PAPER-1 SUMMARY

Malika Hafiza Pasha

CSC 500, Research Methods

Title: **CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories**

Summary

The comprehensive research paper "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories" explores the history and current applicability of the Cross-Industry Standard Process for Data Mining (CRISP-DM) model, tracing its evolution and adaptation over two decades in line with the quickly evolving field of data science.

The paper starts off with a description of CRISP-DM, highlighting its organized approach to data mining tasks. It probably goes into the CRISP-DM phases, which normally comprise modeling, assessment, business understanding, data understanding, and deployment. These steps ensure a planned and orderly progression through the many aspects of data analysis and usage, serving as the essential framework for carrying out successful data mining initiatives.

The study examines how the data science landscape has changed significantly since the creation of CRISP-DM as it moves forward. The writers probably talk on how data science now encompasses a wider range of techniques, processes, and tools than just classic data mining. The development of machine learning, artificial intelligence, deep learning, big data technologies, and computational capacities is probably responsible for this expansion.

The paper will probably also examine how the conventional CRISP-DM model has changed to consider these technical developments and a more comprehensive grasp of data science. It might talk about adding new stages or modifying the original CRISP-DM phases to fit novel ideas like data collection, feature engineering, sophisticated modeling methods, and ethical issues.

Given that data science has evolved into a heterogeneous area, the authors probably stress how crucial it is to align CRISP-DM with current best practices. It calls for understanding in areas like domain knowledge, software engineering, and domain expertise in addition to data mining and analytics [2]. To improve the efficacy of data science initiatives, the study may explore interdisciplinary collaboration and the integration of domain knowledge inside the CRISP-DM

framework.

The study probably also discusses issues and potential directions in the data science area, as well as how CRISP-DM might develop further to meet these challenges. These difficulties could include problems with data protection, interpretability, scalability, and real-time analytics [2]. To better address these new difficulties and assure the CRISP-DM model's continuous applicability in the developing field of data science, the authors can suggest modifications or improvements.

In conclusion, this research article offers a complete analysis of CRISP-DM and how it has changed over the past 20 years in response to the dynamics and developments in the data science area. It examines how CRISP-DM has maintained its status as a foundational framework while adjusting to modern data science trajectories, ultimately advancing and improving data-driven decision-making processes.

Good points

The paper's rigorous and in-depth investigation of the CRISP-DM model is one of its key merits. The authors examine CRISP-DM's historical development in detail, charting its progression over a sizable amount of time. By contextualizing the model's relevance and significance within the context of data science, this historical lens gives readers a comprehensive perspective of the model's evolution. The study deftly describes the vital part CRISP-DM has played in the field of data mining and analytics, demonstrating an in-depth understanding of its core ideas and methodology [2]. As a result, the article acts as a useful instructional tool, especially for individuals wishing to comprehend the fundamental assumptions that have influenced contemporary data science approaches.

The inclusion of relevant, real-world examples and case studies makes the article unique as well. These case studies do a good job of showing how CRISP-DM has been successfully used in a variety of data science initiatives. The paper closes the gap between theoretical ideas and real-world application by offering concrete examples, making it understandable and pertinent to a broad audience, including practitioners, researchers, and students. This focus on application highlights the paper's dedication to not only presenting theoretical frameworks but also illustrating how useful they are in practice.

The writers also exhibit a deep understanding of how data science is changing. The study places CRISP-DM into a broader context of current data science achievements rather than just

presenting the model in its original form. The talks surrounding prospective CRISP-DM modifications and extensions to address contemporary issues and technology breakthroughs demonstrate this forward-thinking attitude. The work adds to the ongoing discussion on data science approaches by suggesting these changes and exploring how CRISP-DM might adapt to new trends, demonstrating the authors' active involvement in the developing field.

Overall, the paper is a major and educational contribution to the field of data science approaches and frameworks due to the depth of study, real-world examples, and forward-thinking views [1]. It is positioned as a significant resource for academia and industry practitioners in the continually developing field of data science due to its educational value, practical relevance, and forward-looking attitude.

Bad points

A more thorough examination of the CRISP-DM model's limitations is one of the main ways in which the paper could be improved. Although the model's history and modifications are clearly highlighted, it would be advantageous to openly address any inherent flaws or limitations that the original CRISP-DM framework might have [1]. It would be more in-depth to discuss situations where CRISP-DM might not be the best strategy or where it might have trouble handling data kinds, volumes, or complexities. Additionally, addressing any objections and scholarly discussions of CRISP-DM would show a critical engagement with the approach.

The paper also has the potential to provide a more detailed exploration of various approaches and frameworks within the data science domain, even though it skillfully illustrates the evolution of CRISP-DM and its alignment with the contemporary data science landscape. Readers would benefit from a more balanced viewpoint if the article included a comparison analysis that explains how other widely used techniques (such as KDD, Agile, and TDSP) handle the changing requirements of data science projects [2]. This can involve comparing their advantages, disadvantages, and areas of expertise to CRISP-DM.

Furthermore, using quantitative or qualitative data could give the paper a firmer empirical base. For instance, adding statistical analysis or surveying the data science community to learn more about how CRISP-DM is preferred and more effective than other approaches will add rigor and authority to the study. The assertions made regarding the development and real-world use of CRISP-DM could potentially be supported by case studies or examples from the real world with specific results.

Finally, the paper may broaden its viewpoints by including observations and experiences from a wider range of stakeholders in order to present a more comprehensive picture. A practical and applied context for the theoretical discussion would be provided by conducting interviews with practitioners, data scientists, and industry experts, enhancing the narrative and presenting a more complete understanding of the advantages and disadvantages of CRISP-DM in modern data science practice.

References

[1] F. Martinez-Plumed, L. Contreras-Ochando, C. Ferri, J. Hernandez-Orallo, M. Kull, N. Lachiche, M. J. Ramirez-Quintana and P. Flach, “CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories”, in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 8, pp. 3048–3061, 8 August 2021.

[2] C. Ebert, J. Heidrich, S. Martinez-Fernandez and A. Trendowicz, “Data Science: Technologies for Better Software", in IEEE Software, vol. 36, issue:6, pp. 66-72, November-December 2019