COMPREHENSIVE REVIEW ON AGILE SOFTWARE DEVELOPMENT INTO BUSINESS INTELLINGENCE AND DATA ANALYSTICS

NAGAM VISHAL RAO[1], PURVI JHALA[2], PURVA PATEL[3]

[1]*Student, Dept. of Electronics and Communication, Medicaps institution of Technology and Management, (M.P), India*

Abstract -*Information technology divisions are facing with numerous complexities like competitive edges with the increased pressure for delivering rapid and high quality solutions. Considering these circumstances, the value of data science techniques must also be process with inspiring new perspectives with high reliability. ASD (Agile Software Development) principle has been developed in 2001 with the principle and objective of promoting data science, Business intelligence (BI) and data analytics. Big data phenomenon, variety, velocity of data, volume has been found to impact he BI and data analytics usage that hinder its future research. Hence the paper provided a comprehensive outlook for the challenges and in-depth indeed for the incorporation of Agile technologies into the Big data field. Big data volume has been measured in more than 100 terabytes and also characterised by its datatypes with high degree of variety. With the implementation of Agile approach, the methodology has been found to be less formal, customer focussed and more dynamic. The dynamics for BI delivery form Agile approach to be good fitted with BI but leads to adjustments and information usage for primary software development. The article explored the implementation of agile in overcoming the challenges of business intelligence and data analytics for system upgradation. BI has been emerging due to the amount of generated data by internet and smart devices that has been exponentially grown for the use of individual information.*

Key Words: **Data science, Agile technologies, Business intelligence, data analytics, Agile challenges etc.**

1. INTRODUCTION

The business intelligence BI has been utilized the recent objectives and principles of the Agile Software development in recent days. Due to the incremental and iterative nature of BI, the application of BI developed naturally. The main motivation of this review is to understand the concept of data science, big data analytics and agile technologies which are incorporated into them and also the exhibiting the overall challenges in data science. The basic ideals of data science are researched by various authors who have been emphasized on the documentation, software tools, customer collaborations and change over plans. Thus the development of software will become more dynamic, less formal and customer oriented. High quality technology solutions are in higher demand and competitive edge are maintained in recent days. Based on payback and investments the technology values are determined. The initiatives of BI needs important upfront and investment for constant scrutiny invite and value maintenance on business scores. For organizations the BI value evaluations struggles due to the BI investment challenges. Thus BI plays major role in enabling the companies to become smarter in work, make better decisions with the help of obtained information[1].

To directly attribute an investment return and over time the enabler role is said to be complex and the information usage is expected and routine. To derive the data and collect the information value is another task for the process of information value chain. BI delivery is centred on the chain of information value[2]. The business context and logic applies is the major term which creates the information and it is then consumed by BI users. After the actions and decisions are made. Therefore, for Agile concept also the information value chain is highly significant which is then applied to BI delivery. Through classic waterfall software development BI deliver is not then attained and through data discovery it is highly concentrated and make understand how the information is used in future. For BI deliver Agile principles are then applied.

The major contributions of this study evolves,

- To review the data science concepts, lifecycle based on the recent developments in data analytics.
- To analyse the big data analytics with business intelligence applications
- To discuss about the Agile technologies and exhibit the challenges of data science associated with big data analytics and business intelligence.

1.1 Paper Organisation

The organisation of the paper is as follows. The paper introduces the review on agile technology in data science. A comprehensive review on recent data science methodologies, an outlook of lifecycle of data science followed by the crisp investigation on business intelligence and data analytics. The paper provided an in-depth review on the existing agile technology incorporation into big data, its challenges and future directions. The review has been concluded in the last section of the paper.

2. Data science

For generating the innovative outputs, sources and gain in strategies and productivity, decision support systems like business intelligence- BI and big data analytics provides capabilities. In this paper, it simplifies the terminology based on big data, analytics, BI, computational science and data science and it supported the projects by the decision support system. The scope of the project, equalities and variations among the types of projects and critical success factors identification are classified using the quantitative techniques. Based on the differential business processes, competence in analytics and proprietary algorithms differentiated with respect to big data analytics and BI. Several items have been correlated with strategies for the big data analytics whereas for revenue and cost performances in operational benefits the BI projects are helpful. BI projects managed the complexity in projects and big data analytics projects have been driven b external market drives. While preparing the project plans these discussed factors should be considered [3].

