This appendix will be available online [7]. We included it as supplementary material for the convenience of the anonymous reviewers.

APPENDIX A
MODEL CONSTRUCTION

In this section, we elaborate on the details of our model construction process. Figure 1 shows an overview of our model construction process. The process can be broken down into following three steps:

1) **Correlation & Redundancy Analysis:** First, we use the Spearman rank correlation test to measure the correlation between factors and remove highly-correlated factors (using a cut-off value of 0.7 [3]–[6]). For each of the highly-correlated factors, we keep one factor in the model. Figure 2 shows the results of our correlation analysis, and the factors that were eventually used in the models. Then we apply a redundancy analysis using the **redun** function in the **rms** R package to remove redundant factors. Finally, we end up with three factors in the project bounty dimension, six factors in the issue report basic dimension, four factors in the issue report bounty dimension, and three factors in the backer experience dimension. Because factors which have a constant value do not contribute explanatory power to a logistic regression model, we remove factors which are constant within a project group when building the corresponding models. For example, we remove *P_B_usage_group* (i.e., bounty-usage frequency) for the first-timer models , the moderate models and the frequent models since *P_B_usage_group* is a constant value for these models. In addition, we remove all project bounty-related factors for the first-timer bounty-projects since for these projects the values of all project bounty-related factors are 0.
2) **Non-linear Term Allocation:** Similar to prior work [4], [6], we add non-linear terms (i.e., NL) in each model to capture the more complex relationship in the data by employing restricted cubic splines [2]. The non-linear factor will be assigned more degrees of freedom (i.e., D.F.). We calculated the Spearman multiple ρ^2 between the dependent factor and each explanatory factor to measure their non-linear relationship. If a factor has a higher ρ^2, it indicates that it has a higher chance of having a non-linear relationship with the dependent factor. We therefore assigned this factor more degrees of freedom. Figure 3 shows the Spearman multiple ρ^2 of the studied factors. By observing the rough clustering of the factors according to their ρ^2, we cluster the factors into four groups according to the Spearman multiple ρ^2 values. The factor marked by the blue diamond is assigned five degrees, factors

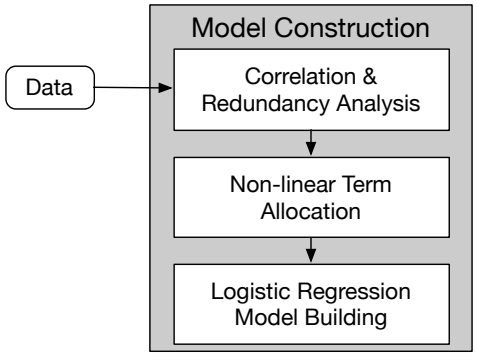

Fig. 1: The overview our model construction process.

