Leveraging Data Analysis in Agile Software Development: Applications, Barriers, and Opportunities

José Cassola Bacallao[0000−0002−1029−1548], Brian Keith Norambuena[0000−0001−5734−8962], and Ítalo Donoso Barraza[0009−0006−0468−9919]

Abstract Data Analysis (DA) integration into Agile Software Development (ASD) has generated substantial research, yet this work remains scattered across multiple studies. This exploratory review consolidates the literature to provide an integrated overview. DA applications in ASD span performance prediction for project management, objective insights for decision-making processes, and continuous quality assessment with fault prediction. Big Data, AI, and ontologies have emerged as critical enablers for productivity gains and data integration. Adoption barriers persist: scarce relevant data, information overload, limited specialized tools, and absence of formal methodological frameworks. Although DA could significantly improve ASD efficiency and quality, realizing these benefits requires further research to develop and empirically validate frameworks.

1 Introduction

Agile software development (ASD) has revolutionized the industry by prioritizing adaptability, collaboration, and rapid value delivery, enabling teams to respond effectively to the growing complexity of projects and changing market demands. In this dynamic environment, data analysis has become an indispensable tool for driving informed decision-making and optimizing the development process [17].

Data collection and analysis throughout the software development lifecycle allows agile teams to track project progress, product quality, and user behavior [5]. Teams use

José Cassola Bacallao
Universidad Católica del Norte, Av. Angamos 0610, e-mail: jose.cassola@alumnos.ucn.cl,

Brian Keith Norambuena
Universidad Católica del Norte, Av. Angamos 0610, e-mail: brian.keith@ucn.cl,

Ítalo Donoso Barraza
Universidad Católica del Norte, Av. Angamos 0610, e-mail: italo.donoso@ucn.cl *Corresponding Author*

this information to identify improvement areas, predict risks, and base decisions on evidence rather than intuition [3, 6].

The adoption of *Data-driven decision-making* (DDDM) approaches in agile software development remains limited. Empirical studies document a gap between recognizing DDDM's value and implementing it in practice, primarily due to scarce relevant data, interpretation challenges, and inadequate tooling [10, 5]. Current research anticipates increased DDDM integration into strategic decision-making, requirements management, and process improvement [27].

This paper aims to examine the role of data analysis in optimizing agile project development in software engineering. To achieve this objective, we pose the following research question: How can data analysis be used to improve agile project development in software engineering? Through a detailed analysis of ten studies in the field, we will explore the various applications of data analysis in agile projects, the benefits it provides, and the challenges that persist.

This paper is organized as follows: first, the fundamental role of data analysis in optimizing agile project development is introduced. Then, a review of the identified application areas in the literature is presented. Next, a synthesis of the literature is provided, highlighting common themes, trends, challenges, and opportunities. Following this, the implications of the findings for both practice and research are discussed. Finally, the main conclusions of this study are summarized.

2 Key Applications of Data Analysis in Agile Development

Distinct key application areas have been identified in which data analysis is transforming agile project development.

2.1 Frameworks for Data Science Project Management

Data science project management presents unique challenges due to the inherent uncertainty in requirements, project success, time estimation, and the quality and relevance of data for the desired predictive model [23]. Traditional approaches like Scrum, Kanban, and CRISP-DM can be inadequate for these challenges, as they were not designed to handle the iterative and uncertain nature of data science projects [23, 1]. These approaches may face difficulties in task estimation, process definition, and integration with other techniques [23].

In [23], the authors propose a new framework called Data Driven Scrum (DDS) to improve the management of data science projects. DDS integrates concepts from Scrum and Kanban but is adapted to the specific challenges of data science by following the four Lean principles for data science. Instead of fixed-time iterations, DDS focuses on completing logical work capabilities, which allows for greater flexibility and adaptability to project uncertainty. Furthermore, meetings in DDS are decoupled from iterations and are held based on logical time intervals, facilitating team communication and coordination. DDS also uses high-level effort estimates to prioritize future iterations, rather than detailed estimates that can be difficult to perform in data science projects.

