

Big Data Platform for Enterprise project management digitization using Machine learning

Ruchi S,
Assistant Professor,
Information Technology Department
NMIMS MPSTME
Mumbai
ruchik.2020@gmail.com

Dr Pravin Srinath,
Professor and HOD,
Computer Department,
NMIMS MPSTME,
Mumbai

Abstract—In recent years, enterprise related data is becoming increasingly digitized. Due to the exponential volume of enterprise data being generated, there is an increased demand in managing this data for efficient decision making. Data Mining and machine learning aids in achieving valuable insight of a business and its growth through the exploration of data by recognizing business project relationships and dependencies. Enterprise project data is a subset of enterprise data being created online or offline. This paper presents the proposed architectural design and related concept explanation for efficient enterprise project management in current data scenario. Also, an analysis of the current enterprise project techniques has been presented along with challenges that need to be addressed.

Keywords—business intelligence; enterprise project; data mining; big data

I. INTRODUCTION

According to PMI (Project Management Institute), pulse survey, the annual global survey of project management practitioners, implementing strategy successfully is inextricably linked to an organization's capability to deliver successful projects and programs [2].

A project is conventionally defined as a “temporary endeavor undertaken to create a unique product or service “[3]. Enterprise project management is a challenging task given the parameters it has to fulfill and the increased failure rate of successful project completion. This task has become increasingly difficult due to digitization of enterprise data and ever increasing volume and velocity of data generated.

The following contributions are made in this paper to achieve digitizing Enterprise Project Management using machine learning and Big data platform for effective Business intelligence:

- Design of EPM framework
- Proposed Architecture for EPM digitization using ML/BI.
- EPM parameters for project consideration
- Analysis of research gap for Big Data based framework adoption

The rest of the paper is organized as follows. Section 2 describes how machine learning can be leveraged for enterprise wide software projects and parameters we have chosen for project evaluation .Section 3 presents the proposed EPM (Enterprise Project Management) from the perspective of Machine Learning. The proposed architecture for digitizing EPM is presented in Section 4 along with its working. Section 5 identifies and analyses the research gap found during course of this research. Finally, conclusion and future work is explained in Section 6.

II. MACHINE LEARNING FOR ENTERPRISE PROJECTS

A decision making system can be made efficient, reliable and robust for enterprise projects only if it is able to detect certain trends related to enterprise planning, resource management and cost estimation by applying data analytics.

1) Industry Research Analytics:

Machine Learning (ML) and data mining (DM) concepts are jointly being applied for research within an industry. This ML/DM combination has become a major research hotspot in recent years.

Research is related primarily in the following sub domains:

- a) Customer identification and classification[7]
- b) Product analysis
- c) Project Feedback Analysis

2) Enterprise Project Management:

Our paper focuses on this sub domain within industry analytics.ML/DM can be used for application of project management practices in a standardized manner. Few characteristics for project evaluation are implemented by separate project tool for each parameter. In [8], cost estimation evaluation is done using regression technique .However the data taken as input is in structured format which may not always be the case especially for software projects in IT industries. Research in [14] introduces a metamodel for software expenditure computing the baseline relationship between and size. In [9], project evaluation is done on construction project having quantitative data

analysis. The authors in research [15] have used GRNN for effort estimation and prediction of software projects. Other machine learning applications using support vectors are described in [10].

Survey conducted in [13] demonstrates the increasing demand of data science and analysis among software engineers in enterprises. Research in [12] highlights the issues that influence the productivity levels of the project teams involved

Enterprise Projects are becoming more complex, making them more difficult to control without a formal management structure. Projects with substantially different characteristics, IT, are emerging.

Project management helps cross-functional teams to become more effective. Also, companies are using project management to develop and test their future leaders [1].