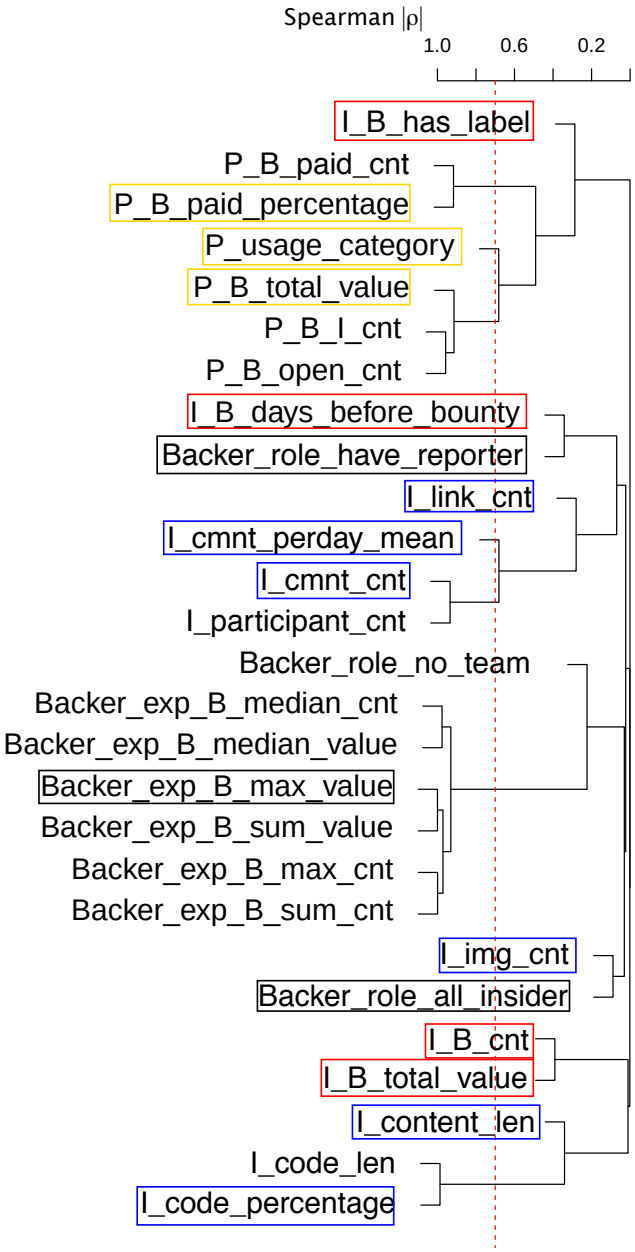

Fig. 2: The hierarchical clustering plot of factors according to the Spearman rank correlation test (using a cut-off value of 0.7). We selected the simplest metrics to compute across each dimension of correlated factors. We ended up with three factors in the project level dimension (marked in yellow), six factors in the issue report basic dimension (marked in blue), four factors in the issue report bounty dimension (marked in red), and three in the backer-related factors dimension (marked in black).

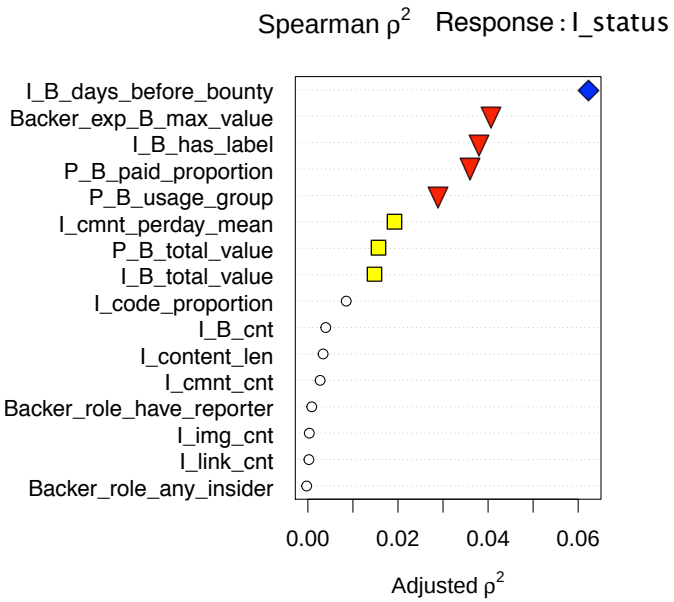

Fig. 3: Dotplot of the Spearman multiple ρ^2 of each factor in all the bounty issue reports. The larger the ρ^2 value, the more likely the factor has a non-linear relationship with the response variable. By observing the rough clustering of the factors according to their ρ^2, we cluster the factors into four groups according to the Spearman multiple ρ^2 values. We assign the first, second, and third groups of factors (categorized by the ρ^2 value) which are highlighted with a blue diamond, red triangle, and yellow square, 5, 4, and 3 degrees of freedom, respectively.

marked by red triangles are assigned four degrees and factors marked by yellow squares are assigned three degrees of freedom. We use the R package **rms**[1] to implement our logistic regression model.

3) **Logistic Regression Model Building:** Finally, we built four groups of logistic regression models (i.e., the first-timer, moderate, frequent, and global models) based on 100 samples and ended up with 400 models.