2.2 Integrating Scrum in Data Science Projects

Building on these frameworks, other approaches seek to adapt Scrum directly into data science contexts. Baijens et al. [2] developed Scrum-DS, which integrates Scrum events, artifacts, and roles into the CRISP-DM data science process stages. Sprint-based work breakdown with regular reviews and retrospectives provides early and continuous feedback, enables rapid problem identification, and supports adaptation to requirement or data changes. The methodology introduces "Sprint Zero" for data preparation, addressing this challenge within a structured timeframe. Expert interviews across three organizations evaluated and refined Scrum-DS, confirming its compatibility with data science projects and its capacity to improve project execution.

2.3 Data Analysis for Decision-Making in Scrum

Beyond project management, data analysis and machine learning enhance decision-making in Scrum, particularly during sprint retrospectives. Sandoval et al. [24] propose supervised machine learning models and analytics techniques to analyze sprint and issue data. Pattern and trend identification in this data helps teams recognize improvement areas and make evidence-based decisions for process optimization [19, 12].

The SEMMA (Sample-Explore-Modify-Model-Assess) methodology structures the machine learning model development. The classification and regression models demonstrate data analysis capabilities for Scrum decision-making. A k-neighbors classification model predicts issue types with 88% accuracy, supporting precise issue categorization and reducing development delays. The multiple linear regression model predicts sprint velocity, enabling accurate estimates and improved sprint planning.

2.4 Big Data and Agile Software Development

At a broader scale, Big Data Analytics (BDA) offers tools for improving ASD [4]. A systematic mapping study examines BDA applications across the agile software development lifecycle [5]. BDA is employed throughout the lifecycle, including code repository analysis, defect correction, testing, project management analysis, and application usage analysis. Improving software development team productivity represents a primary BDA objective in industry [9].

2.5 Continuous Assessment and Improvement of Software Quality

Software quality remains central to ASD, and analytics tools integrating quality models (QM) support continuous evaluation and improvement. Martinez et al. [16] present a case study of four companies using Q-Rapids, a software analytics tool, to evaluate and improve quality in ASD projects. Quantitative and qualitative results indicate professionals view QM integration in analytics tools favorably and find the tools useful. However, adoption challenges emerged, including company-specific adaptation needs and integration requirements with existing tools.

2.6 Ontologies and Data Management in Scrum

Another challenge arises from tool diversity and multiple information sources. Ontologies like the Scrum Reference Ontology (SRO) [25] establish a common framework for data conceptualization and exchange. Common semantic data representation facilitates integration and analysis across different sources, improving decision-making and project management. A case study demonstrated that SRO-based integrated solutions improved estimates, team allocation, and project performance management.

2.7 Intelligent Software Engineering and Agile Development

As applications become more sophisticated, Intelligent Software Engineering (ISE) applies techniques including machine learning and Bayesian networks to improve software development aspects. Perkusich et al. [20] present a systematic literature review on intelligent technique applications in ASD. Reasoning under uncertainty, search-based solutions, and machine learning emerge as the most utilized techniques. These apply to effort estimation, requirements prioritization, and resource allocation, primarily supporting decision-making [18]. The study identifies a trend toward explainable intelligent techniques for increased transparency and trust.

2.8 Failure Prediction in Agile Development

Predictive approaches extend these applications to risk management. Failure prediction in agile development ensures product quality and process efficiency. Batarseh et al. [3] present Analytics-Driven Testing (ADT), which uses data analysis to predict failures in upcoming sprints. ADT employs the Mean Time Between Failures (MTBF) concept. Through statistical analysis and regression models, ADT predicts failure locations and timing, allowing teams to implement preventive measures. Experimental validation showed ADT predicted failures with 93.5% statistical accuracy.

2.9 Data-Driven Decision-Making and Data Availability

Finally, adoption of Data-Driven Decision-Making remains a cross-cutting theme. Svensson et al. [27] conducted an empirical study investigating how software professionals view and apply DDDM. While professionals recognize DDDM capabilities, current adoption remains limited. Primary barriers include data availability, insufficient knowledge about data utilization, and inadequate tooling. Despite limited current adoption, professionals express positive expectations for future DDDM use, particularly in strategic decision-making and requirements management.