5) Fraud Analysis

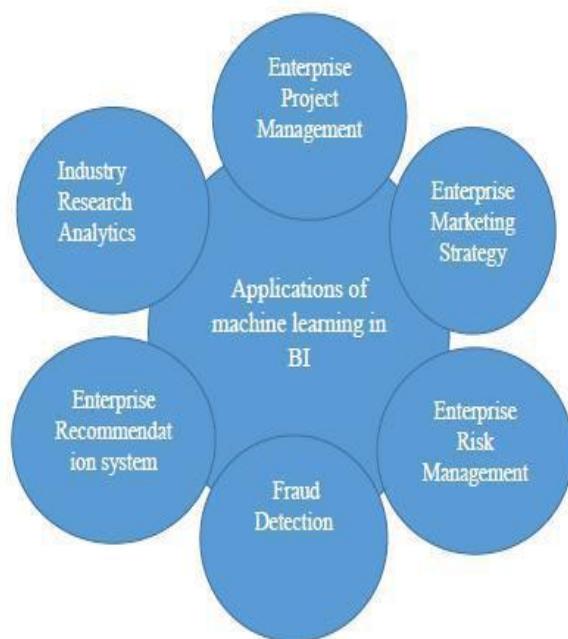


Fig. 1. Applications of Machine Learning for Business Intelligence

3) Enterprise Marketing Strategy:

Machine Learning Algorithms like SVM(Support Vector Machine) , DT(Decision Tree) , MVR(Multivariate regression analysis) are widely used techniques for formulating effective strategies for marketing in industry.

4) Enterprise Risk Management :

This application includes:

Identification of potential risks in enterprise to be able to effectively combat the risks proactively. The idea is to deal with only major and most evident enterprise risks and not all the risks.

Fraud Detection is a major ML/DM application mostly in financial organizations both in public and private sectors.

This includes:

1. Fraud Categorization
2. Fraud Detection[5]
3. Fraud Prediction
4. Prevention of Frauds
5. Enterprise Recommendation System

Machine learning techniques employed are reinforcement learning, deep learning and neural network explained in [6].

TABLE I. EPM PARAMETERS CONSIDERED

Sr no	EPM parameters	
	EPM	Project Management Aspect to be evaluated
1	EPM1	Project Scheduling
2	EPM2	Enterprise Resource Allocation
3	EPM3	Past Project Evaluation
4	EPM4	Project Cost Estimation
5	EPM5	Product opinion consideration of similar projects
6.	EPM6	Project Risk Analysis

Table 1 describes the EPM parameters that would be taken into account for ML/BI framework

Following are the parameters considered for this purpose:

- EPM1: Project Scheduling
- EPM2 : Enterprise Resource Allocation
- EPM3: Past Project Evaluation
- EPM4: Project Cost Estimation
- EPM5: Product opinion consideration of similar projects
- EPM6: Project Risk Analysis

III. ENTERPRISE PROJECT MANAGEMENT FROM MACHINE LEARNING PERSPECTIVE

Following is the brief description of proposed EPM framework as mentioned in Fig. 2:

1) Data Sensing:

This phase of EPM, the capturing of data is to be performed related to enterprises from multiple platforms and

programming environments .This data can be present on various architectures and computing sources.

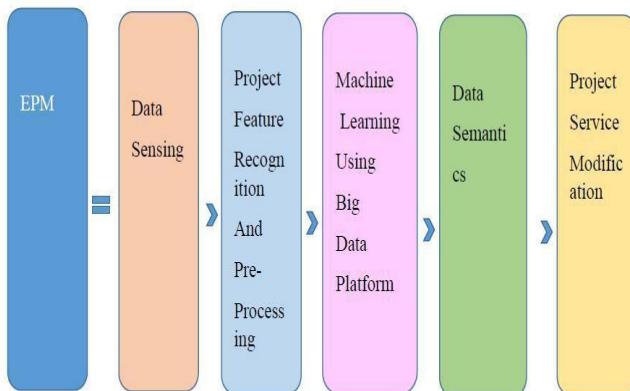


Fig. 2. Proposed EPM framework

2) Project Feature Recognition and Preprocessing:

a) In this phase, data captured from different sensing techniques would be analyzed carefully .Data would then be normalized and preprocessed to convert it into uniform format.

b) The next step in this phase is to recognize relevant enterprise project features by applying machine learning feature extraction algorithms onto processed data from the previous step.

