Data Science: CRISP-DM for Data Science: Strengths, Weaknesses, and Potential

**Yog Chaudhary**

ADTA 5130 – Data Analytics 1

the University of North Texas

Dr. Philpot Denise

July 14, 2023

Abstract

The Cross-Industry Standard Process for Data Mining (CRISP-DM) in the field of data science. It explores the flexibility, adaptability, and nature of CRISP-DM, along with its potential to address emerging trends like explainable AI and ethical considerations. The article acknowledges the limitations of CRISP-DM, including the lack of specific guidance on techniques and integration challenges. It emphasizes the value of CRISP-DM as a structured framework for “data”, science projects, while also recognizing the need for continuous improvement and adaptation to meet evolving industry needs.

Data science has become a pivotal discipline in today's data-driven world, necessitating structured methodologies for managing data-centric projects effectively. The Cross-Industry Standard Process for Data Mining (CRISP-DM) is widely recognized and utilized as a framework in the field of data science. This journal article aims to comprehensively explore the strengths, weaknesses, and the potential of CRISP-DM, providing insights into its applicability and limitations.

It is well-structured, with a clear introduction, progression of ideas, and coherent sections. The language used is appropriate for a scholarly article in the field of “data”, science, and writing style is concise and informative. Headings are effectively used to organize the content and guide the reader through the analysis of CRISP-DM.

CRISP-DM. The strengths of CRISP-DM, such as its flexibility and adaptability to different project types, are discussed in detail. The iterative nature of the framework, enabling continuous improvement throughout the project lifecycle, is highlighted as another key strength. The weaknesses of CRISP-DM, such as the lack of specific guidance on “data”, science techniques and challenges in integrating with other methodologies, are acknowledged, and examined.

CRISP-DM is a data analysis method that improves the quality of machine learning models through data reduction. CRISP-DM is a variant of the cutting-edge domain-specific data mining algorithm CRISP, which is used to discover patterns in large data sets. CRISP-DM works by first reducing the feature space of the data using a feature reduction transformation, and then using a differentiable algorithm to find the best subset of features for use in the machine learning model. This process is often referred to as feature extraction. Additionally, the CRISP-DM process is a data analysis methodology that can be used to help organizations manage and analyze data. The CRISP-DM methodology is a six-step process that can be used to help organizations manage and analyze data. The six steps of the CRISP-DM process are business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

The process is business understanding. In this step, the organization's business objectives are analyzed and understood. The second step of the CRISP-DM process is data understanding. In this step, the organization's data is analyzed and understood. The third step of the CRISP-DM process is data preparation. In this step, the organization's data is prepared for modeling. The fourth step of the CRISP-DM process is modeling. In this step, the organization's data is modeled. The fifth step of the CRISP-DM process is evaluation. In this step, the organization's data is evaluated. The sixth step of the CRISP-DM process is deployment. In this step, the organization's data is deployed.

CRISP-DM methodology offers significant strengths in the field of data science, including its structured and flexible framework, iterative approach, industry acceptance, and emphasis on business understanding. These strengths have contributed to its widespread adoption and effectiveness in managing data-driven projects. However, CRISP-DM also has weaknesses, such as limited guidance on data collection, communication, and visualization support, and its linear process.

The mechanics of writing are strong, employing a well-structured and concise format. The content demonstrates critical thinking and analysis, evaluating the strengths and weaknesses of CRISP-DM, as well as it’s potential to address emerging trends. The article contributes valuable insights to the existing body of knowledge on “data”, science methodologies, benefiting professionals and researchers in the field. It emphasizes the value of CRISP-DM as a structured framework for data science projects, while acknowledging the need for continuous improvement and adaptation to meet evolving industry needs.

REFERENCES

**1.**JS Saltz and I Shamshurin, "Big data team process methodologies: A literature review and the identification of key factors for a project’s success", 2016 IEEE International Conference on Big Data (Big Data), pp. 2872-9, 2016.

2. C Pete, C Julian, K Randy, K Thomas, R Thomas, S Colin et al., Crisp-Dm 1.0, CRISP-DM Consortium, 2000.

3. B Ahmed, T Dannhauser and N Philip, "A Lean Design Thinking Methodology (LDTM) for Machine Learning and Modern Data Projects", 2018 10th Comput Sci Electron Eng Conf CEEC, pp. 11-4, 2018.

4. G. Piatetsky, "CRISP-DM still the top methodology for analytics data mining or data science projects", kdnuggets.com, 2014, [online] Available: http://www.kdnuggets.com/2014/10/crisp-dm-topmethodology-analytics-data-mining-data-science-projects.html.