Related to this, ASD generates a large amount of data that, if properly collected and analyzed, can provide valuable information to inform and improve decision-making processes. The different types of data available are explored in [17]. The authors identify three main categories: business data, software project data, and ”reality mining” data, which is obtained from observing human behavior. The authors emphasize the importance of combining data analysis with human interpretation to make more informed decisions in agile software development.

3 Review Synthesis

The literature review reveals several common themes, trends, challenges, and opportunities regarding the application of data analysis to improve agile project development in software engineering.

Common Themes. Traditional project management approaches may be unsuitable for data science projects due to their iterative and uncertain nature, necessitating new frameworks [23, 2]. In general, there is a consensus on the potential applications of data analysis to improve decision-making in agile software development, from sprint planning to requirements management and software quality assessment [3, 24, 2, 16].

Another recurring theme is the integration of heterogeneous data sources, which allows organizations to obtain a holistic overview of the project and make informed decisions [5, 25]. Artificial intelligence and machine learning techniques, such as Bayesian networks, clustering algorithms, and regression models, are increasingly being used to analyze data and generate useful insights for decision-making processes [3, 20].

Trends. There is a growing trend toward adopting agile methodologies like Scrum in data science projects to improve their management and outcomes [24, 2]. Software analytics tools and methods that integrate quality models are gaining popularity for continuously assessing and improving software quality in agile projects [3, 16]. The application of intelligent techniques and analytics in agile software development is increasing, with a focus on data-driven decision-making processes and task automation [5, 24, 20, 13, 21].

Challenges. Although promising, the widespread adoption of DDDM in agile software development is hampered by several key barriers. Challenges often begin with the data itself, from a scarcity of relevant information to an overabundance that causes information overload [27]. On a technical and organizational level, integrating data from multiple sources and tools poses a challenge [25]. Furthermore, existing tools and methodologies are not always universally applicable, requiring adaptation to meet the specific needs of projects and companies [16, 26, 8]. Moreover, the deficiencies in specialized knowledge required to leverage these tools effectively and the need for greater transparency. The black-box nature of some artificial intelligence techniques, for instance, can forestall adoption if their outputs are not easily explainable [20, 15].

Opportunities. Developing new tools and methodologies could facilitate the collection, analysis and visualization of relevant data for decision-making processes in the agile context [27]. Training software professionals in DDDM and the use of data analysis tools can improve the adoption and impact of these practices [27]. Explainable intelligent techniques reduce adoption barriers in agile software development by addressing transparency concerns [20]. Lifecycle-wide DDDM integration affects project efficiency and quality metrics [17].

4 Discussion

Agile Frameworks for Data Science Projects. The analysis of the literature reveals a wide spectrum of possibilities for how data analysis can enhance agile project development in software engineering. To begin, the adoption of agile frameworks adapted for data science, such as DDS, can address the specific challenges of these projects, including uncertainty in requirements and time estimation. Using capability-based iterations and high-level estimations, DDS offers greater flexibility and adaptability than traditional approaches such as Scrum or Kanban [14].

Data-Driven Decision Making in Agile Lifecycle. Data analysis can empower decision-making processes at different stages of the agile development lifecycle. For instance, during the sprint retrospective phase in Scrum, data analytics and machine learning techniques can help identify areas for improvement and predict future team performance. During sprint planning, historical data analysis can improve effort estimation and resource allocation [7]. Big Data analytics can also be applied across all stages of agile development to obtain insight into code quality, team performance, and user behavior, enabling more informed decisions and process optimization.

Quality Assurance Through Software Analysis Tools. In addition, integrating software analysis tools with quality models can facilitate continuous evaluation and improvement of software quality in agile projects. These tools, such as Q-Rapids, provide valuable information on product quality and the development process, allowing teams to identify and address quality issues early and proactively.

Intelligent Software Engineering Approaches. Intelligent software engineering offers a promising approach to enhance agile development by applying intelligent techniques like machine learning and Bayesian networks. These techniques can automate tasks, analyze data, and provide valuable information for decision-making in areas such as effort estimation, requirements prioritization, and resource allocation.

Predictive Analytics for Failure Prevention. Data analysis can be used to predict failures in agile software development. Methods like the Analytics-Driven Testing approach, which utilize statistical analysis and regression models, can help teams anticipate failures and take preventive measures to avoid them or mitigate their impact.

