

A Novel Approach for Successful Delivery of Big Data Projects in Software Industries

Chandrasekhar Banda*, Phani Kumar S

Department of Computer Science Engineering, GITAM University, Telangana, India

Research Article

Received: 31-Dec-2021, Manuscript No. grcs-22-50690; **Editor assigned:** 03- Jan-20212, Pre QC No. grcs-22-50960 (PQ); **Reviewed:** 05- Jan-2022, QC No. grcs-22-50960; **Accepted:** 10- Jan-2022, Manuscript No. grcs-22-50960 (A); **Published:** 17-Jan-2022, DOI: 10.4172/ 2229-371X.13.1.001

***For Correspondence:**

Chandrasekhar Banda, Department of Computer Science Engineering, GITAM University, Telangana, India

E-mail:

bandachandrasekhar2020@gmail.com

Keywords: Prediction; Big data

analytics; Smart; Project management

ABSTRACT

The changing needs of the people are being supported by the changing technological advancements. The changing technologies have always been a golden duck to the software industry. Big data has forced all the investors to change their mindset and focus on the projects from various domains such as finance, health, and many more on analytics projects based on big data. Though there are various strategies being adopted by most of the software companies, those strategies have failed in successful delivery of big data projects. Hence, in this paper, we have proposed a comprehensive profitmaking framework to implement on Big data and Analytics projects. The proposed model had all the advantageous aspects in delivery the projects successfully within schedule and cent percent customer satisfaction and highest prediction accuracy.

INTRODUCTION

The term big data has gained a prominent position in the current technological era. The world is worried about big data management (viz., storage), but even processing such huge data has become complex. The software industries have started developing various tools which are capable of holding big data. The area where big data arises is Health care, Finance Marketing, and Education, etc. The software market has huge potential for developing projects in these domains ^[1].

Nevertheless, the major challenge is developing projects related to big data. Very little research work was done in this area. As per the Gartner survey, it is found that about 80% of the big data projects have failed. The reasons for

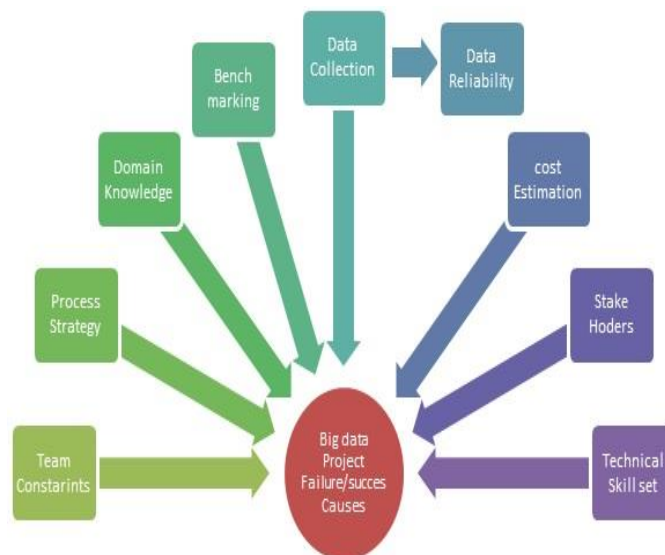
failure have been analysed by various researchers. However, the reasons for failure investigated by them are substantially not enough to reduce the failure rate [2].

Software industries usually followed traditional process models viz., the waterfall model adopted from other sectors like construction or engineering industries. All the traditional process models were observed to be successful when the customer requirements are predefined, static, and known apriori. However, the changing needs, dynamism in requirements, lack of clarity in requirements for the end-user, etc., have forced the software industry to adopt a different approach agile model where the customer is also made as a part of the project development and to a product developed iteratively.

The major focus is to what extent the traditional models suit the current trending projects based on big data. As per the Gartner 2018 report, it is observed that about 60% of the big data projects fail at the initial level only [3]. Though Gartner's report 2018 seems to be too submissive, stating 60% project failure, in reality, it is observed that the failure rate is about 80% [NVP 2019]. The failure of projects might happen for various reasons, such as:

1. Lack of customer needs properly.
2. Not asserting the future requirements of the customer.
3. Software cost estimation is done improperly.
4. Lack of experts in the relevant domain.
5. Data.

Figure 1. Factors influencing the big data projects.



