Application of Data Science within the Army Intelligence Warfighting Function: Problem Summary and Key Findings

Robert Ganger
CUBRC
Buffalo, NY, USA
ganger@cubrc.org

John Coles
CUBRC
Buffalo, NY, USA
john.coles@cubrc.org

Jacob Ekstrum
CUBRC
Buffalo, NY, USA
jacob.ekstrum@cubrc.org

Timothy Hanratty
US Army CCDC ARL
APG, MD, USA
timothy.p.hanratty.civ@mail.mil

Eric Heilman
US Army CCDC ARL
APG, MD, USA
eric.g.heilman.civ@mail.mil

Jason Boslaugh
US Army ICoE
Ft. Huachuca, AZ
jason.l.boslaugh.mil@mail.mil

Zachary Kendrick
US Army ICoE
Ft. Huachuca, AZ
zachary.d.kendrick.ctr@mail.mil

Abstract— **Army Intelligence operates in a data rich environment with limited ability to operationalize exponentially increasing volumes of disparate structured and unstructured data to deliver timely, accurate, relevant, and tailored intelligence in support of mission command at echelon. The volume, velocity, variety, and veracity (the 4 Vs) of data challenge existing Army intelligence systems and processes, degrading the efficacy of the Intelligence Warfighting Function (IWfF). At the same time, industry has exploited the recent growth in data science technology to address the challenge of the 4 Vs and bring relevant data-driven insights to business leaders. To bring together the lessons from industry and the data science community, the US Army Research Laboratory (ARL) has collaborated with the US Army Intelligence Center of Excellence (USAICoE) to research these Military Intelligence (MI) challenges in an Army AR 5-5 Study entitled, "Application of Data Science within the Army Intelligence Warfighting Function." This paper summarizes the problem statement, research performed, key findings, and way forward for MI to effectively employ data science and data scientists to reduce the burden on Army Intelligence Analysts and increase the effectiveness of data exploitation to maintain a competitive edge over our adversaries.**

Keywords—Data Science, Army Application, Intelligence Implementation

I. Introduction

Data and information are currently proliferating our world. The Internet of Things (IoT), mobile phones, smart automobiles, and internet-connected houses have vastly increased the amount of data produced and available for collection by systems. The data enriches our world by enabling services that can assist our daily lives through prediction, recommendations, and retrospective analysis that help us better use our resources in the future. However, more data also creates more challenges for processing that data in meaningful ways. In general, there is an increasing gap between amounts of data available over time and the growth of tools and capabilities to process that data. This same disparity holds true in the battlespace, where devices, machines, and sensors are producing a greatly increased volume of data in the modern era.

The purpose in performing the AR 5-5 Study was to examine the practices employed in Military Intelligence (MI), determine how current data science practice fits into military intelligence processing for enhanced data understanding, and provide a path that will enable MI analysts to use data science practices to achieve overmatch against our adversaries. The study focused on a dual military and industry overview of the large data processing problems and recommended ways forward in adopting current and projected data science practice within the MI organization along multiple lines of effort (LOEs). The study was not intended to develop an implementation manual, but rather to suggest a general strategy that can be tailored by military experts into an initial Concept of Employment (CoE) for integration of data science into the Intelligence Warfighting Function (IWfF).

II. Industrial Approach

Academia and industry are constantly iterating on different techniques and practices to find what may help a data science team be most effective. Our work focused on CRoss InduStry Process for Data Mining (CRISP-DM) [1] used by industry as a baseline practice for executing data science analyses. Also considered were other industry practices such as The IBM

978-1-7281-5584-5/19/$31.00 ©2019 IEEE
DOI 10.1109/CSCI49370.2019.00039

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 08:32:18 UTC from IEEE Xplore. Restrictions apply.

Foundational Methodology for Data Science [2] and the Microsoft Team Data Science Process [3] that provide valid examples of industry practices. We acknowledge that these techniques offer refinements that may resonate with some teams. However, these mirror CRISP-DM in the essential tasks necessary to focus efforts of a team towards producing improved results. The steps of CRISP-DM are: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment all surrounding a core set of data. These steps are not just linear, but include feedback loops to previous steps as any given round of data analysis may illuminate additional data characteristics supportive of an iterative analyses.

