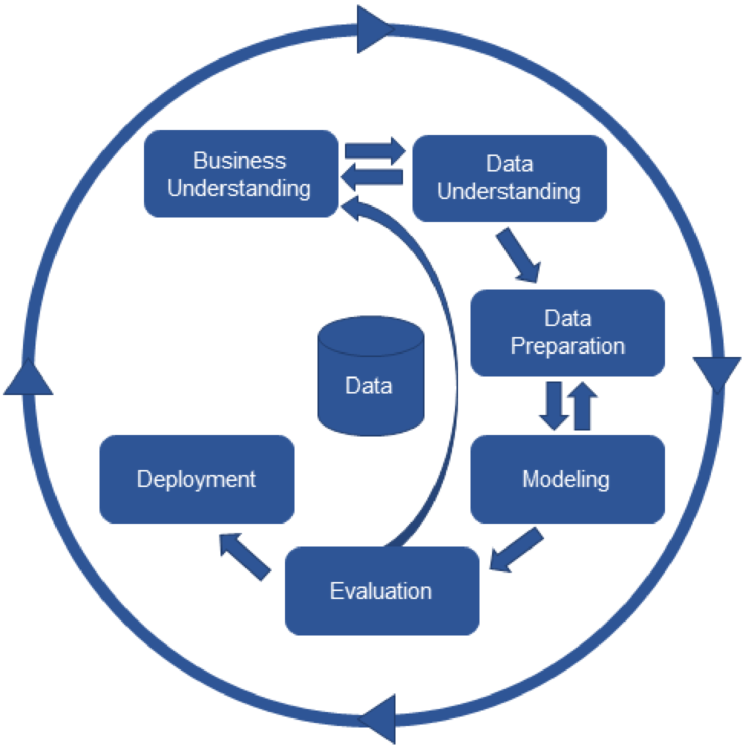
CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology

Business Problem: Quality assurance for ML models

Introduction

The Cross Industry Standard Process for Data Mining (CRISP-DM) is a process model with six phases that naturally describes the data science life cycle.



CRISP-DM methodology mainly focuses on data mining and it is most suitable for industrial projects. It has a huge support by industrial sector but lacks to address machine learning specific tasks such as quality assurance, system behaviour. The system behaviour would be derived from training data, so this leads to a black-box (as the system is not having any human intervention).In light of this situation, the Japanese industry has jointly worked on a set of guidelines for the quality assurance of AI systems. As there is no specific model for ML applications, organizations are using other models which are closely related to their needs such as CRISP-DM. So, the “CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology” proposes a process model for the development of machine learning applications, which guide throughout the life cycle of a machine learning application to meet business expectations.

Shortcomings of CRISP-DM

First, CRISP-DM focuses on data mining and does not cover the application scenario of ML models inferring real-time decisions over a long period of time.

Second, CRISP-DM lacks guidance on quality assurance methodology.

Other Methodologies

Few research teams have come up with certain methodologies which address the short comings of CRISP-DM. One of them is Microsoft research team, that have come up with a process model with nine different phases and address the challenges in ML projects. However, their process model lacks quality assurance methodology and does not cover the business needs.

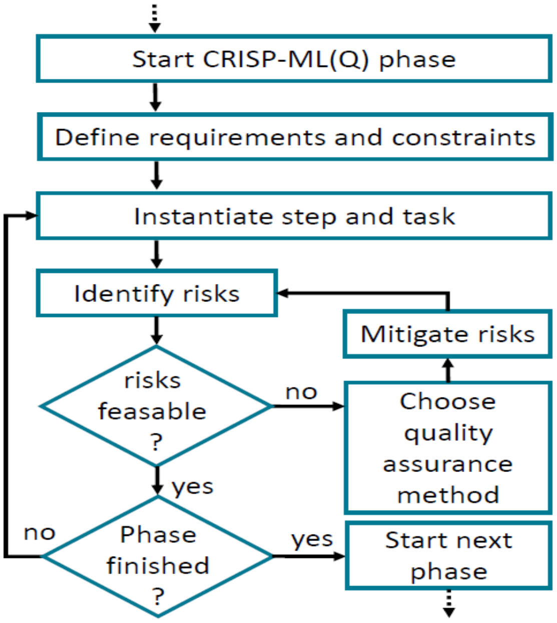
Similarly, Breck et al. proposed specific tests to quantify issues in the ML pipeline and these tests estimate the production readiness of a ML application, i.e., the quality of the application. However, their tests do not completely cover all project phases, e.g., excluding the business understanding activity.