3) Machine Learning using Big Data Platform:

There are two steps proposed in this phase .Step 1 would be to adapt Big Data Technologies like Hadoop , Hive, Pig, etc to cope with high velocity, high volume data generated within an organization or at the most within boundaries of partner firms only .

Second step is to formulate and apply machine learning algorithms to predict trends within project data generated before the product is available to the consumer.

There would multi fold benefits of this approach .Firstly, project data can be efficiently stored and computed. Moreover, forecasting of the trends would lead to huge decrease in losses of the concerned industry.

4) Data Semantics:

In this phase, projects related data would be analyzed using semantic ontology specifically developed for organization belonging to a particular domain.

This can be done on Big Data or cloud environment applying business services OWL- S (Ontology Web

Language –Semantic) and BPEL (Business Process Execution Language) for service composition.

5) Project Service Modification:

In this phase, enterprise services would be modified depending on 2 factors:

a) Output of data semantic phase and machine learning application phase.

b) Changing market scenario leading to corresponding change in enterprise services even though project process and initial requirement specification remains the same.

IV. PROPOSED METHOD

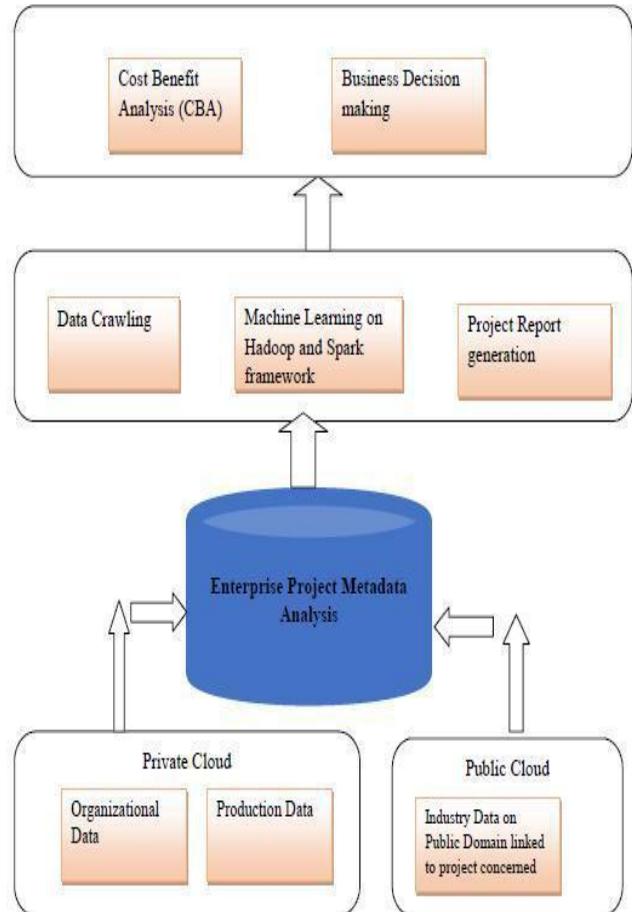


Fig. 3. Proposed architecture for EPM using machine learning

Fig 3 describes the EPM architecture proposed for machine learning. The platform considers the project data deployed on 2 platforms namely public cloud and private cloud.

Phase 1:

The project data sources considered would be production data, Organizational data extracted from transactional logs from within the enterprise. Project related data is scattered online that needs to be extracted for effective EPM. Hence, this project data available hosted on public cloud model from a pool of online resources would be handled by virtual machines (VM) across diverse data platforms.

Phase 2:

This phase consists of three components: Data crawling for aggregating scattered data relevant for current project being implemented. The output of data crawling module would be input for Machine learning on Hadoop and Spark platform. Finally, last component would be the project report generation applying business visualization techniques.

Phase 3:

There are two components in this phase : Cost Benefit Analysis(CBA).Here, the EPM parameters discussed in Section II would be analyzed for result calculation .The second component is Business Decision Making implemented using MCDA(Multi criteria Decision Analysis) in Fuzzy approach and neural network computational analysis depending upon the criterion spaces.

