

Female Student Participation in Software Engineering Projects: Opportunities to Model Project Evaluation and to Improve Early Prediction of Teamwork Failure

Ismail Al-Taharwa

Department of Computer Information Systems, The University of Jordan
77110 Aqaba, Jordan

Abstract

Software engineering project is the preferred mean to measure competency and practical skills among learners in IT-fields. This study investigates the opportunities to model software engineering project final evaluation and improve early prediction of academic software engineering project failure by considering female student participation as a teamwork member, regardless of being a teamwork leader or a teamwork regular member. Four distinct arrangements of software engineering development teamwork are advised and studied. Those arrangements range from female-less participation teamwork to female-dominated participation teamwork. Machine learning techniques are leveraged to build prediction models. Teams are evaluated from two distinct perspectives. First, software products submitted at the end of each project life cycle milestone, namely product perspective. Second, the degree of obeying the good practices of software engineering project development, namely process perspective. Results reveal significant differences due to female student participation. Arrangement of female-less participation attains the worst modeling and prediction performance compared to the other arrangements of female student participation.

Keywords: Gender diversity, E-learning, Software engineering, Project failure, Machine learning

DOI: 10.7176/JEP/11-35-06

Publication date: December 31st 2020

1. Introduction

Software engineering is a branch of computer science that includes the design, development, evaluation, and implementation of computer software (Page and Six, 2017; Sommerville, 2015). The covid-19 pandemic caused extra suffering for software engineering students, particularly software engineering project teamwork. Students struggle to form teamwork, to communicate with each other, to follow up with their instructor, to submit the project's deliverable on time, and to adhere to the guidelines and comments made by their instructor. Teamwork members may ask to change the project's idea, miss submitting deliverable on time, or even could not submit the project's final product. Dispute may happen among team members and may be developed into a kind of enmity. A teamwork member may leave the teamwork or even ask to reform the teamwork. Some teamwork may ask to change the teamwork leader or even the teamwork supervisor (Kennedy and Vossen, 2017; Dhir, Kumar and Singh, 2019). All such issues shed the light on the pressing demand to improve early prediction of academic software engineering project failure.

This research studies the issue of female student participation in academic software engineering projects for the undergraduate level. Real-world software engineering projects reflect a pronounced variation among female participation. The industrial sector of information technology and software engineering lack female diversity, even when considering leading technology companies (Google, 2020). Literature attributes such imbalance in gender representation to the inherent divergence in undergraduate study (Buquet, 2011). Educators advise that seeking an equal representation of female students in undergraduate study will lead to a better representation of women in the information technology market (García-Holgado et al. 2018). The key objective of this study is to model software engineering project evaluation considering female student participation. Further, we investigate opportunities to improve the early prediction of software engineering projects by treating projects differently according to female student participation.

A recent publicly released data set collected by a dedicated e-learning system is utilized, SETAP data set (Petkovic et al. 2016). Projects are done in a teamwork fashion, such that each teamwork is formed by three to eight students. Male and female students can participate in the teamwork of their choice. Thereby, female student participation varies significantly among the different teams. Box and whisker plot distribution is utilized to categorize SETAP data into four distinct arrangements according to female student participation. Those arrangements are 1. Teams that do not include female students at all, namely Female-less participation (FLP). 2. Teams that are dominated by male students, namely male-dominated participation (MDP), 3. Teams that include moderate levels of female student participation, namely gender equality participation (GEP), and 4. Teams that are formed by female students majority, namely female-dominated participation (FDP). The four arrangements are modeled and evaluated using decision tree classifier. As the key objective is to improve the early prediction of

software engineering project failure, time intervals that cover the early stages of software prediction are emphasized. This paper tries to answer the following research questions:

- How does female student participation affect the modeling of software engineering teamwork evaluation?
- How does female student participation affect the early prediction of academic software engineering project failure?

The rest of this paper is organized as follows, Section 2 overview the background and related works. The methodology is elaborated in Section 3, Results are presented and discussed in Section 4 and Section 5, respectively. Finally, this paper is concluded in Section 6.