According to the literature available (Figure 1), various agencies conducted many surveys; there should be a much more sophisticated, feasible, and robust model that should help the software industries handle big data projects to successfully deliver the products with maximum confidence of the product and the customer well [4].

In this paper, we are attempting to reconcile the aspects which are a point of focus for the software industries while accepting the projects and predicting the software project cost estimation and failing in cost estimation, along with various other factors that are impacting upon the failure of big data projects.

The remaining sections are organized as follows: Section-II elaborates the available work done till date, along with the causes of failure, and Section-III of the paper highlights what other factors that are to be considered for big data projects a success ^[5]. Section IV of the paper clearly emphasizes the importance of data and data management in the software industries to reduce big data project failure. Finally, Section-V concludes the paper.

MATERIALS AND METHODS

Related works

The term Big coined about two decades ago and gained the highest importance in the current technological era. After successfully implementing the big data projects, which are preliminary, most companies have focused more on these big data projects. The software companies have started gradually adopted big data analytics ^[6].

Few projects used to fail due to predicting the cost estimation process has researched the prediction of software cost estimation. Earlier, during the 70s, software cost estimation was done purely based on measurable parameters such as Lines of Code (LOC) and other direct measures ^[2]. Nevertheless, the cost estimation of the software is sufficient on just the LOC ^[7]. However, it depends on various other indirect measures derived in the mid-1970s using mathematical technique. He has used a few graph theory concepts termed cyclomatic complexity metric, an alternative to size based metrics also developed metrics for effort estimation based on statistical techniques. Shen, et al. 1984 also proposed certain metrics that are considered the number of bug density. It is observed that, if the density of bugs increases, there is a high probability that the product fails.

Chidamber, et al. 1994 have evaluated object-oriented metrics called CK metrics. These CK metrics were Weighted Methods Per Class (WMC), Depth of Inheritance (DIT), Number of Children (NOC), and Coupling between Object Classes (CBC), Response for a Class (RFC), and a Lack of Cohesion in Methods (LCOM).

Fei et al. 2004 incorporated a few improvements to Halstead's complexity metrics by adding weights. The authors have suggested different weights for operands and operators.

The cost estimation models were classified into two major categories based on the approach/procedure, such as Algorithmic and Non-Algorithmic. However, all these models are best suited only for traditional development methods ^[8].

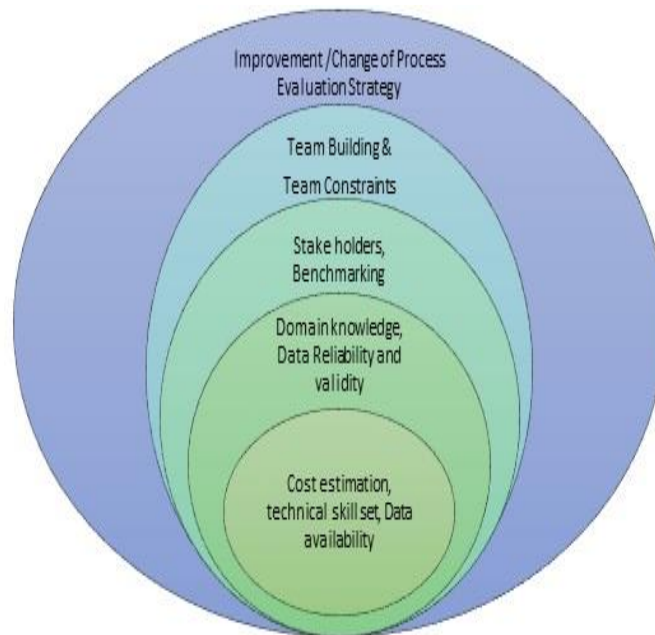
Michael Bloch, et al. 2012 surveyed, along with the University of Oxford, observed that about 45% of the projects run over budget and 7% over time, leading to the highest risk of cost and schedule overruns.

IT executives identify four categories of issues that impact most of the project failures.

- Missing focus
- Content issues
- Skill issues
- Execution issues

Assessment of the above four issues is termed to be a "Value Assurance" that indicates the project's wellbeing. This approach had a solid track record. Figure 2 below shows the success factors for each category of "Value Assurance" wherein, there are 5 factors which are considered to be as Value Assurance.

Figure 2. Value assurance.