Practical and Research Implications. The results of the literature review reveal significant implications for the use of data analysis in agile projects. They highlight the need to adopt frameworks and methodologies that adapt to the iterative and uncertain nature of agile software development, especially in data science projects.

Similarly, the findings underscore the value of data analysis in enhancing decision-making processes throughout all stages of the development lifecycle, from planning to continuous evaluation and improvement of software quality. This review also emphasizes the importance of integrating data from multiple sources and tools to obtain a holistic view of the project and make more informed decisions. Thus, the results point to the capabilities of artificial intelligence and machine learning techniques to automate tasks, analyze data, and generate actionable insights for decision-making processes. Finally, a summary of the analysis can be observed in Table 1.

Table 1: Summary of the Analysis in the Selected Publications

Article	**Application Area**	**Applied Techniques/Outcomes**	**Benefits**
Saltz et al. (2022) [23]	Data science project management.	Iterations based on capabilities, decoupled meetings, high-level estimations.	Greater flexibility and adaptability to the uncertainty of data science projects, improved team communication and coordination, effective prioritization of iterations.
Sandoval-Alfaro and Quintero-Meza (2021) [24]	Decision-making in the Scrum retrospective stage.	Supervised machine learning models (regression and classification), data analytics techniques, SEMMA methodology.	Improved task estimation, prediction of team velocity, more informed and evidence-based decision-making in retrospectives.
Baijens et al. (2020) [2]	Execution of data science projects.	Scrum-DS methodology (integration of Scrum and CRISP-DM), Sprint Zero.	Higher success rate in data science projects, early and continuous feedback, rapid identification and resolution of problems, adaptation to changes, better work organization and prioritization, structured approach for data preparation.

Continued on next page

Table 1 – continued from previous page

Name	**Application Area**	**Applied Techniques/Results**	**Benefits**
Biesialska et al. (2021) [5]	ASD (analysis of code repositories, defect correction, testing, project management, application usage analysis).	Predictive, descriptive, and adaptive analysis; machine learning (SVM, Random Forest, Naive Bayes, neural networks).	Identification of improvement areas in project development, process optimization, risk prediction, more informed decision-making, improvement of software quality and development efficiency.
S. Martinez-Fernandez et al. (2019) [16]	Continuous assessment and improvement of software quality in ASD projects.	Software analysis tools with integrated quality models (Q-Rapids).	Valuable information about product and development process quality, early identification of quality problems, data-driven decision-making to improve software quality.
Santos Júnior et al. (2021) [25]	Data management in Scrum projects.	SRO, semantic integration of applications (Azure DevOps and Clockify).	Improved project management, more informed decision-making, automatic data exchange between applications, reduction of manual work and errors, greater focus on development activities.
Perkusich et al. (2020) [20]	ASD (effort estimation, requirements prioritization, resource allocation, and requirements management).	Reasoning under uncertainty (Bayesian networks), software-based search, machine learning.	Task automation, data analysis, decision-making support, improvement of software efficiency and quality.
Batarseh and Gonzalez (2018) [3]	Failure prediction in ASD.	Statistical analysis, MTBF, regression models, ADT .	Failure anticipation, improvement of software quality, reduction of development costs.
Svensson et al. (2019) [27]	Decision-making in ASD.	Questionnaire for software professionals.	Identification of barriers to DDDM adoption (lack of data, lack of knowledge, lack of tools), positive outlook on the future use of DDDM in strategic decision-making and requirements management.

Continued on next page

Table 1 – continued from previous page

Name	**Application Area**	**Applied Techniques/Results**	**Benefits**
Matthies and Hesse (2019) [17]	Decision-making in ASD.	Analysis of business data, software project data, ”reality mining” data.	More informed decision-making at all project levels, identification of problems and failures, improved team communication and coordination, process optimization, improvement of software quality.

5 Conclusions

Data analysis enables agile teams to make decisions based on evidence from multiple sources, optimize processes, and improve product quality. Data-Driven Scrum addresses data science project challenges through adapted frameworks. Sprint retrospective analysis identifies improvement areas and predicts team performance. The Scrum-DS methodology combines Scrum's iterative structure with data science's exploratory nature.

