceptance and usability
   4. Minimize risks of unforeseen errors
   5. Deployment strategy
6. **Monitoring and Maintenance:** 
   1. Identify Risks for performance degradation:
      1. Non-stationary data distributions/data drift
      2. Degradation of hardware
      3. System updates
   2. Monitor
      1. Monitor all input signals and compared to trainings data -> in this way, updates in the input data could be caught. What to do with anomalies in the input?
      2. Monitoring History of Performance
   3. Update