Software Analytics: A must for every organization

A brief report of importance of software Analytics in an organization

Joydeep
IIIT,Bangalore
joy.deep@iiitb.org

JayPrakash L T
IIIT,Bangalore
jtl@iiitb.org

***Abstract*—In software development life cycle, a huge amount of data is generated. In spite of such huge potentially powerful data which could give deep insights about the company and its employees, there are less recognized tools to do that. In this paper, we present what are the advantages and disadvantages of software analytics in an organization indulged in software development.**

This is a report made after referring many papers published in the field of software analytics. The type of paper referred were mostly implementation oriented.

***Index Terms*—Data, Software Analytics, Software development life cycle.** *(key words)*

I. Introduction

Software development life cycle includes artifacts like source code, bug reports, check-in histories, test cases etc, contains knowledge about the organization and its employees. If those artifacts are analyzed in the best possible way, questions like what kind of projects are well suited for the organization, strengths and weaknesses of each employees, etc.

Software Analytics[1] is to utilize the data-driven approach to enable software practitioners to perform data exploration and analysis in order to obtain insightful and actionable information for completing various tasks around software systems, software users, and software development process.

Most organizations don't invest in software analytics because they are small[8]. The objective of software analytics is to obtain *insightful* and *actionable* information from software artifacts to help software practitioners accomplish their targets around software systems, software users, and software development process[software analytics in practice]. In other words, software analytics aims to improve software systems in terms of quality, software users in terms of experience and software development process in terms of productivity.

II. Why Analytics should be brought in every company?

In a small organization, the best intuitive model of software development is code and fix. This model had no guidelines regarding requirement analysis, modularizations, and presence of many software artifacts which enhance the quality of software, which leads to delay in delivery, change in requirements etc.

In an average sized company it is important to analyze the loopholes in software development and take proper actions on the information, like more concrete POC's or alternatives towards completing the task.

In a large organization, it's important to be updated with respect to domain of functionality of the company and quality of software systems and user experience.

So, from the above we can say that every company needs to focus on the following

- Long term strategic decisions.
- Short term tactical decisions.
- More transparency in software development process.[amisoft].

III. Procedure

A. Data sources

- Time sheets are one of the most important artifacts during software development. This artifact could be used to analyze the current status as well as predict the amount of time required to finish a project.
- Requirements are one of the important pillars in software development life cycle and should be analyzed properly because it gives the objective of the project. Due to lack of proper requirement analysis, an organization may lead to unsatisfied customer, delay in project delivery, extra cost etc.
- Emails plays vital role of communication in software development especially when teams working on a project have different locations. Emails can also predict the social status of the developers.
- Source code should be analyzed in terms of pattern or anti pattern. Focus should be on functionality and style of code, for example, proper comments at proper position makes readability and bug trace easier.
- Client review document analysis should be done with utmost care. This artifact could help the company to be in sync with the client's requirements. This document could be utilized to get insights of client behavior and organization's

functioning related to requirements specified earlier.
- Other artifacts like design document, test case document, structural document needs to be analyzed for in depth analysis of the insights.
- Test management tools like TestLink can be used during development for management and quality purpose.

These artifacts can be human generated, application generated or system generated and should be preprocessed to do analysis.

B. Metrics

After data collection metrics needs to be defined and targets are defined with respect to them [amisoft]. EVM could be one of the metrics. These metrics may vary depending upon the type and size of organization. Few additional metrics are

- Requirement volatility is defined as the ratio of new, modified, or deleted requirements to total number of requirements.
- Software events
- Human Resource
- Email mining

Along with the metrics, the organization should also decide upon the metrics that are not to be collected. This is important for the organization to focus on the desired goal.

C. Data Analysis

After deciding upon the data sources and metrics, data analysis algorithms like hunt's algorithm for classification, or statistics principles can be applied to get human recognizable pattern and these patterns can be used to obtain knowledge about the data produced during software development. The data can be used for predictive or descriptive analysis.

IV. Benefits of Software analytics

The objective of software analytics is to obtain insightful and actionable information from software artifacts to help software practitioners accomplish their target tasks around software systems, software users and software development process[software analy in prac]. Insightful information provides hidden valuable information about the task from which the data has been generated. In general, it's very difficult to deduce human recognizable patterns from such huge data without the use of analytical technologies. For example work load on a particular type of employees who have similar skills set. Actionable information provides information on particular areas where enhancements are required. For example, one module in software is not performing as it is meant to be, so software practitioners should come up with alternative solutions or better proof of concept.

If software artifacts are analyzed from the insightful information and proper actions are taken on the actionable information, the quality of the software enhances. For example, if scheduling, requirement volatility, verifications and validations are analyzed properly and actionable decisions are performed, development will not incur extra cost due to these factors.

Since the artifacts like test case results will be analyzed for insightful information and actionable information, the performance of the software will be more justified.

Client interaction is one of the software artifacts that can enhance the software under development in terms of customer experience. The requirement volatility will be more stable and the software will be more specific as per guidelines provided during the customer interactions. We have also discussed

V. Cost of Software analytics

Though software analytics helps in enhancing quality in terms of software system, experience in terms of Software experience and productivity in terms of software development process, there is a significant amount of cost associated with it.

For software analytics, software artifacts need to be collected for analysis. In an organization, collecting data for time sheets requires time of each employee. These time sheets are then reviewed by manager of the employee. This requires time of the manager. The time sheets produced by employees of a project are analyzed by the data analyst to check for a pattern and the results are forwarded to governing authority. This requires time of data analyst. All these are the cost associated in an organization where software analysis has been already implemented.

For an organization will to adopt software analytics techniques need to have a metric analyst to determine metrics to be collected?

VI. conclusion

Software analytics is a data driven approach that uses software artifacts to get insightful and actionable information which would help the practitioners to complete a specific task in software development process. In this paper we have presented the requirement of software analytics in an organization indulged in software development process. We have also discussed the procedure of achieving it and its benefits. At the end, we have presented the costs involved in adopting the technique.

References

[1] R. Robbes, R. Vidal, M. C. Bastarrica, "Are Software Analytics Efforts Worthwhile for Small Companies? The Case of

Amisoft", *IEEE Software*, vol.30, no. 5, pp. 46-53, Sept.-Oct. 2013, doi:10.1109/MS.2013.92

[2] Dongmei Zhang, Shi Han, Yingnong Dang, Jian-Guang Lou, Haidong Zhang, Tao Xie, "Software Analytics in Practice", *IEEE Software*, vol.30, no. 5, pp. 30-37, Sept.-Oct. 2013, doi:10.1109/MS.2013.94

[3] C. Bird, A. Gourley, P. Devanbu, et al. (2006). Mining email social networks, Proceedings of the 2006 international workshop on Mining software repositories, Shanghai, China, 137-143.

[4] Mike Harris, “What are “Software Analytics” and who should care about them”