Big Data analytics applies across the agile development lifecycle to enhance productivity. Software analysis tools with integrated quality models support continuous quality evaluation in agile projects. Ontologies like the Scrum Reference Ontology facilitate cross-tool data integration, improving project management decisions.

Intelligent software engineering applies machine learning and Bayesian networks to automate tasks and analyze data for effort estimation, requirements prioritization, and resource allocation. Analytics-driven testing demonstrates failure prediction capabilities in agile development.

Current DDDM adoption in agile software development remains limited by several barriers: data scarcity, information overload, insufficient analytical knowledge, and inadequate tooling. Addressing these requires developing appropriate tools and methodologies, training professionals, and researching explainable intelligent techniques.

6 Future Work

Several areas require further research in data analysis for agile development. DDDM adoption barriers need systematic solutions. Data scarcity, information overload, analytical skill gaps, and tool inadequacies limit practical implementation. Tool development should focus on data collection, analysis, and visualization specific to agile contexts. Professional training in data analysis could improve team decision-making and project outcomes.

Heterogeneous data integration from multiple sources and tools presents ongoing challenges. Solutions for seamless integration and effective visualization could enhance agile project decision-making [22, 28].

Research into explainable intelligent techniques should continue. Understanding system reasoning and decision processes affects professional adoption. Methods that clarify intelligent system decisions and their application to development processes merit investigation [11].

Early failure detection methods could reduce development costs. Research should identify reliable failure-predicting patterns and metrics, developing predictive models for integration into agile tools and processes.

Acknowledgment

Brian Keith is supported by Project 202311010033-VRIDT-UCN.

References

1. S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, and T. Zimmermann, "Software Engineering for Machine Learning: A Case Study," *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pp. 291–300, 2019.
2. J. Baijens, R. Helms, and D. Iren, "Applying Scrum in Data Science Projects," *Proceedings - 2020 IEEE 22nd Conference on Business Informatics, CBI 2020*, vol. 1, pp. 30–38, 2020.
3. F. A. Batarseh and A. J. Gonzalez, "Predicting failures in agile software development through data analytics," *Software Quality Journal*, vol. 26, no. 1, pp. 49–66, 2018.
4. T. R. Batista, M. Bronzo, and M. S. Barbosa, "The effects of big data analytics capabilities on agile software development practices and project performance: a dynamic capabilities view perspective," *International Journal of Managing Projects in Business*, pp. 112–137, 2025.
5. K. Biesialska, X. Franch, and V. Muntés-Mulero, "Big Data analytics in Agile software development: A systematic mapping study," *Information and Software Technology*, vol. 132, p. 106448, 2021.
6. H. K. Dam, T. Tran, T. T. M. Pham, S. W. Ng, J. Grundy, and A. Ghose, "Towards effective AI-powered agile project management," *Proceedings of the 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*, pp. 41–44, 2019.
7. L.-D. Radu and D. Popescul, "Data Analytics in Agile Project Managements," *Proceedings of the 2024 7th International Conference on Information Science and Systems*, ACM, pp. 7–12, 2024.
8. T. Detofeno, S. S. Reinehr, and A. Malucelli, "Technical Debt Guild: When experience and engagement improve Technical Debt Management," *Proceedings of the XX Brazilian Symposium on Software Quality*, ACM, Article 25, pp. 1–10, 2021.
9. J. Iqbal, M. Omar, and A. Yasin, "An Empirical Analysis of the Effect of Agile Teams on Software Productivity," *Proceedings of the 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, IEEE, pp. 1–8, 2019.
10. A. Fawzy, A. Tahir, M. Galster, and P. Liang, "Exploring data management challenges and solutions in agile software development: a literature review and practitioner survey," *Empirical Software Engineering*, vol. 30, no. 1, article 77, 2025.
11. A. Fellah, "Integrating AI into Scrum: Enhancing and Refining Agile Software Development Practices," *The 2025 OkIP International Conference on Artificial Intelligence Frontiers*, Oklahoma City, USA, 2025.

