2023

Towards an Agile CRISP-DM Process

Natalie Gerhart

Russell Torres

Laurie Giddens

Towards an Agile CRISP-DM Process

Natalie Gerhart
Creighton University
NatalieGerhart@Creighton.edu

Russell Torres
University of North Texas
Russell.Torres@unt.edu

Laurie Giddens
University of North Texas
Laurie.Giddens@unt.edu

ABSTRACT

Data mining is an increasingly important capability for businesses to remain competitive. For decades, analysts have implemented the data mining process using the CRoss Industry Standard Process for Data Mining (CRISP-DM) methodology. The six phases of the CRISP-DM process (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment) allow for an organized process to move modeling forward. However, as analytics becomes more prevalent, teams of analysts are growing and changing, resulting in changes in the process. In some instances job functions are narrowing in on certain phases of the process. In other instances, analyst teams are requiring more input from business stakeholders due to an overwhelming amount of data that needs to be narrowed down for business reasons, not just statistical reasons. As a result, the communication in the CRISP-DM process, and the understanding between additional parties, is more necessary than ever. Despite this, iteration of the process is increasingly important as the available data grows. In this research in progress, we propose the CRISP-DM process should be updated to follow an Agile-inspired methodology focusing more on the parties and interactions between those parties. Specifically, employing short sprints throughout the CRISP-DM process could engage stakeholders more and result in better outcomes. In addition, beginning the process with expectations for change could also improve the modeling process and reduce frustration from both analysts and stakeholders when there is an inevitable need to change.

Keywords

Data Mining, CRISP-DM, Agile, Analytics.