Various nations have been contributed and associated with initiatives of big data analytics. The economic growth has improved in public sectors b these initiatives. The general key deliverables recognised are productivity growth with shared benefits, cost savings, innovations improvement and information communication technology. The major goal of this study is illustrated with respect to big data analytics frameworks for supporting the public sectors and big data organization. Several different platforms of technology have been seen which associated with the public and private sectors and procedures and policies have been generated strengthening the data governance. Using the business intelligence maturity model, the performance analysis has been carried out. Using the data science techniques and recent analytics several data sources have been utilized [4]. For delivering competitive value profits, BI and analytics have been promoted always. The absorptive capacity's interplay described based on the underpinning assets resulted with delivering value process with respect to creation of knowledge. Thus the absorptive capacity significance in business intelligence and analytics domain have been established. With respect to small sized firm's context, the technical knowledge benefitted from analytics for small scale business and in future it benefitted for even in larger firms. Across various cultural frameworks the business intelligence and analytics have been varied [5]. The organizational performance and improved decision making is supported by the big data analytics which is quantifiable and transparent. However, it is not focused on the strong concerns, inconsistencies and chances. The business data analytical methods have been described in Fig.1 [6].

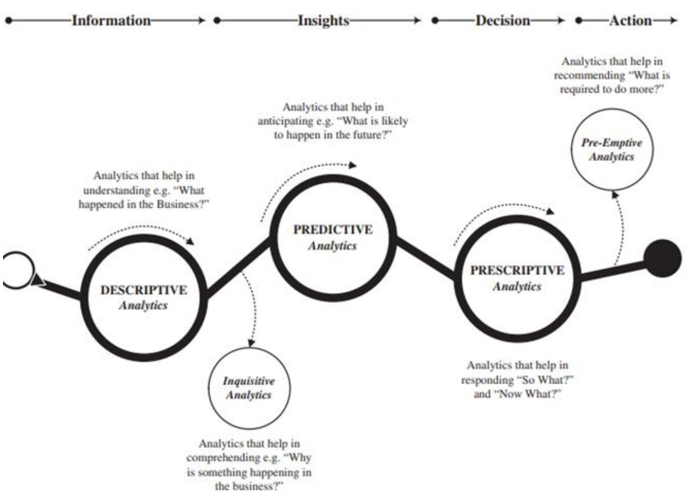

Fig.1. Big data analytical methods and its types [6]

In this study, the big data analytics are utilized and the business intelligence are improved. For this purpose, big data analytics based architecture with service-oriented has developed name BASOA. The enterprise information systems and business intelligence are improved. The interrelationships among the big data analytics and BI are analyzed and the features such as expect ability, relativity and temporality are investigated. Big data science, BI, IoT and big data computing facilitates into further researches. For modelling the above mentioned features Google analytics can be used in future [7]. The business analytics interaction has been provided in this study, comparative with resources of the enterprises. The business analytics values have been generated with accurate conditions. On the IT infrastructure, the BA values are established. The infrastructure of flexible IT is necessary and thus the agility of the organization is effectively enhanced. As part of the organization assets the BA is viewed. If BA is not used the agility of ever companies suffered. The current and the advanced business applications are enabled by the internal communication. For different IT infrastructure, attention should be placed. While implementing the BA models, inconsistency in outcome is achieved. Due to the contributions, BA has possessed various strategic scores [8]. The below table represented the document types statistics in dataset. The articles, book reviews and editorial material are presented in datasets. Between the Big data publications, 77 percentages were articles, editorial materials are 10 percent and reviews are of 6 percent. 89 percent are articles in business intelligence of 1144 publications [9].