Applying the CRISP-DM method in industry includes the processes that govern how a team functions to assign and complete work. The current business practice of promoting an agile way of thinking is revolutionizing business practice. A finding for the MI is that the data science team's organization and interaction with outside stakeholders should reflect industry-standard agile processes as these maintain collaboration, communication, and the delivery of incremental value. At the heart of the Agile system is a "Scrum" Framework [4] that prescribes methods for effective team collaboration on complex projects. The team organizes its work around "sprints," which define the team's tasking over a set period of time; and "daily scrums," where the trajectory of the work is evaluated for blockers and course-corrected as necessary to reach the sprint goals. The outcome of a sprint is an incremental product for internal evaluation or perhaps for external release as a usable data science-based model.

Data science as practiced today in industry contains numerous lessons and challenges for current and future implementation in MI processes. Implications include adaptation of business standard practices, incorporation of agile methods, and creation of a readily available data science tool set. An examination of these aspects of data science will provide a model for MI team formation and empowerment resulting in improved operation within the military environment. A significant challenge is that industry data science teams function using processes that are vastly different than those used traditional military environments. A study hypothesis is that a large number of organizational and cultural changes will be of concern when adapting CRISP-DM team formation and operations to the military environment. Our work acknowledges and describes these challenges and offers a strategy for moving forward with MI data science teams. Existing MI organizations will have the task of designing specific processes to enable their data science teams to perform in an effective, agile way while still adhering to the inherent hierarchical structure of the US Army.

III. Data science to enhance analytic rigor in military intellligence analysis

An intelligence estimate is a conjecture and should be discarded if not compatible with a known (verified) fact [5]. But even if the conjecture has no evidence to discard it, it is still an estimate with a level of confidence (which is in reality an unknown quantity). As stated in the article cited above, no degree of research can "make an analyst a prophet." However, analytic rigor is a practice that *can* increase the confidence in an estimate to a higher level and also provide improved intelligence products. The Eight Practices/24 Standards of Analytic Rigor [6], implemented in current MI, summarizes MI analysis methodology. These describe how the application of low, moderate, and high rigor is recognized in the activities of an MI analyst. Comparisons of industrial and MI analytic techniques suggests that the integration of industrial data science applications within the IWfF will enhance analytic rigor [10].

As MI data science teams form to learn and practice their craft, aspiring to high levels of analytic rigor will be an important aspect of their vision and goals. By applying industrial data science principles, MI teams will produce highly effective and relevant results (at least to the highest level possible given the intelligence questions that they seek to answer).

IV. Developing a data team

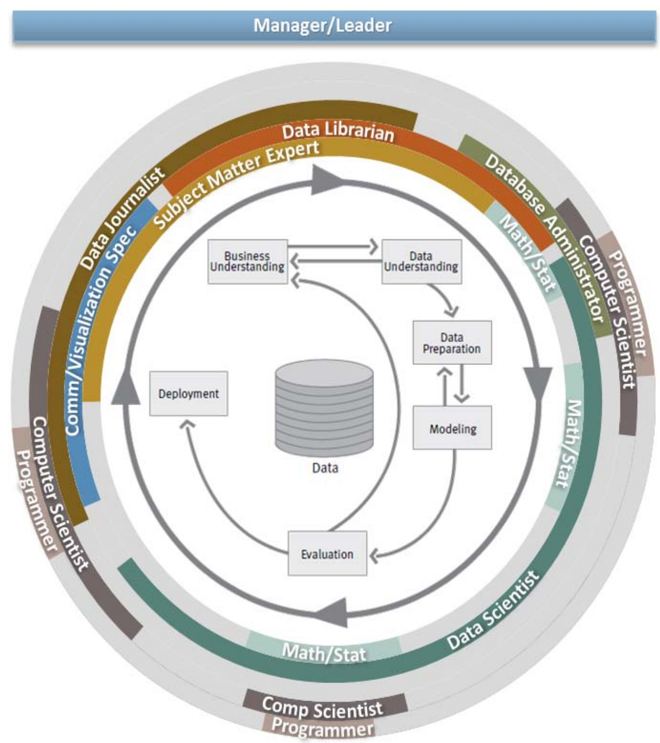

Figure 1Roles of the data science Team within CRISP-DM Steps [10]

The key resource for effective data science in an organization is people. The team composition and inherent skill sets are crucial to successful data science, even more so than the tools employed. To frame the importance of skill diversity within the team, we provide a mapping of the CRISP-DM steps, tasks and requirements to skill sets shown in Figure 1. There is also an implied step not formally shown in CRISP-DM, but never the less an important part of the practice: *Feedback from Stakeholders*. Model deployment is optimal when implemented in an iterative fashion with feedback from stakeholders such that

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 08:32:18 UTC from IEEE Xplore. Restrictions apply.

model efficacy is constantly being evaluated and models will be in a mode of continuous improvement.