2. Background and Related Work

Academic software engineering projects are increasingly gaining attention as a reliable method to assess competency levels among ICT students (Raibulet and Fontana, 2018; Ju and Fox, 2018). In (Kennedy and Vossen, 2017), the authors considered a student peer assessment scale in addition to the lecturer's overall scale to enhance the reliability of the evaluation. Quite recent studies shed the light on the importance of incorporating maintained practices across the project development process in addition to the final product (Ju and Fox, 2018; Frezza, Daniels and Wilkin, 2019). The principal purpose is to quantify to which extent do students pay attention to the good practices of software engineering. In (Yelmo and Fernández-Corugedo, 2011), authors go beyond team evaluation to evaluate the whole educational process through an advised collaborative environment for software engineering project management.

Statistical hypotheses utilizing both qualitative and quantitative data have been proposed and evaluated to enhance software engineering project's success opportunities. A quite recent study investigated the gap between the software industry and software engineering education (Oguz and Oguz, 2019). Frezza, Daniels and Wilkin (2019) attempted to assess how information technology students obey the professional values of software engineering in a global project setting. Also, (Dhir, Kumar and Singh 2019) investigated success and failure factors among academic software engineering projects. A collaboration of multiple academic institutions investigated the issue of early prediction of software engineering projects from different perspectives (Petkovic et al. 2014; Petkovic et al. 2016). Follow up research explored the opportunities to improve early prediction of academic software engineering project failure by considering teamwork distribution (Petkovic et al. 2018; Al-Taharwa, 2020).

Gender-diversity is a desirable characteristic of the working environment. Catolino et al. (2019) investigated the actual benefits of achieving gender-diversity in software engineering project teamwork. Recent studies reveal the severe situation when considering information technology professions including software engineering (Burnett et al. 2016; Vorvoreanu et al. 2019; Wang and Redmiles, 2019). Google's diversity report for 2020 illustrates the existing shortage in female IT specialist representatives (Google, 2020). Research literature emphasis that a better understanding of gender mismatch in undergraduate study is desirable to bridge the gap in industrial sectors (Buquet Corleto, 2011; García-Holgado et al. 2018). Hirshfield and Koretsky studied the discourse among student teams working in an engineering problem-based learning (PBL) environment (Hirshfield and Koretsky, 2018). Surprisingly, they found that gender-status and team makeup has no explicit relation with team participation. But, they noticed that female students were more responsive in answering non-technical questions than male students. Similarly, an empirical study of female participation in software engineering projects found that female students were the best in leading project management, and requirement related tasks (Nguyen-Duc, Jaccheri and Abrahamsson, 2019).

All aforementioned studies tackled the issue of gender diversity from different perspectives. However, the literature lacks a comprehensive investigation that connects the dots of online collaboration and gender diversity among academic software engineering projects. Petkovic et al. (2014) proposed and constructed a dedicated online environment for project development, namely SETAP. Their system considers characteristics of each deliverable handed during the project life cycle, namely product characteristics. Also, it considers team activity measures (TAM) such as how teamwork members stick to the good practices of software engineering, namely process characteristics. Interestingly, the SETAP platform records the representation of female students in teamwork utilizing the percentage of female students participating in each teamwork, namely *femaleTeamMembersPercent*. We leverage this characteristic to model the project's final evaluation and to improve the early prediction of academic software engineering project teamwork failure.

3. Methodology

In this study, a real-world data set collected by a dedicated online environment is utilized (Petkovic et al. 2016). Pre-processing described by Al-Taharwa (2020) is undertaken first. Box-and-whisker plot (Thirumalai, Vignesh and Balaji, 2017) is leveraged to categorize the data into groups of teamwork that reflect variant female student participation arrangements. Decision trees technique of machine learning is applied to each arrangement of female student participation. Both process and product perspectives are considered.

3.1 Data Categorization

Female student participation depicts female student representation in academic software engineering teamwork. As software engineering project teamwork varies in size, the ratio of female student members to the overall number of teamwork members is chosen as the criteria to categorize teamwork into different groups. Interestingly, SETAP data exhibits extremely varying representations of female student representation. Some teams do not include female student members at all, while other teams include varying ranges of female student participants. To attain an objective categorization of the SETAP data set according to female student participation, box-and-whisker plot is utilized as a reference instrument for categorization. Categorization by box-and-whisker plot reflected a skewed representation towards the groups of low female student participation. Categories of low female student participation were larger than categories of moderate and dominated female student participation. Additionally, some arrangements of female student participation exhibit an imbalanced representation of teamwork evaluation. Minor changes are incurred to guarantee better representation among all considered categories of female student participation. Specifications of the final data categorization according to female student participation, and the corresponding teamwork evaluation are shown in Table 1 and Figure 1 respectively.