2.2 Lifecycle of data science

Scope: **-** the data science and fast analytics came into existence because of big data. For some time, the data science and fast analytics have been emphasized in this study. Based on data mining and fast analytics, visual analytics have been figured out which are utilized with the data science. The latest version of the methods of data analysis have been considered as data science and fast analytics. During the data mining discover phase, the visualization has been utilized in which the complementary methods are the data mining and visual analytics. The raw data is explored by the visualization tools from big data analysis. For the visualization, new tools have emerged in the business intelligence enterprises. In the data science process, the raw data is explored b analysis of data. For analysis the data is first acquired and is considered as data science for fast analytics scope. The data product namely classifier, recommendation engine and prediction engine have been generated. In analysis of scope various data sources have been introduced. To the unstructured data, data sources have not been restricted. For the analytical data, BI program analytics can possess the scores.

Data discovery/ acquisition: - to obtain data without the structure understanding new technologies have been made it possible which is the alternate of the lifecycle of BI from which the data is evaluated before the data repository loaded. Without data modelling, Hadoop file system initialized at Google and utilized as open source format from the enterprises. For visualizes, assess and accessing to inventing the data sources usage and scores the analysts used the fast analytics. Data lake considered as new data repositories in which the technology enables the processing power and storage supporting the larger unstructured datasets.

High-frequency keywords in related publications.

High frequency keywords			
Literature of "Big Data"		Literature of "Business Intelligence"	
Key words	Occurrences	Key words	Occurrences
Model	549	Management	109
Algorithm	480	Data warehouse	104
System	401	Big Data	84
MapReduce	389	Data Mining	91
Cloud computing	376	Systems	81
Management	363	Model	79
Networks	347	Performance	66
Information	308	Data analytics	61
Classification	307	Knowledge management	54
Data mining	280	Information	52
Machine learning	262	Social media	46
Performance	262	Information systems	44
Social media	255	Design	40
Privacy	240	OLAP	39
Internet	233	Web	38
Surveillance	215	Decision support system	37
Data analytics	205	Competitive intelligence	29
Hadoop	186	Information technology	28
Prediction	181	Business analytics	27
Optimization	176	Decision making	26
Internet of Things	167	Design science	26

Fig.2. Data keywords in recent publications[9]

Evaluate/Visualize: - visualization and analysis are the iterative process for both data science and fast analytics. To supporting the analysis, the visual analytics is the basic goal with the fast analytics. Visual product refinement generates the new knowledge. New scorecards or dashboards utilized and produced b iterative fast analytics in the presenting BI or the one-time analysis tools produced in supporting innovation knowledge benefit. The parameters are identified and useful in the descriptive analysis. In analytical models the parameters and variable relationships are identified.

Model/ Development: -in fast analytics the used data is describes as data science with analytical modelling and general data modelling. The predictive, prescriptive and descriptive evaluation with the help of machine learning approaches are comprised in the analytical modeling. The machine learning approaches are the clustering, regression and classification. The data science is correspondingly presented in the process of verified the analytical model in which the modelling error is reduced.

Validate: - the analytical model validation is represented in this phase in which the errors are reduced. For the integration of analytical modeling, new parameters with fast analytics are recognized. This kind of process is named as fitting the model. Little values are exhibited from the analytical models, scorecards and dashboards.

Feedback/support: -as the organizational environment modifications, analytical products has to be revised and supported. Based on the rate of change in industry and organization the analytical model lifecycle is exhibited. An applicability over time and feedback users determines the lose value of analytical models

Fast analytics/Data science synthesis lifecycle and agile: the three phases namely design, development and discovery phases benefitted from the small time box increments, co-located resources, iterative cycles and stakeholder collaboration. Both data science and fast analytics are agile inherently and everything follows the small teams, iterations and collaboration is necessary among the technical resources and business subject matter experts [10].

3. Big data analytics

In latest technologies in information and communication, big data analytics needs computing resources in greater amount and big data adopting technology is said to be inexpensive for various small to medium organization. Through cloud computing the big data analytics are deployed and the benefits and challenges have been outlined. The data mining processes are improved from these dominant technologies which enables the decision making process enabled. While using the specific cloud service model, certain issues are also analyses in this study [11]. Certain critical big data features have been presented and different was are recommended in business intelligence. The big data analytics have been analyses in this study which are helpful in building the business big data based analytical models and the business intelligence are further improved [12].