Appendix B
Model Analysis

In this section, we elaborate on the details of our model analysis process. The process includes two parts:

1) **Model Assessment:** For a logistic regression model, we use the Area Under the Receiver Operating Characteristic Curve (i.e., AUC) and a bootstrap-derived approach [1] to assess the explanatory power of the models following prior studies [3], [4], [6]. The AUC ranges from 0 to 1 (0.5 is the performance of a random guessing model) and a higher AUC means that the model has a higher ability to capture the relationships between the explanatory factors and the response factor.
For each sample, we use a bootstrap-derived approach [1] to validate the performance of models. We first train a model with a bootstrapped sample and calculate the AUC (i.e., the *bootstrapped_AUC*) on the bootstrapped sample. Then we apply the same model to the original sample and calculate the AUC (i.e., the *original_AUC*). After that, we use the optimism value, which is the difference between the *bootstrapped_AUC* and *original_AUC* to evaluate the degree of overfitting of the model. A small optimism value indicates that the model does not suffer from overfitting. We repeated the bootstrap-derived approach for 100 iterations for each sample and used the median *bootstrapped_AUC* and the median optimism value to represent the performance of models for that sample. Finally, we built 10,000 (100 samples * 100 bootstrap-derived iterations) models for each group of models. For each group of models, we use the median optimism value and the median AUC of all samples to evaluate the stability of the models. In order to condense our writing, we use the *median AUC* and the *median optimism value* to express the above concepts.

2) **Explanatory Variables Analysis:** We further study the impact of each factor on the issue-addressing likelihood by using the **anova** function in the R **rms** package to compute the Wald χ^2 value. The larger the Wald χ^2 value of a factor is, the larger impact of the factor on the issue-addressing likelihood. For each sample, we computed the Wald χ^2 value for each factor. Then we use the median Wald χ^2 value of each factor to represent the impact of that factor.
In addition, to further understand how a factor influences the value of the response variables, we use the **Predict** function in the **rms** R package to plot the estimated issue-addressing likelihood against a factor.

1. https://cran.r-project.org/web/packages/rms/index.html

Appendix C
Results of Sensitivity Analysis

In this section, we visualize the result of our analysis in the threats to validity section regarding different bounty-usage frequency thresholds (i.e., 40 and 60). Tables 1 and 2 present the results of the built models with using different thresholds. Our findings are not affected by our choice for the threshold for the bounty-usage frequency.

References

[1] B. Efron. How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81(394):461–470, 1986.
[2] F. E. Harrell, Jr. *Regression Modeling Strategies*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
[3] S. Kabinna, C.-P. Bezemer, W. Shang, M. D. Syer, and A. E. Hassan. Examining the stability of logging statements. *Empirical Software Engineering*, 23(1):290–333, 2018.
[4] S. McIntosh, Y. Kamei, B. Adams, and A. E. Hassan. An empirical study of the impact of modern code review practices on software quality. *Empirical Software Engineering*, 21(5):2146–2189, 2016.
[5] G. K. Rajbahadur, S. Wang, Y. Kamei, and A. E. Hassan. The impact of using regression models to build defect classifiers. In *Proc. of the Int'l Conf. on Mining Software Repositories*, pages 135–145, 2017.
[6] S. Wang, T.-H. Chen, and A. E. Hassan. Understanding the factors for fast answers in technical Q&A websites. *Empirical Software Engineering*, pages 1–42, 2017.

TABLE 1: The results of the sensitivity analysis of global, moderate and frequent models (under the threshold of 40). We highlight the most important factors of models of each group in bold. The **NL** indicates the non-linear term and the **D.F.** indicates the degree of freedom.