Table 1. Arrangements of female student participation in software engineering projects teamwork

Female student ratio (FSR)	Size (Teams)	Label
$FSR = 0.0$	24	Female-less participation (FLP)
$0.0 < FSR \leq 0.1667$	20	Male-dominated participation (MDP)
$0.1667 < FSR \leq 0.25$	14	Gender equality participation (GEP)
$0.25 < FSR < 1.0$	16	Female-dominated participation (FDP)

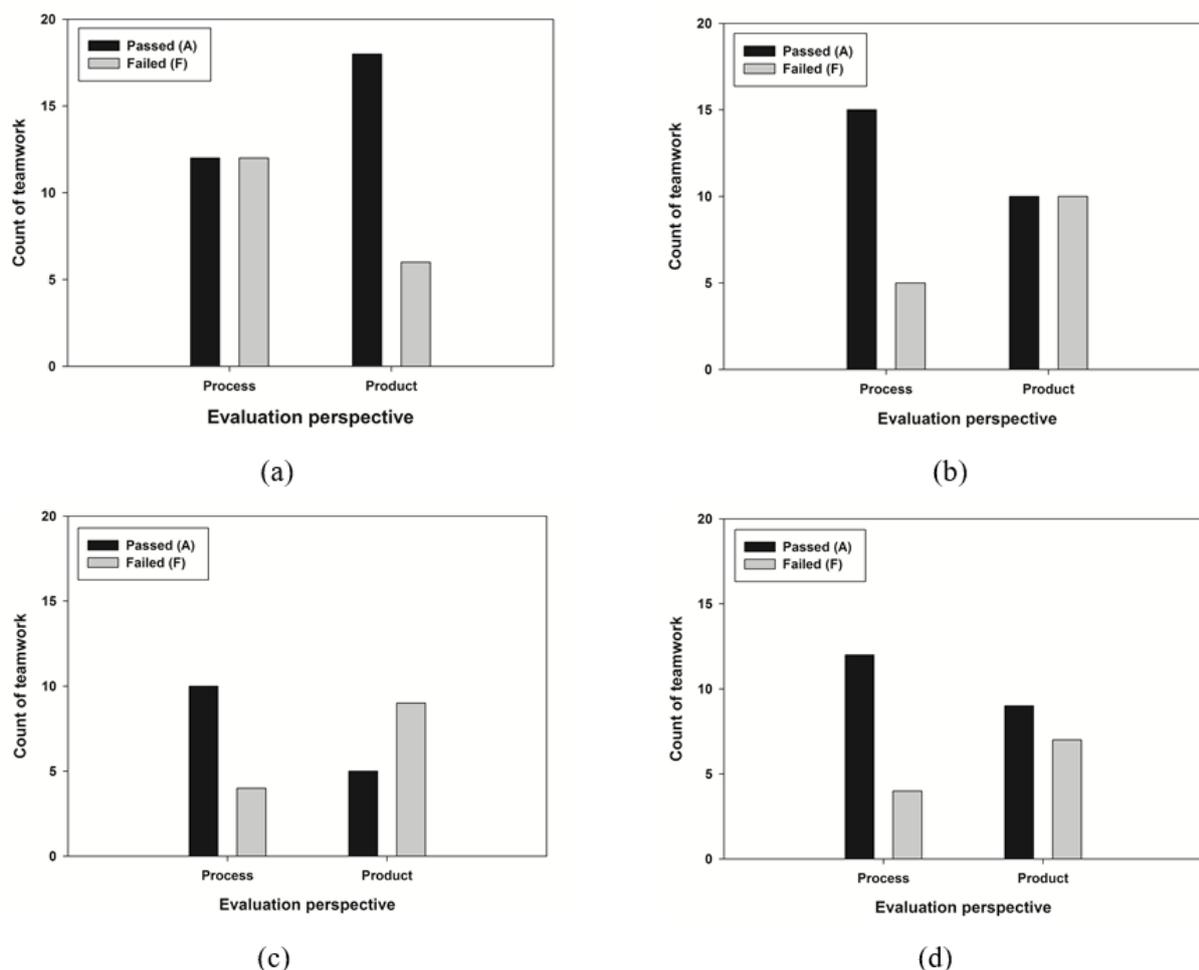


Figure 1. Evaluation of software engineering project teamwork considering arrangements of female student participation: (A) Female-less participation (FLP), (b) Male-dominated participation (MDP), (c) Gender equality participation (GEP), (d) Female-dominated participation (FDP)