The function of the data science team is highly collaborative in nature, and roles have overlapping task requirements within the CRISP-DM process. It is unlikely that a single person can effectively perform all of the steps of the process. We recommend that MI data science teams contain sufficient skill sets to address the requirements shown in Figure 1. If it is not possible to cover all requirements of a team with available staff, then team's membership should provide coverage for as many skills as possible and that training be applied to widen the team's experience in missing skills. A team size of 4-12 members seems optimal; fulfilling a minimum of skill diversity, while keeping the maximum size below a level where management efficiency begins to be lost.

The steps of CRISP-DM are best performed by a diverse talented team who collaborate to address each step using their individual talent contributions and engage in iteration of steps to refine and enhance the analysis as needed. Our study identifies the *industry-titled* roles of Managers/Leaders, Subject Matter Experts, Data Librarians, Database Administrators, Computer Scientists, Programmers, Mathematicians/Statisticians, Data Scientists, Visualization Specialists, and Data Journalists as an optimum set of team members for a CRISP-DM analysis process. In our work, we expand each of these roles and suggest transformations to better align with MI requirements and existing MI skill sets.

While the promise of data science is realized in a wide range of industries, it is important to provide a personnel plan that leverages the unique environment, mission, and current structure of army MI. We envision that the incorporation of formal data science approaches can have a transformative impact on MI practice. Currently, the MI community, realizing the emergence of data science as a positive factor in analysis is implementing the practice in an ad hoc manner. A key point in our study is that the formalization and acceleration of data science practices. Our work highlights the individual contribution of each MI role integrated into the data science process. In parallel, the Army will need to develop a clear data science career path within the current military framework. An industry data science team goes through changes over time as members hone their skills and advance in their career tracks. The Army team members will likewise go through similar transitions, so the Army must formally prepare to manage team role changes and advancement.

V. INCORPORATING AN ADAPTABILITY PROCESS

We suggest that the data science team's overall organization and its interaction with outside stakeholders reflect agile processes in order to maintain collaboration, communication, and the delivery of incremental value. The Agile and the Scrum Frameworks [4] provide effective team collaboration on complex projects. The central organizational activity for the Scrum Framework is a biweekly or weekly sprint. There, the team discusses work prioritization and assignment. The team will be able to react quickly to the changing intelligence environment by reevaluating intelligence questions and current data science analyses during each sprint, and course-correct as needed to stay on track for high-value delivery to the customers.

We recommend that the data science team add all intelligence questions to their product log, manage and prioritize the log with each sprint (or mid-sprint for highest priority questions), and use daily scrums to evaluate the progress of all analyses. The team can then adjust project priorities to reflect critical intelligence requirements. The team should work closely with the users of their data science models, communicating and collaborating with them often to receive feedback on the model performance. This practice will enable the team to generate continuous model improvements.

In the ideal agile structure, tasking occurs at the start of the sprint and stays static throughout the sprint duration. For the applied military approach, we recommend an additional degree of flexibility for the agile data science team to account for immediate "hot topics" that may arise during a sprint, especially during times of conflict. During these times, we speculate that the data science team will need to quickly evaluate new questions in daily or ad hoc scrums to determine if they can be answered within the required timeframes using data science techniques, or if they should be passed to a traditional intelligence team. If the agile team takes on the questions, they will adjust tasking mid-sprint and start new model development.

For handling data science projects that take longer than a single sprint, the team should use an expanded process that includes aspects of SAFe Agile [7]. Particularly, this technique contains "Agile Release Trains" (ARTs) that each represent an effort to produce a single analysis model or product. The timeframe for an ART is defined by the product requirements so that it can exist across multiple sprints. ARTs have a single point of deployment for the customer, followed by maintenance/sustainment, or may have multiple points of deployment if incremental product releases make sense to the intelligence customer.

Figure 2 shows the hybrid Scrum/SAFe process for a continuously operating high-performance team. Strategic and tactical warfighters and decision makers, i.e. the intelligence customers, give questions and requirements to the team. The intelligence questions then become part of the sprint project log for the team from which these gain a prioritized task assignment for a team sprint. When a task becomes active in a sprint, a team, following the CRISP-DM process, creates a data science product. Deployment of a data science product to a customer then begins a process of incremental feedback and upgrades worked as part of an ART until developmental completion.