By the traditional data mining approaches, the big data brings out the problems like larger number of variables and imbalanced data distribution which are not managed efficiently considered as drawback. From one single data source most of the traditional methods are utilized. Combining the data from various sources, certain difficult business issues are handled. On the purchase behavior, the transaction records and social media data have been combined. To study the similar marketing issue with various data sources, the heterogeneity between data sources are handled. Due to the various analysis methods and data collection, various structure, objectivity, granularity and quality have been utilized to review the similar marketing issue [13]. On optimization and simulation of big data analytics and business intelligence refer as several intelligences, this stud is focused. The knowledge emergence is in the form of framework of business intelligence which is appropriate for the particular enterprises. Using the big data analytics, the BI model is framed. The data architecture and knowledge repository are in the model of the diagnostics dashboard [14]. To the industries and academia, business intelligence and big data analytics are appealed considerably. As an emerging frontier, this current stud focused on the big data analytics and thinking and also artificial intelligence. This study is based on the industries and academic development of the traditional approaches and it provides the analytics intelligence and analytics thinking details. The interrelationships among them have been discussed in detail. Big data analytics internal relationships have been discussed. AI- driven big data analytics intelligence have been presented with the unified technologies. Thus further, the development of big data analytics intelligence, thinking, data science, AI and BI are highly concentrated in recent days [15].

3. Agile technologies in Data Science

Agile technologies retain the consistency of the engineering process and helps both the software engineers and stakeholders to build, maintain and deploy complex software. This has been introduced formally by the integration of four core values and twelve principles. Agile has been used characteristically in several software Engineering (SE) periodicals like IEEE Software, IEEE Transactions on SE, Journal of Systems and Software, Information and Software Technology and Empirical Software Engineering. This research has been recognised in flagship conferences like and other such reputable conferences[16].

The paper [17] presented the BDA big data analytical setup influencing the parallel processing and data distribution across resource clusters. The analytical system introduced several new challenges particularly for analytics. This comprehensive lifecycle portion characteristically followed a waterfall process that completes one step previously initiation. Numerous efforts have been performed for mapping several analytic types to specific agile methodology in which the steps are described often as the breaking activities to smaller tasks. The overall process is consistent still with stepwise waterfall manner. BDA modifies a number of the activities in the analytics lifecycle, as well as their ordering. The agile analytics goal for reaching the point of optimal in between value generation from the data and the time spent. This paper discussed the consequences of an agile methodologies for BDA in transformation, cleansing, and analytics.

The study [18] developed an enhanced process for handling the government data analytic process by considering the knowledge lacking in this domain. The study also proposed an organised method for the detection of business issues. The study also allowed earlier cooperation of the business people with the data scientist. The study implemented the suggested framework in governmental organisations. This work also helps in the effective mapping of the framework with the core work with the use of real time dataset. The study has been passed via all the phase with the use of various real datasets from the HR system that manages 65000 employee record. The study explained the implementation with the use of Turnover dataset by the employment of data dimension and business objective dimension, design thinking, data mining goal, pre-processing, modelling, evaluation and deployment. This framework into various datasets may also provide our validation to be more strength. Further this data set could be applied to combined dataset with higher dimensions.

The study[19] presented a dispel4py a multipurpose data intensive kit as a standardized Python library. It enables the scientists to test and experiment ideas with the use of their desirable prototyping environment. This package delivered mappings to different computing infrastructure, comprising HPC architectures, cloud technologies, and particular data-intensive machine, to seamlessly move into production with large scale data loadings. These mappings are completely automated and hence the encoded data handling and data analyses are unchanged completely. The supporting model is a lightweight arrangement of fine grained operations on the data that has been integrated together by data streams using the available lowest cost technology. Such fine grained flow of the work is interpreted locally during the development and being mapped to several nodes and systems like Storm and MPI for synthesis. The study[19] explained the importance of the approach for data-driven research rapidly innovating and exploiting the growing wealth of data by adapting the current techniques. The study depicted the provision of provenance management to improved reproducibility and understanding, and the process of registration support sharing and consistency. The study obtained three application domains for measuring on multiple infrastructures and optimisations. In conclusion the study presented the succeeding steps for achieving performance and scalability.