3.2 Decision Trees

A decision tree is a tree-like decision support tool that enables analysis of decision alternatives when dealing with

a high number of decision criteria and alternatives. Usually, decision trees may expand vertically and horizontally as the number and values of the considered features increase and vary. Once a decision tree model is generated, Data mining tools utilize a rule-based system to classify unseen cases. It is very common to end up with a very long deeply nested rule-based system. Logical conjunctions are utilized to join multiple conditions in order to simplify rules. Tree pruning is a desired alternative to keep top-level highly discriminating features in order to minimize decision trees. Thereby, it results in considerably more concise and simplified rule-based systems. Additionally, it is an effective approach to avoid overfitting models (Safavian and Landgrebe, 1991).

4. Experiments and Results

This section presents the experimental setup and the corresponding results. J48, Weka data mining tool implementation of decision tree classifier is maintained in all experiments (Hall et al. 2009). In order to assert the unbiased performance of prediction models, 10-folds cross-validation is applied. Results are investigated from two perspectives. First, modeling of software engineering teamwork evaluation. Second, early prediction of software engineering teamwork failure. Complying with the nature of the problem, i.e., binary classification of software engineering projects final evaluation, either pass (A) or fail (F), set of failure projects, i.e., class F, considered as the positive class. Consequently, all reported results are in terms of positive class in order to emphasize the ability to predict failure among software engineering projects. Same experimental setups are applied to the four arrangements of female student participation.

4.1 Modeling of Software Engineering Teamwork Evaluation

As the key motive behind this research is to promote modeling of teamwork evaluation and prediction of project failure as early as possible, earlier stages of development are emphasized here. Considering process perspective, the second time interval (ProcessT2) is investigated. Considering product perspective, the third time interval (ProductT3) is investigated. Teamwork members start to submit code-deliverable, i.e., software products, at the third time interval. The first two time intervals are reserved for non-code deliverable, i.e., documentation, restrictively. Tables 2, and 3 compare the results of prediction models for the four different arrangements of female student participation (FSP) in terms of confusion matrix (Powers, 2007) from process and product perspectives respectively.

Table 2. Results of Software Engineering Project evaluation: Process Perspective, Second Time Interval (ProcessT2)

FSP arrangement	FLP		MDP		GEP		FDP	
Actual \ Predicted	A	F	A	F	A	F	A	F
A	8	4	13	2	8	2	12	0
F	4	8	1	4	1	3	3	1
Accuracy	0.667		0.85		0.786		0.813	

Table 3. Results of Software Engineering Project evaluation: Product Perspective, Third Time interval (ProductT3)

FSP arrangement	FLP		MDP		GEP		FDP	
Actual \ Predicted	A	F	A	F	A	F	A	F
A	11	7	7	3	3	2	5	4
F	3	3	2	8	0	9	1	6
Accuracy	0.584		0.75		0.857		0.688	

Accuracy is the preferred metric to measure the appropriateness of the prediction model as all types of classification errors are considered, Eq. 1. According to the results shown in Tables 2, and 3, the Female-less participation (FLP) arrangement was the worst in terms of accuracy. It attained 0.667 and 0.584 from process and product perspectives respectively. These results indicate that heterogeneous arrangements of female student participation, i.e., include both female and male students, outperforms female-less teamwork in terms of modeling software engineering project final evaluation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

4.2 Early Prediction of Software Engineering Project Failure

Recall metric is an alternative measure that provides a more comprehensive explanation in terms of false-negative errors as illustrated in Eq. 2. The higher the recall rate the lower number of failed projects wrongly classified as passed ones. Figure 2 compares the performance of the project failure prediction models in terms of recall rate. Interestingly, these results indicate that prediction models of product perspective are better than those of process perspective except for the FLP arrangement of female student participation. Additionally, the arrangements of female student engagement outperform the arrangement of female-less participation except the FDP when considering process perspective.