		Global Models		**Moderate Models**		**Frequent Models**	
Median AUC		0.71		0.70		0.81	
Median Optimism Value		0.01		0.01		0.01	
Factors		Overall	NL	Overall	NL	Overall	NL
I_B_days_before_bounty	D.F.	4	3	4	3	4	3
	χ^2	**93.19*****	**13.27****	**28.24*****	**0.122**	**59.53*****	**17.33****
P_B_usage_group	D.F.	2		-		-	
	χ^2	**63.57*****		-		-	
I_B_total_value	D.F.	2	1	2	1	2	1
	χ^2	**12.57****	**11.84****	8.11*	1.25	11.38**	1.87
I_code_proportion	D.F.	1		1		1	
	χ^2	5.71*		**21.89*****		9.05*	
I_B_has_label	D.F.	1		1		1	
	χ^2	**23.02*****		1.91		**19.60*****	
Backer_exp_B_max_value	D.F.	2	2	3	2	3	2
	χ^2	**17.37****	**19.70****	3.47	2.94	7.40	6.80
P_B_paid_proportion	D.F.	3	2	3	2	3	2
	χ^2	4.06	3.30	**11.35***	**0.42**	**15.26****	**14.69****
P_B_total_value	D.F.	2	1	2	1	2	1
	χ^2	11.95**	11.43**	0.19	0.01	1.40	1.05
I_img_cnt	D.F.	1		1		1	
	χ^2	2.66		2.04		**20.22*****	
I_link_cnt	D.F.	1		1		1	
	χ^2	0.00		1.05		3.48	
I_content_len	D.F.	1		1		1	
	χ^2	0.00		**7.22**		2.41	
I_cmnt_perday_mean	D.F.	2	1	2	1	2	1
	χ^2	1.24	1.06	2.30	2.25	0.25	0
I_B_cnt	D.F.	1		1		1	
	χ^2	4.89*		2.00		**13.37*****	
I_cmnt_cnt	D.F.	1		1		1	
	χ^2	1.01		1.00		10.85**	
Backer_role_any_insider	D.F.	1		1		1	
	χ^2	0.24		0.01		0.04	
Backer_role_have_reporter	D.F.	1		1		1	
	χ^2	0.04		**7.27****		4.88*	

P-value of the χ^2 test: '***' < 0.001; '**' < 0.01; '*' < 0.05

[7] J. Zhou. Supplementary material for our paper. https://github.com/SAILResearch/wip-18-jiayuan-bountysource-SupportMaterials/blob/master/appendix.pdf, 2018.

TABLE 2: The results of the sensitivity analysis of global, moderate and frequent models (under the threshold of 60). We highlight the most important factors of models of each group in bold. The **NL** indicates the non-linear term and the **D.F.** indicates the degree of freedom.

		Global Models		**Moderate Models**		**Frequent Models**	
Median AUC		0.73		0.70		0.81	
Median Optimism Value		0.00		0.01		0.01	
Factors		Overall	NL	Overall	NL	Overall	NL
I_B_days_before_bounty	D.F.	4	3	4	3	4	3
	χ^2	**106.22*** **	**7.64**	**41.62*** **	**2.13**	**39.19*** **	**13.00**
P_B_usage_group	D.F.	2		-		-	
	χ^2	**31.79*** **		-		-	
I_B_total_value	D.F.	2	1	2	1	2	1
	χ^2	**15.21*** **	**14.02*** **	**3.56**	**1.18**	**16.09*** **	**9.36***
I_code_proportion	D.F.	1		1		1	
	χ^2	8.01		1.89		0.93	
I_B_has_label	D.F.	1		1		1	
	χ^2	**31.77*** **		**13.54*** **		2.96	
Backer_exp_B_max_value	D.F.	3	2	3	2	3	2
	χ^2	**15.54** **	**13.81** **	2.72	2.18	**16.48** **	**13.72** **
P_B_paid_proportion	D.F.	3	2	3	2	3	2
	χ^2	14.30	5.20	**23.17*** **	**1.05**	**16.77** **	**15.30*** **
P_B_total_value	D.F.	2	1	2	1	2	1
	χ^2	0.59	0.07	0.12	0.10	3.90	3.89*
I_img_cnt	D.F.	1		1		1	
	χ^2	13.88		1.78		**17.00*** **	
I_link_cnt	D.F.	1		1		1	
	χ^2	1.14		0.25		0.36	
I_content_len	D.F.	1		1		1	
	χ^2	0.01		1.89		0.67	
I_cmnt_perday_mean	D.F.	2	1	2	1	2	1
	χ^2	0.01	0.03	0.31	0.23	3.31	1.21
I_B_cnt	D.F.	1		1		1	
	χ^2	1.99		2.52		10.81**	
I_cmnt_cnt	D.F.	1		1		1	
	χ^2	10.83		**2.90**		1.26	
Backer_role_any_insider	D.F.	1		1		1	
	χ^2	4.47*		0.67		0.16	
Backer_role_have_reporter	D.F.	1		1		1	
	χ^2	4.69*		1.92		5.60*	

P-value of the χ^2 test: '***' < 0.001; '**' < 0.01; '*' < 0.05