12. R. Hoda, H. Dam, C. Tantithamthavorn, P. Thongtanunam, and M. A. Storey, "Augmented Agile: Human-Centered AI-Assisted Software Management," *IEEE Software*, vol. 40, no. 4, pp. 106–109, 2023.
13. M. M. John, H. H. Olsson, and J. Bosch, "Towards MLOps: A Framework and Maturity Model," *Proceedings of the 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 1–8, 2021.
14. E. Kula, A. van Deursen, and G. Gousios, "Context-Aware Automated Sprint Plan Generation for Agile Software Development," *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, pp. 1745–1756, 2024.
15. M. Laiq, N. bin Ali, J. Börstler, and E. Engström, "Software analytics for software engineering: A tertiary review," *arXiv preprint arXiv:2410.05796*, 2024.
16. S. Martinez-Fernandez et al., "Continuously Assessing and Improving Software Quality with Software Analytics Tools: A Case Study," *IEEE Access*, vol. 7, pp. 68219–68239, 2019.
17. C. Matthies and G. Hesse, "Towards using Data to Inform Decisions in Agile Software Development: Views of Available Data," *Proceedings of the 14th International Conference on Software Technologies*, SCITEPRESS - Science and Technology Publications, pp. 552–559, 2019.
18. S. Wang, L. Huang, A. Gao, J. Ge, T. Zhang, H. Feng, I. Satyarth, M. Li, H. Zhang, and V. Ng, "Machine/Deep Learning for Software Engineering: A Systematic Literature Review," *IEEE Transactions on Software Engineering*, vol. 49, no. 3, pp. 1188–1231, 2023.
19. D. E. Strode, T. Dingsøyr, and Y. Lindsjørn, "A teamwork effectiveness model for agile software development," *Empirical Software Engineering*, vol. 27, no. 2, article 56, 2022.
20. M. Perkusich, L. C. e Silva, A. Costa, F. Ramos, R. Saraiva, A. Freire, E. Dilorenzo, E. Dantas, D. Santos, K. Gorgônio, H. Almeida, and A. Perkusich, "Intelligent software engineering in the context of agile software development: A systematic literature review," *Information and Software Technology*, vol. 119, p. 106241, 2020.
21. S. V. Subramanya, S. Sierla, and V. Vyatkin, "From DevOps to MLOps: Overview and Application to Electricity Market Forecasting," *Applied Sciences*, vol. 12, no. 19, p. 9851, 2022.
22. T. Rique, E. Dantas, M. Perkusich, K. Gorgônio, H. Almeida, and A. Perkusich, "Use Cases for Software Development Analytics: A Case Study," *Proceedings of the XXXVI Brazilian Symposium on Software Engineering*, ACM, pp. 388–397, 2022.
23. J. Saltz, A. Sutherland, and N. Hotz, "Achieving Lean Data Science Agility Via Data Driven Scrum," *Proceedings of the 55th Hawaii International Conference on System Sciences*, pp. 7472–7481, 2022.
24. O. E. Sandoval-Alfaro and R. R. Quintero-Meza, "Application of data analytics techniques for decision making in the retrospective stage of the agile scrum methodology," *2021 Mexican International Conference on Computer Science (ENC 2021)*, IEEE, pp. 1–8, 2021.
25. P. S. Santos Júnior, M. P. Barcellos, R. de A. Falbo, and J. P. A. Almeida, "From a Scrum Reference Ontology to the Integration of Applications for Data-Driven Software Development," *Information and Software Technology*, vol. 136, p. 106565, 2021.
26. D. Ståhl and J. Bosch, "Modeling continuous integration practice differences in industry software development," *Journal of Systems and Software*, vol. 87, pp. 48–59, 2014.
27. R. B. Svensson, R. Feldt, and R. Torkar, "The Unfulfilled Potential of Data-Driven Decision Making in Agile Software Development," in *Agile Processes in Software Engineering and Extreme Programming*, P. Kruchten, S. Fraser, and F. Coallier, Eds., Springer, Cham, pp. 69–85, 2019.
28. C. Uwasomba, A. Deshpande, H. Sharp, P. Gregory, R. Willis, L. Barroca, M. Uwadi, and K. Taylor, "Data-Driven Agility: Assessing Agile Culture transformation in a technology organisation," *Information and Software Technology*, vol. 177, p. 107543, 2025.