The membership of a data science team is ideally co-located to enable efficient scrums and collaboration; however, the nature of MI operations may not make this practical/possible in all cases. In these cases, the team exists as a "hub-and-spoke" team consisting of centralized leadership and distributed team

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 08:32:18 UTC from IEEE Xplore. Restrictions apply.

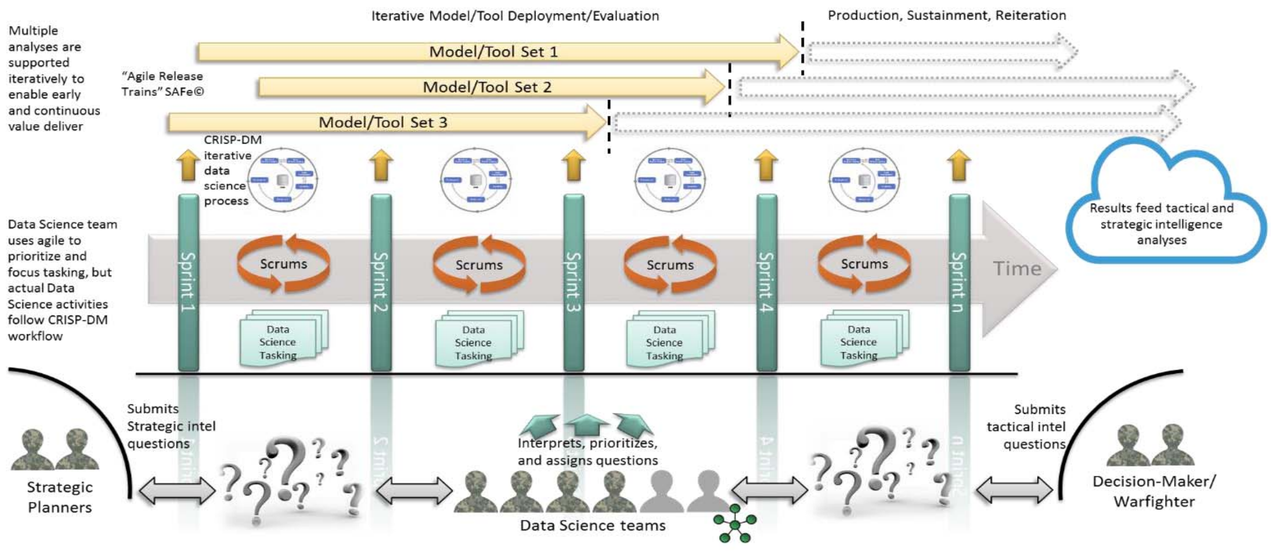

Figure 2: Data Science and Agile

members. The team will function the same way as a co-located team, except that daily scrums and collaboration will happen intentionally and creatively so that the benefits of agile are not lost. This could include the use of teleconferences (e.g. Skype, Zoom), chat/collaboration applications (e.g., Slack, Microsoft Teams), and screen sharing (e.g., Google Hangouts).

The Data Science team employs agility not just in the day-to-day tasks of answering intelligence questions, but also in keeping up with the data science community on tooling, techniques, and research. We recommend that the team participate in data science on-line communities to discover and quickly implement new developments in the field. The ability to experiment with new tools will be necessary so that the team can gauge whether or not to incorporate them into the team's process. A potential way that the team could do this without compromising Army networks would be to utilize commercial unclassified networks for testing tools using unclassified example data sets. To incorporate newly found tools into the production system quickly, we suggest the development of a sufficient, but streamline information assurance process.

As discussed in this study, a critical member of the data science team is the Subject Matter Expert (SME) who brings their understanding of the domain and environment to interpret intelligence questions and results. SMEs collaborate with the strategic and tactical analysts and decision-makers who generated the intelligence questions so that the team can fully address the intent of the intelligence questions and maximize the efficacy of data science products. We see this role as the main connection point between the team and the customers (though it is also reasonable for other team roles to collaborate directly with the customers to understand their data or to evaluate model performance). Our full study depicts a collaborative process based on the agile framework between the data science team (this would be the organic team), and the warfighter/decision-maker.