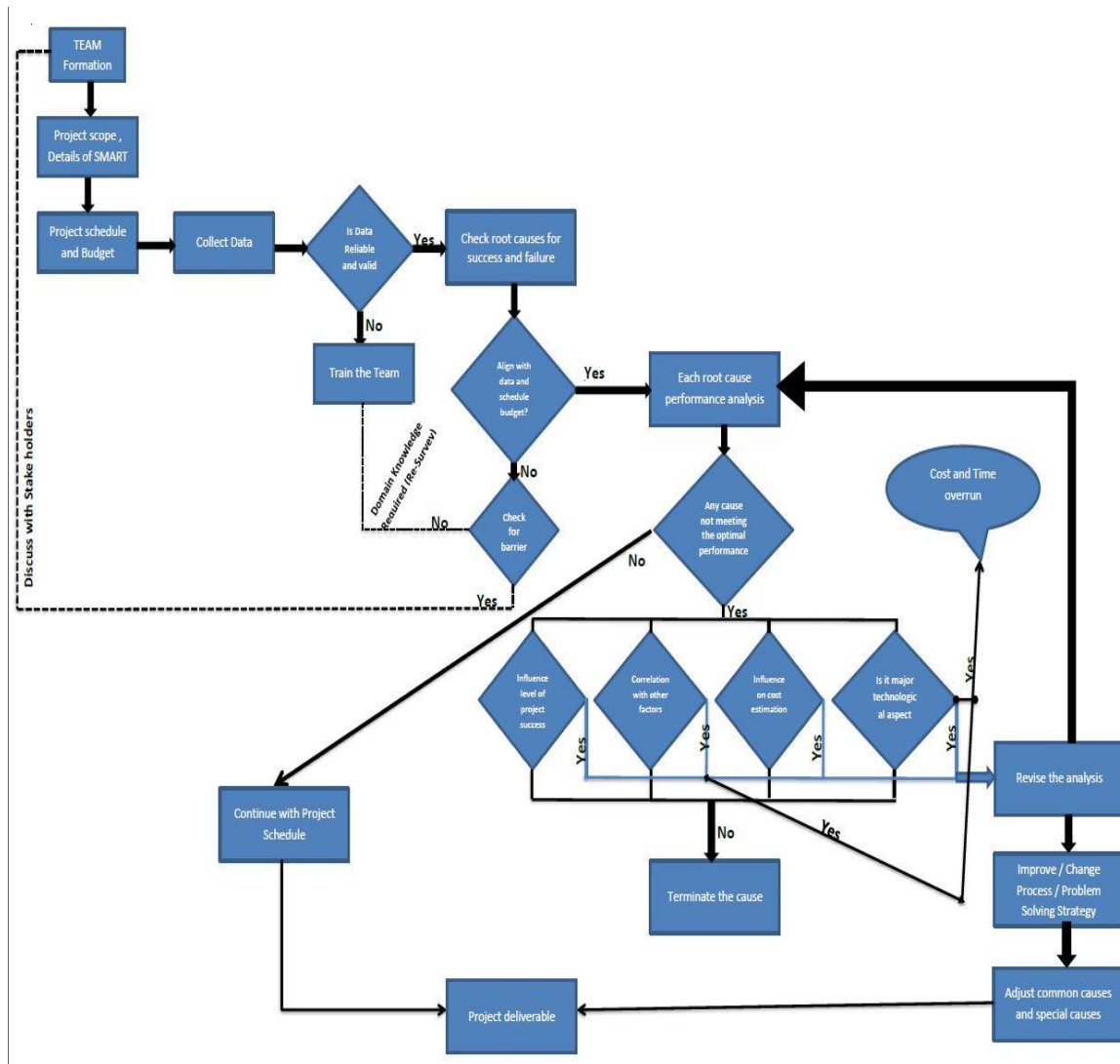
Gabriel Kabanda 2019, in his research findings has stated that the highly critical aspects of big data projects are finding the right persons for the project's success and the customer satisfaction and the projects' assignments to the working teams. He anticipated that AI and ML usage in the projects' assignment would aid the projects' success (Figure 2).

The third major aspect of big data projects is the data itself. The major issue is, "Is the Data available sufficient? Is the Data perfect/ready to use? Are the fields in the data sufficient for analyzing the desired aspects?" Answering these questions shall reduce the risk of failure to the extent of 50% which has been discussed in the next section.