$$\text{Recall Rate} = \frac{TP}{TP + FN} \quad (2)$$

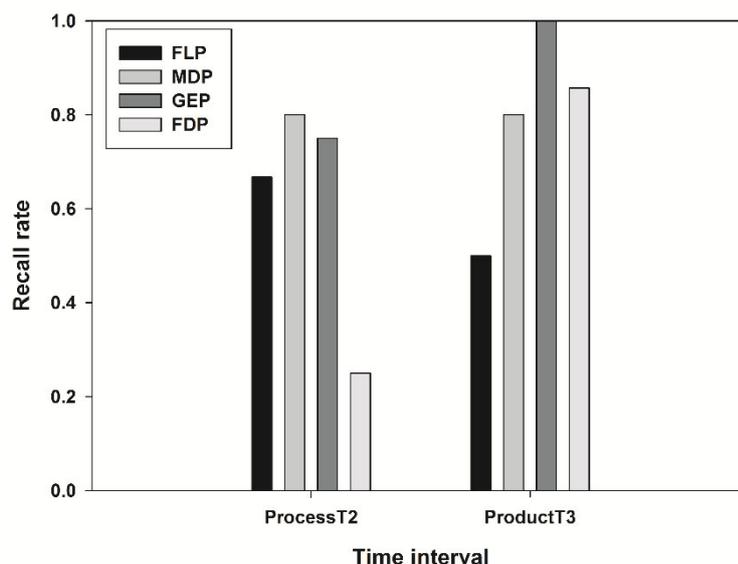


Figure 2. Comparison of female student participation models in terms of project failure prediction considering process (processT2) and product (productT6) perspectives

5. Discussion

Results presented in Section 4.1 indicate a better modeling for teamwork that include female students compared to female-less participation arrangement. Considering the process perspective, the only explanation is attributed to the conventional roles of female students in software engineering projects. Female students are better at handling non-technical tasks, such as project scheduling, system designing, and progress documenting. Further, they pay more attention to the good practices of software engineering development compared to male students. Therefore, the missing of female students lead to quite heterogeneous patterns of teamwork behaviors across the arrangement of FLP. As shown in Figure 1, considering process perspective, the number of passed teamwork exceeds the number of failed ones significantly for all arrangements of female student participation except FLP arrangement.

Considering the product perspective, the relatively low modeling results of FLP arrangement is quite questionable. As male students are known to do better than female students in terms of technical issues of software engineering. This remarkable observation attributed to the relative variance in FLP arrangement labeling, passed projects accumulate three times of failed ones. This effect of an imbalanced data set is clear as the model of early prediction of project failure is considered. FLP arrangement attains the worst prediction results from the product perspective (i.e., ProductT3) as shown in Figure 2.

6. Conclusion

This study investigated the issue of female student participation in academic software engineering projects. A real-world released data set was utilized. Teams are categorized according to the ratio of female student participation into four distinct arrangements. Those arrangements range from female-less participation to female-dominated participation. Prediction models are built and evaluated considering two perspectives. First, modeling of teamwork final evaluation, i.e., either passed or failed. Second, early prediction of software engineering project failure. Interestingly, arrangements of female student participation that include female students outperform female-less participation arrangement in terms of modeling project's final evaluation from both process and product perspectives. Similarly, the female-less participation arrangement was the worst in terms of early prediction of software engineering project failure from the product perspective.