4. Challenges

While big data technology has been originated as a striving new scientific field, complexities and questions arise for the need of data science. Optimal discussions concerning the intelligence and intrinsic complexities in data science challenges and gaps are discussed below.

Data science issues needs methodologies, systematic thinking, and methodologies to support the deep learning and machine learning branches. The conceptual data science landscape assists the data scientists to represent, understand, and develop the intelligence and complexities the prevailing problems. Data scientists attempt to invent intelligence and data driven machines to reinforce, learn, simulate, and transform imagination, human like perception, creative thinking and curiosity through human data cooperation and interaction. The data science core is the complex exploration that has been inherently extracted in the business, data and problem solving frameworks.

Complexity denotes to intrinsic characteristics in the data science networks. The paper treat data science problems as X-complexities, domain, behaviour, context, social factors, deliverables and learning. Data complexity is also reflected on the basis of data characteristics and circumstances comprising high dimensionality, extreme imbalance, large scale, real-time and online interaction, mixed sources, strong dynamics, cross-media applications, uncertainty, noise mixed with data, high frequency, unclear hierarchy, unclear structures, strong sparsity, heterogeneous distribution, and unclear availability of specific sometimes critical data[20].

Another critical problem comprises the complex relations in data that are difficult to understand the forces hidden in data. These complex relations comprise comprehensive couplings that could not be explained through prevailing dependence, association, causality correlation, theories and systems[21]. These comprise implicit and explicit, non-structural and structural, syntactic and semantic, vertical, and hierarchical, global and local, non-traditional relations and traditional as well as evolution and effect.

Behaviour complexity denote the challenges in the understanding of business activities through the integration of semantic analysis and behavioural subjects in real time that are simplified in the data world developed by physical activity in the data transformation process. The X-intelligence and X-complexity in the complex data science systems and broadening the gap in-between invisibility in attaining the yielding capability to new research challenges motivating the data science development.

In understanding the mathematical and statistical analysis, the objective is to enable the data scientists to describe, disclose, represent, and the capturing the intelligence complexities for extracting the action insight. Prevailing computational and analytical theories has to be investigated for insufficient or problematic for redeveloping the address the complexities for heterogeneous, supporting multiple, large scale hypothesis survey design, testing learning inconsistency as well the uncertainty of multiple dynamic data sources[22]. Apart from these challenges the following are also considered as main challenges.

X-analytics.

Challenges arises due to tools, analytic theories, and systems that not available readily to represent, implement, discover, and data managing intelligence and supporting data science.

Social Issues and Data quality.

For the purpose of identification, specification, and respecting the social issues in the domain-specific data, the business understanding and processes comprising the usage, security, privacy, and trust that make possible social based data science tasks that are not previously well processed.

Decision making challenges:

For the development of decision support theories and systems for enabling data-driven decisions and transformation, incorporation of prescriptive strategies and actions that could not be managed through the prevailing system and technologies[23]. The companies cannot avoid the agile wave, but also faces various challenges. Social process featured by collaboration and communication between the member communities is crucial for the success of agile methodologies.

5. CONCLUSIONS

Investigation of the Agile principles provides a comprehensive understanding on usage of an Agile approach with BI delivery. For reiterating the principles, interactions and individuals over the tools and processes, working of software over the comprehensive documentation, over contract negotiation, customer collaboration and change over response the paper provides a detailed review on Agile technology. Variety, velocity of data, volume has been found to influence the BI and data analytics use that demotes the beliefs of BI and data analytics. Hence this paper provided a comprehensive outlook for the challenges and urgent need for the incorporation of Agile technologies into the data mining and data science

REFERENCES

[1] D. Wang, J. D. Weisz, M. Muller, P. Ram, W. Geyer, C. Dugan, et al., "Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI," Proceedings of the ACM on Human-Computer Interaction, vol. 3, pp. 1-24, 2019.