Proposed framework

Though there has been not much research done in this area, the researchers have worked from various angles and stated the reasons for big data projects' failure in industries ^[9]. This problem throws a big challenge for the software industries and the AI-ML-based start-ups to successfully deliver the projects. The major aspect of these big data analytics projects is gaining a level of confidence. The credibility of the industry lies in the accuracy obtained in the predictive analytics for a particular project ^[10]. Thus, in this paper, we have proposed a new hybrid frame framework for successfully delivering big data analytics projects. Though various researchers have stated that the projects' failure for submission is one or two, the statistics still are astonishing.

Figure 3. Proposed framework for success of big data analytics projects.



The framework has four broad categories such as:

- i. Prediction of cost estimation metrics,
- ii. Human resources,
- iii. Data management and
- iv. Project management.

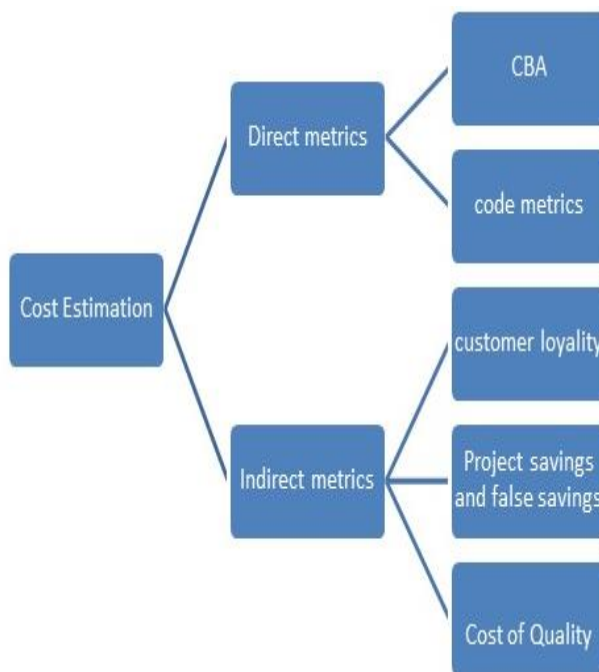
The above four categories are to be implemented in detail for every aspect inspected at our location for the BDA projects (Figure 3).

RESULTS

Prediction of cost estimation metrics

Cost estimation also referred to as "Effort Estimation," (Figure 4) plays a vital role in making the companies gain profits [11-14]. According to the study directed, it is discovered that about 45% of the projects fail due to the cost overrun or due to over budget [15]. Generally, there are two models, namely Algorithmic and Non-Algorithmic models, to estimate software costs. There are well-defined cost estimation tools available that can predict the cost for general projects. Also, there are many projects which have been already developed by the company which had a template for cost estimation [16,17]. So, the team can estimate the cost easily if similar projects are already developed. Also, there is a possibility that the cost estimation can also be done using Expert Judgement (A non-algorithmic) when similar projects were not developed earlier, and the template is not available and given the condition that the members of the team have past expertise in developing the similar kind of projects has proposed a hybrid method for both object-oriented metrics and traditional metrics [18]. As most of the current BDA projects are developed using an object-oriented approach, this model suits well, yielding about 95% accuracy.

Figure 4: Factors for cost estimation in big data projects.



The projects are initially categorized based on the domain to which they belong, and every domain is estimated with a certain parameter like, number of methods in class, inheritance depth, number of coupling, number of children classes, complexity, average cyclomatic complexity, etc., and rated High, Medium and Low with certain weights. This shall make the project's cost estimation easier and highly accurate. The Table 1 below illustrated show an example of 3 different kinds of domains and for both Traditional and OO Metrics Analysis.

Table1. Traditional metrics analysis.

Category	Stability	Paths	Cyclomatic complexity	Nesting	SLOC Maths	MCDC	Complexity	LOC	Time
Finance	H	L	M	H	M	M	L	H	H
Gaming	L	H	H	M	H	H	H	M	M
Communication	M	M	L	L	L	L	M	L	L

Table 2. Object-oriented metrics analysis.

Category	#Methods in the class	Inheritance depth	Coupling #	Children classes #	Complexity	Avg. CC
Finance	L	M	M	L	L	M
Gaming	H	H	H	H	H	H
Communication	M	L	L	M	M	L

Apart from the Table 2, it is essential to understand the data's size to be analyzed also. The aspects of data size are discussed in the later sub-sections.