The proposed architectural platform will be beneficial to both technical team and management team for joint decision making.

V. GAP ANALYSIS

Following research gaps that have been identified in this sub domain:

1. Effective standardization of project management principles across enterprises as specified by PMI and SEI(Software Engineering Institute)
2. Lack of integration of ML/DM techniques with existing project management framework.
3. Computational Complexity of certain ML algorithms like CNN (Convolutional Neural Network) making it difficult to be applied to EPM.
4. Domain dependence exists for available techniques applied to enterprise projects making EPM platform complex to implement.
5. Effective training for employees to implement integrated framework
6. Resistance of adoption of integrated platform for project management.

VI. CONCLUSION

In this paper, we have proposed an EPM platform using Big Data and machine learning concepts. The advantage of this architecture is the effective digitization of project related information and transactional logs in various formats available both within and outside the firm concerned. Standard project management tools are standalone softwares focusing on few of project management parameters .We have also formulated six EPM parameters for project evaluation using Big Data platform. Also, this paper identifies and addresses research gaps of EPM in detail. This platform would mitigate the gap between the different stakeholders involved in development of the project.

REFERENCES

- [1] Nicholas G. Hall," project management: recent developments and research challenges", Journal of Systems Science and Systems Engineering, Springer, June 2012.
- [2] PMI: The High Cost of Low Performance | Pulse of the Profession 2016-<http://www.pmi.org/-/media/pmi/documents/public/pdf/learning/thought-leadership/pulse/pulse-of-the-profession-2016.pdf>
- [3] Project Management Institute, "A guide to the project management body of knowledge",4th Edition
- [4] Ammar Alsaig, Alaa Alsaig, Mubarak Mohammad," Context-based project management",5th International conference ICCASA ,Springer,2016
- [5] Abdelali El Bouchti; Ahmed Chakroun; Hassan Abbar; Chafik Okar, "Fraud Detection using deep reinforcement learning", Seventh International Conference on Innovative Computing Technology (INTECH), 2017
- [6] Aji Mubalaike Mubarek; Eşref Adalı,"Multilayer Perceptron neural Network for Fraud detection", International Conference on Computer Science and Engineering (UBMK), 2017
- [7] T.K. Das, "A customer classification prediction model based on machine learning techniques, International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2015
- [8] Sphurti, S. Arage, Nagaraj V. Dharwadkar, "Cost estimation of civil construction projects using machine learning paradigm", International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017.
- [9] Min-Yuan Cheng, Hsing-Chih Tsai, Erick Sudjono "Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry" Expert Systems with Applications, 37 : 4224–4231, 2010
- [10] K. Iwata, T. Nakashima, Y. Anan, and N. Ishii, "Estimating interval of the number of errors for embedded software development projects," International Journal of Software Innovation (IJSI), vol. 2, no. 3, pp. 40–50, 2014.
- [11] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," Statistics and Computing, vol. 14, no. 3, pp. 199–222, Aug. 2004. [Online]. Available: <http://dx.doi.org/10.1023/B:STCO.0000035301.49549.88>
- [12] Z. Jiang, C. Comstock, "The Factors Significant to Software Development Productivity". World Academy of Science, Engineering and Technology, pp. 160 – 164, 2007.
- [13] J. W. Bailey and V. R. Basili, "A meta-model for software development resource expenditures," in Proceedings of the 5th International Conference on Software Engineering, ser. ICSE '81.

Piscataway, NJ, USA: IEEE Press, 1981, pp. 107–116. [Online].
Available: <http://dl.acm.org/citation.cfm?id=800078.802522>

- [14] C. Lopez-Martin, A. Chavoya-Pena, M. E. MedaCampana, "Software Development Effort Estimation in Academic Environments Applying a General Regression Neural Network Involving Size and People Factors", Lecture Notes in Computer Science Volume 6718, Springer, DOI: 10.1007/978-3-642-21587-2_29, 2011