[2] H. Atwal, Practical DataOps: Delivering Agile Data Science at Scale: Springer, 2019.

[3] G. J. Miller, "Quantitative Comparison of Big Data Analytics and Business Intelligence Project Success Factors," in Information Technology for Management: Emerging Research and Applications, ed: Springer, 2018, pp. 53-72.

[4] S. Manikam, S. Sahibudin, and V. Kasinathan, "Business intelligence addressing service quality for big data analytics in public sector," Indonesian Journal of Electrical Engineering and Computer Science, vol. 16, pp. 491-499, 2019.

[5] K. Božič and V. Dimovski, "Business intelligence and analytics for value creation: The role of absorptive capacity," International journal of information management, vol. 46, pp. 93-103, 2019.

[6] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," Journal of Business Research, vol. 70, pp. 263-286, 2017.

[7] Z. Sun, L. Sun, and K. Strang, "Big data analytics services for enhancing business intelligence," Journal of Computer Information Systems, vol. 58, pp. 162-169, 2018.

[8] X. Chen and K. Siau, "Business analytics/business intelligence and IT infrastructure: impact on organizational agility," Journal of Organizational and End User Computing (JOEUC), vol. 32, pp. 138-161, 2020.

[9] T.-P. Liang and Y.-H. Liu, "Research landscape of business intelligence and big data analytics: A bibliometrics study," Expert Systems with Applications, vol. 111, pp. 2-10, 2018.

[10] D. Larson and V. Chang, "A review and future direction of agile, business intelligence, analytics

and data science," International Journal of Information Management, vol. 36, pp. 700-710, 2016.

[11] B. M. Balachandran and S. Prasad, "Challenges and benefits of deploying big data analytics in the cloud for business intelligence," Procedia Computer Science, vol. 112, pp. 1112-1122, 2017.

[12] F. Khoshbakht and S. Quadri, "A Study on Analytics of Big Data and Business Intelligence-A Review."

[13] S. Fan, R. Y. Lau, and J. L. Zhao, "Demystifying big data analytics for business intelligence through the lens of marketing mix," Big Data Research, vol. 2, pp. 28-32, 2015.

[14] M. Jayakrishnan, A. Mohamad, and M. Yusof, "Understanding Big Data Analytics (BDA) and Business Intelligence (BI) Towards Establishing Organizational Performance Diagnostics Framework," Int. J. Recent Technol. Eng, vol. 8, pp. 128-132, 2019.

[15] Z. Sun, "Big Data Analytics Thinking and Big Data Analytics Intelligence," Big Data Analytics and Intelligence, 2020.

[16] R. Hoda, N. Salleh, and J. Grundy, "The rise and evolution of agile software development," IEEE software, vol. 35, pp. 58-63, 2018.

[17] N. W. Grady, J. A. Payne, and H. Parker, "Agile big data analytics: AnalyticsOps for data science," in 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 2331-2339.

[18] W. Qadadeh and S. Abdallah, "An Improved Agile Framework For Implementing Data Science Initiatives in the Government," in 2020 3rd International Conference on Information and Computer Technologies (ICICT), 2020, pp. 24-30.

[19] R. Filguiera, A. Krause, M. Atkinson, I. Klampanos, and A. Moreno, "dispel4py: A python framework for data-intensive scientific computing," The International Journal of High Performance Computing Applications, vol. 31, pp. 316-334, 2017.

[20] A. Shukla, "Role Oriented Challenges for User Experience Design in Agile Scrum."

[21] K. Conboy and N. Carroll, "Implementing large-scale agile frameworks: challenges and recommendations," IEEE Software, vol. 36, pp. 44-50, 2019.

[22] R. G. Cooper and A. F. Sommer, "New-Product Portfolio Management with Agile: Challenges and Solutions for Manufacturers Using Agile Development Methods," Research-Technology Management, vol. 63, pp. 29-38, 2020.

[23] R. B. Svensson, R. Feldt, and R. Torkar, "The unfulfilled potential of data-driven decision making in agile software development," in International Conference on Agile Software Development, 2019, pp. 69-85.