Human resources: In the traditional projects, the human resource estimation is performed based on the #LOC, #Days to deliver, etc., and then used to deploy the employees in such projects. The Team Leaders or Project Managers used to assess the time left and the quantity of work to be done, and in tough situations, extra people were used to deploying into the team to avoid schedule overrun [19]. Poor estimation of work products and workforce required leads to schedule overrun. However, such cases listed are too few. The unexpected aspect concerning the workforce is either people quitting in the middle of the project due to better opportunities elsewhere. However, the BDA projects also had a lot of such issues raising such as skills of employees, technology adaptation, team building, management of employees' in-terms of perspectives, delivery cycles (Long or Short), and management practices as most of the employees in the companies are not skilled enough to handle BDA projects. It is essential to provide them the required training in the technologies [20]. High performing and experienced experts shall help understand both the technical and business concerns that are highly essential for BDA projects. The team members' expertise shall increase the performance up to 100 percent with their judgmental ability to interpret the data and its pattern. As there are a wide variety of technologies available that can address the BDA projects from various domains, it is essential to have expertise in both technology and domain to identify the best possible technologies and feasible ways to solve the problem and understand the insights of the projects which helps in identifying the gaps if any, while gathering the business requirements.

Table 3. Performance of different types of IT projects varies significantly.

Project type	Avg. cost overrun	Avg. schedule overrun	Avg. benefits shortfall
Software	66	33	17
Non-Software	43	3.6	133
Total	45	7	56

In the Table 3, the software projects have a serious issue with the schedule overrun, and non-software projects have the highest shortfall of average benefits. The factors influencing the cost overrun are frequent change or ambiguities in requirements leading to cost overrun, poor estimation of expenditure on the project, etc. The shortfall of benefits is due to the lack of technical sound persons, poor communication, ambiguities in requirements, etc. Most of the big data projects fail due to a lack of clarity in the requirements leading to a shortfall of benefits. Hence, it is essential to depute skilled personals in big data projects.

The following are the most important factors which are highly influencing the failure of big data projects.

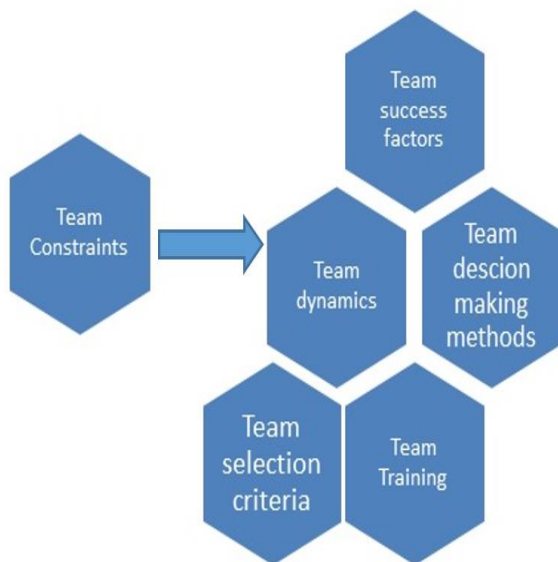
Skill: Skillset is an important aspect for any person to complete the given task successfully within the stipulated time. The data scientists are in no case different from any other IT professional. With the advent of big data being huge in size of data, the skills of managing the traditional data lead to many failures due to the reason that the traditional data is well organized, whereas the big data is both structured and unstructured. So, it is essential to draw insights into the statistics. The ability to extract relevant information from the humungous amount of data that is being processed every minute. Hence, the most important skills that are required for any big data analyst are Analytical thinking, Logical thinking, Technical expertise, Communication skills, domain knowledge, business skills, and above all, Quantitative Aptitude and statistics.

Technology adoption: In software industries, especially projects like big data analytics, where the technologies that can be used are frequently changing, they need immediate attention from the team members. The team members should be adaptable to the new technologies at the earliest. At the same time, companies need to provide the employees with the required training, and the cost of training also has to be considered while predicting the software cost estimation [21]. Few companies have deployed separate training wings in the company to train their employees. The company should support employees by providing sufficient training on advanced technologies to enhance their skills, cope with the requirements, provide necessary tools, and encourage employees based on their performance. In today's online era, people are habituated to update their skills of the most advanced technologies from home only. There are enormous platforms that provide online training viz., plural sight, edureka, udemy, etc., on certain advanced technologies provided to the employees.

