Survival Analysis of Questions Posted on the iFixit Answers forum

L.K. Oshita[a,*], A. Pileggi[b], S. Pileggi[a]

[a]*Department of Statistics, California Polytechnic State University, San Luis Obispo, CA 93407, United States*
[b]*iFixit, 1330 Monterey St, San Luis Obispo, CA 93401, United States*

Abstract

Community-driven online question and answer forums (CQA), platforms that allow users to ask and answer questions, contain an extensive amount of crowd-sourced information. One such example of a CQA is iFixit's *Answers* forum. This forum features user-asked questions related specifically to device repair, which are answered by both repair experts and everyday users. An important aspect of these CQAs, for both users and the forum, is question response time. This paper presents a survival analysis of the time until a question posted on iFixit's *Answers* forum receives it's first response. A Cox proportional hazards model was developed to identify variables associated with response time. Though several predictors were significant, the model's predictive accuracy was low ($R^2 = 0.15$). Significant predictors included the device category of the question (questions pertaining to Apple products received answers faster than others (HR = 2.54, 95% CI = (2.32, 2.79))), factors related to the question's title (e.g., if it was phrased as a question (HR = 1.29, 95% CI = (1.21, 1.38))), and the day of the week the question was posted (questions posted over the weekend received answers slower than those posted on a weekday (HR = 0.87, 95% CI = (0.82, 0.93))). Future studies can investigate whether factors identified as significant in this analysis can be generalized to other CQAs.

Keywords: Cox regression, CQA forum, response time, survival analysis, survival probabilty

1. Introduction

Community-driven online question and answer forums have become widely-used sources of information. These online platforms feature thousands of user-posted questions and answers and receive millions of visits every month. The CQA analyzed in this paper is iFixit's *Answers* forum. Founded in 2003, iFixit

---

*Corresponding author
Email address: l.oshita@yahoo.com (L.K. Oshita)

provides over 35,000 free online repair guides and sells the specialized tools and parts needed for such repairs. This company's mission is to equip users with the knowledge and tools necessary to repair their broken devices, as part of an effort to help them save money and reduce electronic waste.

As not all possible repairs are covered in the published guides and users may have additional questions related to existing guides, iFixit's *Answers* forum is an important resource for users. This platform features questions pertaining to over 10,000 devices, ranging from jammed zippers to shattered phone screens and faulty air conditioners, with over 120,000 user-posted solutions. As thousands of users rely on this forum for information, it is important that users receive timely answers. Fast response times have been shown to be associated with positive user satisfaction and, in turn, increase web traffic, which is valuable to the reputation and longevity of the CQA [1]. Analysis of response times can reveal factors that affect how quickly questions receive answers, which can lead to suggestions for how users can ask better questions to minimize response times and for how the forum design can be improved.

However, analysis and prediction of response times on CQAs have not been thoroughly investigated. There is need for further analysis of forum response times, as the majority of existing research focuses on assessing and predicting question or answer quality. This paper presents a survival analysis on the time until a question receives its first answer on iFixit's *Answers* forum, in order to determine factors significantly related to response time and to predict survival probabilities of questions.

2. Related Work

With the recent increase in the popularity and use of CQAs, these data-rich platforms have been the subject of a multitude of studies. Regarding analysis of response times, Bhat et al. [2] developed a classification model using both textual and non-textual features to estimate response times for questions posted on *Stack Overflow*. Mahmud et al. [3] used parametric methods and proposed models to predict response times based on exponential distributions. Asaduzzaman et al. [4] instead focused exclusively on unanswered questions and created a model to predict how long a question would remain unanswered. However, all research mentioned restricted analysis to either questions for which response times were observed or to questions that remained unanswered. The research presented in this paper uses survival analysis, a method not yet applied to CQAs, to allow for analysis of both answered and unanswered questions in a unified framework.

The majority of existing research has been focused on predicting question and answer quality. Weimer et al. [5] developed a classification algorithm to assess the quality of posts on *Nabble.com* using primarily textual features of questions. Blumenstock [6] also determined that textual features (e.g. word count) was the most accurate predictor of Wikipedia article quality (Wikipedia articles can represent the same kind of user-generated content featured on CQAs). Fu et al. [7] instead found that non-textual features, like revision

and comment count or the number of points a user has, were the most useful indicators of quality. A number of other studies have also developed classification algorithms using textual and non-textual features with the similar goal of predicting question or answer quality [8, 9, 10, 11]. The present analysis seeks to utilize both textual and non-textual features of questions on iFixit's *Answers* forum to investigate if factors that indicate that a question is of high quality also lead to faster response times.