References

- 1 Page, C. and Six, H.W., (2017). Software engineering. *Larsen and Keller Education*.
- 2 Sommerville, I., (2015). Software Engineering. 10th. In *Book Software Engineering. 10th, Series Software Engineering*. Addison-Wesley.
- 3 Kennedy, I.G. and Vossen, P.H., (2017). Teamwork assessment and peerwise scoring: Combining process and product assessment. *Bildungsräume 2017*.
- 4 Dhir, S., Kumar, D. and Singh, V.B., (2019). Success and failure factors that impact on project implementation using agile software development methodology. In *Software Engineering* (pp. 647-654). Springer, Singapore.
- 5 Google diversity annual report (2020), *Google*, <https://diversity.google/commitments>
6. Buquet Corleto, A.G., (2011). Transversalización de la perspectiva de género en la educación superior: Problemas conceptuales y prácticos. *Perfiles educativos*, 33(SPE), pp.211-225.
7. García-Holgado, A., Mena, J., García-Peñalvo, F.J. and González, C., (2018), April. Inclusion of gender perspective in Computer Engineering careers: Elaboration of a questionnaire to assess the gender gap in Tertiary Education. In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1547-1554). IEEE.
8. Petkovic, D., Sosnick-Pérez, M., Okada, K., Todtenhoefer, R., Huang, S., Miglani, N. and Vigil, A., (2016), October. Using the random forest classifier to assess and predict student learning of software engineering teamwork. In *2016 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.
9. Raibulet, C. and Fontana, F.A., 2018. Collaborative and teamwork software development in an undergraduate software engineering course. *Journal of Systems and Software*, 144, pp.409-422.
10. Ju, A. and Fox, A., (2018), July. TEAMSCOPE: measuring software engineering processes with teamwork telemetry. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (pp. 123-128).
11. Frezza, S., Daniels, M. and Wilkin, A., (2019). Assessing students' IT professional values in a global project setting. *ACM Transactions on Computing Education (TOCE)*, 19(2), pp.1-34.
12. Yelmo, J.C. and Fernández-Corugedo, J., (2011), April. An experience of educational innovation for the collaborative learning in Software Engineering. In *2011 IEEE Global Engineering Education Conference (EDUCON)* (pp. 567-574). IEEE.
13. Oguz, D. and Oguz, K., (2019). Perspectives on the Gap Between the Software Industry and the Software Engineering Education. *IEEE Access*, 7, pp.117527-117543.
14. Petkovic, D., Sosnick-Pérez, M., Huang, S., Todtenhoefer, R., Okada, K., Arora, S., Sreenivasen, R., Flores, L. and Dubey, S., (2014), October. Setup: Software engineering teamwork assessment and prediction using machine learning. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1-8). IEEE.
15. Petkovic, D., Barlaskar, S.H., Yang, J. and Todtenhoefer, R., (2018), October. From Explaining How Random Forest Classifier Predicts Learning of Software Engineering Teamwork to Guidance for Educators. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.
16. Al-Taharwa, I., (2020). Teamwork Distribution: Local vs. Global Software Engineering Project Development Teamwork. *International Journal of Emerging Technologies in Learning (iJET)*, 15(18), pp.183-201.
17. Catolino, G., Palomba, F., Tamburri, D.A., Serebrenik, A. and Ferrucci, F., (2019), May. Gender diversity and women in software teams: How do they affect community smells?. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)* (pp. 11-20). IEEE.
18. Burnett, M., Stumpf, S., Macbeth, J., Makri, S., Beckwith, L., Kwan, I., Peters, A. and Jernigan, W., (2016). GenderMag: A method for evaluating software's gender inclusiveness. *Interacting with Computers*, 28(6), pp.760-787.
19. Vorvoreanu, M., Zhang, L., Huang, Y.H., Hilderbrand, C., Steine-Hanson, Z. and Burnett, M., (2019), May. From gender biases to gender-inclusive design: An empirical investigation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).
20. Wang, Y. and Redmiles, D., (2019), May. Implicit gender biases in professional software development: An empirical study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)* (pp. 1-10). IEEE.
21. Hirshfield, L. and Koretsky, M.D., (2018). Gender and participation in an engineering problem-based learning environment. *Interdisciplinary Journal of Problem-Based Learning*, 12(1), p.2.
22. Nguyen-Duc, A., Jaccheri, L. and Abrahamsson, P., (2019), May. An empirical study on female participation in software project courses. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)* (pp. 240-241). IEEE.
23. Thirumalai, C., Vignesh, M. and Balaji, R., (2017), April. Data analysis using Box and Whisker plot for Lung Cancer. In *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-6). IEEE.
24. Safavian, S.R. and Landgrebe, D., (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3), pp.660-674.

25. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H., 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), pp.10-18.
26. Powers, D., (2007). Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation (Tech. Rep.). *Adelaide, Australia*.