Team building

Team building is the most critical and crucial phase in the project implementation process. The team should be a blend of all members having different skillset, with varied experience levels [22]. However, many surveys have concluded that there exists certain constraint in forming the teams, as shown in Figure 5. The team building had many factors for consideration. The literature had various models adopted in the process of team building. The constraints which need to be considered are dynamics, selection criteria to be adopted.

Figure 5. Team constraints.



Stakeholders: Stakeholders play a pivotal role in finalizing the requirements from time to time and assessing the project's performance and usability. The project development inputs have to be gathered from various sources such as the end users, customers, and various business organizations working with similar projects. This process shall help enhance the project's performance and provide scope for providing additional features into the projects that shall, in turn, help improve the client's satisfaction levels (Figure 5).

Figure 6. Sources of data for the analytics projects.



In today's analytical world, data is treated to be a vital component in the analytics based companies. So, most of the businesses are turning out to be data-driven. Getting data from various sources is also challenging task. Especially, identifying the right person for gathering the data is the most challenging issue.

“The voice of the customer guides world-class leaders’ every action and decision”

-Yu Sang Chang, George Labovitz, and Victor Rosansky.

The Figure 6 above gives an idea of gathering data from various sources like, client's voice is the first most important aspect, then gathering the information from the business organizations and involving the most important target group i.e., end user who is the ultimate beneficiary. Identifying and selecting an appropriate data collection methods (*viz.*, interviews, surveys, focus groups, observations) helps in improving the projects deliveries more efficiently and successfully.

DISCUSSION

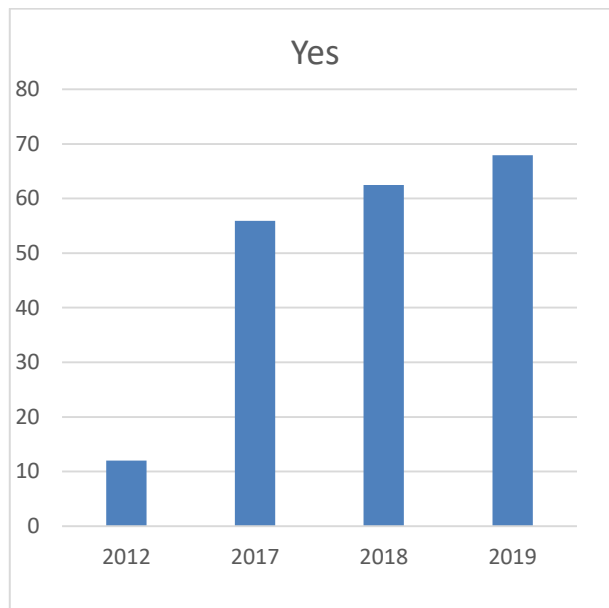
Data management

Specialized persons: Successful delivery of the software projects highly depend on the expertise and skillset of the employees. Certain projects shall need to deploy highly efficient and specialized persons. Similarly, the projects related to big data analytics shall also need to have deployed specialized persons like Chief Data Officer, Analytics officer. As per our proposed model, the team should consist of a perfect balance of the right people. It is observed that apart from the skilled persons, the big data projects are vulnerable to receive huge amounts of data at a very higher rate with varieties of information, etc. Using the data as is received shall not serve the purpose of the big data projects and may lead to wrong predictions as there might be lot of missing values or other issues which are discussed in the next subsection. Hence, to handle all these issues related to data, it is essential to deploy an expert in data management. As per a survey conducted by NVP, 2017, it is observed that those companies which have employed Chief Data Officer (CDO), Chief Analytics Officer (CAO), Chief Information Officer (CIF), and Head of Big Data (HBD) shall reduce the risk of improper data leading to mispredictions. Deploying experts into the project shall accelerate the pace of the project completion and also increases the accuracy in prediction ^[23]. The recruitment of CDOs has been increased drastically since 2012. Table 4 clearly demonstrates that the deployment of specialized person shall increase the probability of the project getting successfully delivered. The graph in Figure 7 shown represents the increasing in the specialized persons in the company for big data analytics persons.

Table 4. Statistics of appointing specialized person in company.

Appointment of a expert	Chief data officer
2012	12
2017	55.9
2018	62.5
2019	67.9

Figure 7. Graph showing increase in appointing CDOs in companies.