The purpose of intelligence is to understand the commander's information requirements, analyze information from all sources, conduct operations to develop the situation, facilitate situational understanding, and support decision-making [8]. Data science processes for MI must be able to (at least indirectly) provide capabilities to meet each of these aspects. The Intelligence Cycle is notably iterative and contains steps that can map to the steps of a CRISP-DM process in the following manner: Planning and Direction → Business and Data Understanding; Collection → Data Preparation; Processing and Exploitation → Modeling; Analysis and Production → Evaluation; Dissemination and Integration → Deployment. The similarity of steps leads us to believe that the overall CRISP-DM process employed by a data science team to solve an intelligence question is similar and complementary to the approach being used by everyday regular MI analysts to address intelligence questions. This synergy will help the data science team comfortably work in a normal manner with intelligence analysts throughout the process. Army leadership should be aware that the data science team is performing a variation of the intelligence cycle to systematically develop models that leverage the latest technology and improve analytic rigor while answering the intelligence questions.

The data science team uses agile principles to organize work and execute the CRISP-DM process for intelligence analyses. Modeling and Deployment are the core steps where intelligence questions are answered with data science techniques. CRISP-DM understanding steps, those of business questions and data understanding, are best addressed by those on the team that are immersed in the domain at the tactical level – they know their environment. The Evaluation step is as critically important as the feedback loop to ensure the optimal performance of data

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 08:32:18 UTC from IEEE Xplore. Restrictions apply.

science products. Feedback from users as they put products into practice is essential to determine the efficacy of answering intelligence questions (and at what level of confidence they were they answered). From feedback, the data science team can refine the models and update the products.

As data science and user teams employ agile principles they will enhance collaboration and communication. For example, SMEs of the data science team may conduct a weekly (or daily for fast-moving projects) remote scrum with customer analysts to discuss product develop progress, potential blockers, and to receive feedback on product performance. We believe that a tight communication cycle between the developers and customers is essential to both asking the right intelligence questions and generating correct answers in a more expedient manner.

VI. CONCLUSIONS AND END GOAL

As the Army modernizes under the leadership of the Army Futures Command, AI and data science have become primary areas of concentration. To not address these areas will leave the Army behind the curve and vulnerable to future adversaries. Maintaining technical overmatch means that the Army must adapt to newer methods of AI and data science.

Our work focused on the current state of data science in the Army We interacted with experts in the military, industry, and academic domains, researched the literature, and developed a set of recommendations for implementing data science in the IWfF. Our primary recommendation is the creation of organic data science teams at Army higher echelons to address challenging intelligence questions. These teams will provide improved intelligence analysis by employing modern data science methodologies. Additionally, the teams will form a bridge by collaborating closely with intelligence analysts both in the field and stationed at strategic levels. The teams will include diverse skills that cover the tasks and requirements of the data science analysis processes and are more adaptable to questions in multiple domains across warfighting landscapes.

The most important underpinning of this capability is people. While a subset of data science-related skills exist in US Army organizations, implementation of all aspects of CRISP-DM will require significant knowledge growth in data science engineering perspectives and subject matter expert participation. The collaborative power of CRISP-DM teaming with customer feedback is paramount for a successful implementation. Army leadership views data science techniques embedded within the MI infrastructure as an all-around improvement that will lead to battlefield success.

ACKNOWLEDGMENT AND DISCLAIMER

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-15-2-0043. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] P. Chapman *et al.*, "CRISP-DM 1.0," SPSS, Inc., 2000 1999.

[2] J. B. Rollins, "Foundational Methodology for Data Science," Domino Data Lab, Inc., Whitepaper, 2015.

[3] Microsoft, "What is the Team Data Science Process?," 10-Jan-2019. [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview. [Accessed: 10-Jan-2019].

[4] "What is Scrum?," *Scrum.org*. [Online]. Available: https://www.scrum.org/resources/what-is-scrum. [Accessed: 09-Jan-2019].

[5] I. Ben-Israel, "Philosophy and methodology of intelligence: The logic of estimate process," *Intelligence and National Security*, vol. 4:4, pp. 660–718, 1989.

[6] R. Sensenig, "Eight Practices / 24 Standards of Analytic Rigor," presented at the Data Science Study IPR, USAICoE, 14-Dec-2018.

[7] "SAFe 4.5® Introduction: Overview of the Scaled Agile Framework® for Lean Enterprises," Scaled Agile, Inc., Whitepaper, Aug. 2017.

[8] "ADP 2-0 Intelligence." Headquarters, Department of the Army, 31-Aug-2012.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on June 14,2020 at 08:32:18 UTC from IEEE Xplore. Restrictions apply.