Proposed methodology

The proposed methodology is CRoss-Industry Standard Process model for the development of Machine Learning applications with Quality assurance methodology (CRISP-ML(Q)).It follows the principles of CRISP-DM, but is modified to meet the requirements of ML applications and proposes quality assurance methodology. It's focus primarily is on the technical tasks needed to produce evidence that every step in the development process is of sufficient quality to warrant the adoption into business processes.

Quality Assurance in Machine Learning Projects

The proposed model has six phases and quality assurance methodology is introduced in each phase. The flow chart shows the instantiation of one specific tasks in a development phase, and the dedicated steps to identify and mitigate risks.



The six phases in CRISP-ML(Q) are

1. ***Business & Data Understanding***: In this phase we define the business objectives and translate it to ML objectives, along with this we also collect data and verify the quality. Other important steps to take care while working in this phase are –
    - Define scope of the application, which clearly states the functions and features of product that is being developed
    - Define Success Criteria considering the three aspects – Business (defining specific failure rate), ML (define an acceptable level of performance to meet the business goals), Economic (keep track of Key Performance Indicator (KPI))
    - Requirements on the application
    - Version control on the data
    - Data Quality Verification
2. ***Data Preparation***: This phase includes
    - Feature selection - select features which contribute most to the prediction variable
    - Data selection – removing certain samples which do not satisfy the quality check
    - Clean data – noise reduction and data imputation
    - Feature engineering – extract the features that improve the performance
    - Data augmentation – create new data using transformation techniques
    - Standardize Data – use generic standard file format and normalize the features
3. ***Modeling***: The goal of the modeling phase is to craft one or multiple models that satisfy the given constraints and requirements. The following steps are followed in this phase
    - Define quality measures of the model - Performance, Robustness, Scalability, Explainability, Model Complexity, Resource Demand are the quality measures.
    - Model Selection – depends on business objectives, the data and the boundary conditions of the project the ML application
    - Model training
    - Ensemble methods – these techniques that create multiple models and then combine them to produce improved results
    - Reproducibility - make changes to the experiment to reproduce data, still with the aim of achieving the same results
4. ***Evaluation***: This phase includes
    - Validate performance
    - Determine robustness – make sure it clears all the quality checks
    - Compare results with defined success criteria
5. ***Deployment***: The deployment phase of a ML model is characterized by its practical use in the designated field of application. Check if following conditions are met while deploying
    - Assure user acceptance and usability
    - Model evaluation under production condition
    - Minimize the risks of unforeseen errors

    - Make a deployment strategy to reduce the risk of undetected errors during the process

6. ***Monitoring & Maintenance***: As the ML models are used over long period of time and they have a life cycle, it is important to maintain them in order to avoid the risk of performance degradation over a period of time. The following criteria are to be checked while monitoring and maintaining the deployed model
    - Degradation of hardware
    - System updates
    - Non-stationary data distribution – as data distributions change over time and this can result in a stale training set.
    - While monitoring we need to ensure whether the model is maintaining the predetermined desired level of performance

Conclusion

CRISP-ML(Q) provides quality assurance methodology which helps the organizations to increase efficiency and the success rate in their ML projects. This methodology provides maintenance & monitoring of deployed models and quality-oriented methods to mitigate the risks.