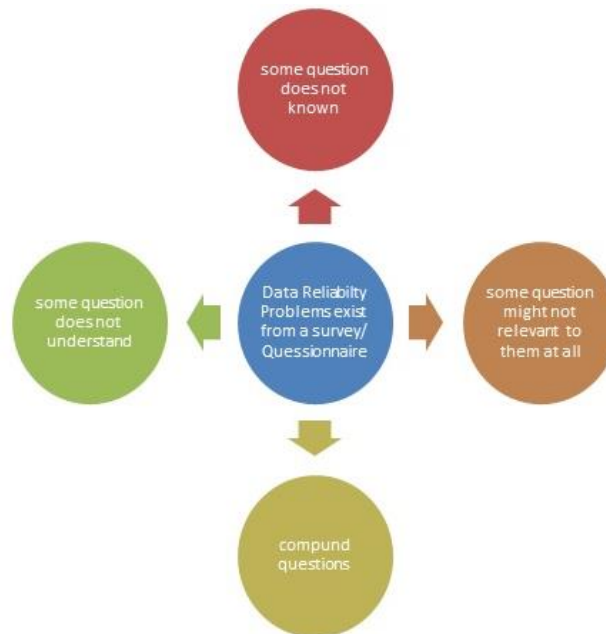
Data treatment: The term Big Data was used initially coined in about 1990s and has gained wider prominence since 2012 ranging from terabytes to zettabytes transforming to heavier data. Now, the data is increasing drastically with a higher velocity and veracity having various varieties of data [24]. Data has to be obtained from various sources such as from Focus groups, interviews, observations on the domain from the domain experts, etc., as shown in Figure 8. It is essential to identify the right stakeholders for getting the data.

Figure 8: Data collection from various sources.



After acquiring the data, it is also essential to check for reliability of the data. As the data is acquired from survey/questionnaire, there is a chance of encountering reliability issues shown in Figure 9. Reliability issues are found due to the issues like few questions might be irrelevant or ambiguous and few questions may not be understood by the responder. Few questions may be compound. Also, it is very important to analyze the answers/responses in the survey form because of the reason that the answers might be biased or misinterpreting. Hence, data treatment has to be performed to avoid mispredictions.

Figure 9. Causes for data reliability issues.



Required parameters: Prediction is defined to be a statement made by someone thinks shall happen in future based on certain assumptions or statistics or following certain patterns. But, it is essential assess the necessity of parameters which shall influence the prediction analysis. Prediction accuracy is completely dependent on the parameters used and the patterns identified. The reliability on any analytics projects lies in the accuracy of prediction. The higher the accuracy is, the higher will be the customer satisfaction which shall in-turn help in expanding the business. Retaining or withstanding the confidence of the customer is the most difficult task in the Hence, in this paper we suggest the best way of identifying the parameters.

In our proposed model, the parameters may be classified into 3 levels such as direct linked parameters, indirectly linked parameters and hidden parameters. The details of parameter categorization are discussed in the next section [25-28].

Project management: Project management is the most crucial aspect in handling the big data & analytics projects. Any company shall have an assessment model for the PMV (Project Management Values) for the measurement of the performance of project management. The stronger the project management, the higher the cost to benefits of the project. The effectiveness of the project management is determined by the various aspects which can be classified as:

- a. Financial aspects: Financial aspects are those measures which show the variation in Return on Investment (ROI), Productivity, Cost savings, Earnings and cash flow per share, Economic value addition, growth of sales in terms of both increase in volumes of product and the revenue inflow.
- b. Customer aspects: Customer is treated to be the highly prioritized person in the entire project development process. He gains utmost importance among all the people involved in a project. Customer satisfaction, his profitability, customer retention, acquisition of new customers, market share of customer's and his use.
- c. Project or Process aspects: The success of project does not just depend on the financial aspects and customer aspects, but are even dependent on Project and process measures. Project performance, projects risk, etc.
- d. Employee aspects: Successful delivery of any project lies also with employees involved in respective projects. Employee Buy-in is the most essential aspect that shall definitely lead to successfully completing the project leading to timely delivery and 100% customer satisfaction. The growth of any company lies with few more aspects like employee productivity, empowerment, employee's turnover and motivation.

CONCLUSION

The increase in investments of industries on analytics based projects in software industries signifies that most of the sectors are gearing up predicting the risk well in advance to reduce the loss and also look for increase in profits. As per the statistics revealed by various organizations like Gartner, NVP etc., it is observed that the failure rate of big data based analytics projects is more than 85% which is impacting on the economy of the companies investing on big data projects. There is need for an efficient model which can lead to successful delivery of the big data projects with higher prediction accuracy.

The proposed comprehensive model shall help in driving to successful project development with higher customer retention ratio and reliable tool. The proposed model is based on certain industrial practices and had opted for novel aspects.

The overall project implementation framework has been suggested for the companies investing on big data analytics projects. However, the project management strategy is left for the future scope. The suggested model is being experimented along with the novice PM methodology for future work.

REFERENCES

1. Bagriyanik S, et al., Big data in software engineering: A systematic literature review. J Glob Inf Technol Manag. 2016; 6:107-116.
2. B. W. Boehm. Software Engineering Economics, Prentice Hall, Englewood Cliffs, NJ, USA. 1981.
3. Putnam L.H. A general empirical solution to the macro software sizing and estimating problem. IEEE TRANS. SOFTW. ENG. IEEE T SOFTWARE ENG.1978; 4:345-361. [Crossref][Google Scholar]
4. Reihaneh H, et al. Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data. 2019; 6:44.

5. Feras A, et al. Predicting failures in agile software development through data analytics. *Softw Qual J.* 2018; 26:49-66.
6. Vikram S, et al. Project Analytics to Improve Project and Portfolio Decision Making, Project Management National Conference, India. 2017.
7. Mikalef P, et al. Big data analytics capabilities: A systematic literature review and research agenda. *INF SYST E-BUS MANAG.* 2018; 16:547-578.
8. Thurber M. A Holistic Framework for managing data analytics projects. Whitepaper. 2017.
9. Premjith BPK. How to make an Agile Teamwork for Big Data Analytics, Whitepaper. 2017.
10. Yonatan H. My best tips for Agile Data science research, Whitepaper. 2018.
11. Ori Cohen Data-Science? Agile? Cycles? My method for managing data-science projects in the Hi-tech industry, Whitepaper. 2018.
12. Lewis M. Project Management Methodologies for Big Data Analytics, Whitepaper. 2019.
13. Jiwat. What roles does Big Data have in shaping the future of Project Management? (Part-A), IPMA. 2017.
14. Jiwat Project Eco-system: Designing future developments through Big Data analytics (Part-B), IPMA. 2017.
15. Michael Bloch, et al., Delivering large-scale IT projects on time, on budget, and on value. *McKinsey on Business Technology.* 2012; 27:1-7.
16. Max Henrion. Why most of the Big Data analytics projects fail, *ORMS Today.* 2019.
17. Thomas H, et al. Big Data Executive Survey 2017, New Vantage Partners LLC. 2017.
18. Hassan N, et al. Predicting Software Projects Cost Estimation, Based on Mining Historical Data. *ISRN Software Engineering.* 2012; 1-8.
19. V R Basili, et al., Software Errors and complexity: an empirical investigation. *Commun. ACM.* 1984;27:42-52.
20. Halstead M H *Elements of Software Science.* Elsevier North-Holland. 1977.
21. Chidember SR, et al. Metrics suite for object-oriented design. *IEEE TRANS. SOFTW. ENG. IEEE T SOFTWARE ENG.* 1994; 20:476-493.
22. Stefan Studer, et al. Towards CRISP-ML (Q): A Machine Learning Process Model with Quality Assurance Methodology. *CoRR abs/2003.05155 (2020), arXiv: 2003.05155.*
23. McCabe T.J. A Complexity measure. *IEEE TRANS. SOFTW. ENG. IEEE T SOFTWARE ENG.* 1976; 2:308-320.
24. Fei YY, et al. Improvements about Halstead model in software science. *J Comput Appl.* 2004; 130-132.
25. Kosarenko Y. The majority of business analytics and AI projects are still failing, in data driven investors. 2020.
26. <https://www.datadriveninvestor.com/2020/04/30/the-majority-of-business-analytics-and-ai-projects-are-still-failing/#>.
27. Zicari RV, et al., Setting Up a Big Data Project: Challenges, Opportunities, Technologies, and Optimization. In: Emrouznejad A. (eds.) *Big Data Optimization: Recent Developments and Challenges. Studies in Big Data.* 2016:18.
28. Center for Business Practices Measures of Project Management Performance and Value, A Benchmark of Current Business Practices. 2005.