3. Materials

Categorical Variables	Device category the question pertains to. Categories include: Android/Other Phone, Apple Products, Camera, Electronics, Game Console, Home, Other, PC, Tablet, Vehicle
	If the question was posted on a weekend or on a weekday
	If the question's text contains end punctuation marks (. ? !)
	If the question's text is in all lower case
	If the question's title contains at least one word considered to be "frequently-used" among answered questions
	If the question's title contains at least one word considered to be "frequently-used" among unanswered question
	If the question's title ends in a question mark
	If the user edited or added information to the question's text after posting it
	If the user made an effort to solve the problem prior to asking the question
Continuous Variables	Average number of characters in each question's tags (average tag length)
	Average frequency for all of a question's tags (average tag frequency)
	Number of characters in the question's text (text length)
	Number of characters in the user-defined device name (device name length)
	Ratio of the number of line breaks to the number of characters in the question's text (line break to text length ratio)

Table 1: Categorical and continuous variables derived

The data analyzed contained 8,025 questions posted from April 8, 2017 to July 7, 2017 (date of data download). Variables in the data included: device name, device category, question title, text, tags, new user status, date and time the question was posted, and date and time the first answer was received. Variables derived can be found in Table 1. The Appendix contains details on variable derivation, as well as an example of an answered and unanswered question in Table 5.

4. Methods

Questions analyzed were restricted to those posted in English. The time until event variable used in survival analysis was defined as the time since posting until a question received its first answer. For questions that did not receive an answer by the download date, time until event values were defined as the time since posting to the time the data was downloaded. Such questions were considered right-censored, meaning that exact answer times for these questions are greater than the observed download time (questions may receive answers after the download date) [12].

Survival was defined as the event that a question did not receive an answer beyond a certain time, t. Estimates of survival probability were generated with the Kaplan-Meier method, which adjusts to the presence of right-censored, or unanswered questions [13]. From these estimates, survival curves were constructed to examine the survival experience of questions. Mean and median survival times were also generated.

As the probability distribution for response times is unknown, a non-parametric Cox proportional hazards model was developed to predict the survival probability of questions [14]. To build the model, five-fold cross-validation was used [15]. Univariate analysis, performed on one training set, was used to identify variables to include in the final predictive Cox model [16]. Each predictor was entered into univariate Cox models and the evidence of association with response times was assessed. Those with partial likelihood ratio p-values of less than 0.001 were included in the final model. Continuous predictors were entered into univariate Cox models both with and without square root and log transformations to determine the form of the predictor to include in the final model. Univariate analysis was again performed on each continuous predictor to investigate the use of splines. Each predictor, with the transformation found necessary in the previous analysis, was fit to three separate Cox models with restricted cubic splines of three, four, and five knots. AIC statistics were used to determine whether or not to include splines, as well as the optimal number of knots [17]. All predictors found to be significant, along with transformations and splines found to be necessary, were included in the model for cross-validation. No variable selection was used after identifying significant predictors in univariate analysis, as Harrell [17] determined that reducing the number of predictors in the model decreased predictive accuracy more so than retaining all predictors.

In each iteration of cross-validation, the full model was built on the training set and used to generate predicted hazard ratios on the corresponding test set. To assess prediction performance, predicted hazard ratios were entered into separate Cox models as the single quantitative predictor with response times as survival time. The resulting Nagelkerke's R^2 statistic, concordance statistic, Somers' Dxy, partial likelihood ratio and p-value, were assessed as indicators of the model's performance [18]. Significant results from this model would indicate high predictive accuracy. The partial likelihood ratio assessed is a ratio of the log partial likelihood function evaluated at the parameter estimates, equivalent to a goodness of fit measure of the model with predictors, and the

4

log partial likelihood function for the null model without predictors and only the baseline hazard function, equivalent to a goodness of fit measure for the null model. The p-value for this statistic was calculated from the χ^2 distribution [19]. Concordance statistics and Somers' Dxy are a measure of the model's discriminative ability. Concordance is defined as the probability that for any two randomly chosen questions, the question with the shorter response time also has the higher predicted hazard. Concordance statistics close to 1 indicate high discriminative ability, while statistics close to 0.5 indicate discordance, or random predictions. Somers' Dxy is the difference between the model's concordance statistic and discordance, 0.5. A Somers' Dxy statistic of 0 indicates random predictions, while a statistic equal to 1 indicates perfect predictions [17]. These metrics were computed for each iteration on every training and test set. Averages of metrics across training and test sets were evaluated. The final model was then fit to the full data and the same metrics were computed and compared.

Correlations between scaled Schoenfeld residuals, differences between observed and expected predictor values for questions that received answers, and a function of time, were examined to assess the proportional hazards assumption that the effect of predictors on hazard does not depend on time [20]. Significant correlations would indicate a predictor has violated this assumption.

All computation was executed under R version 3.4.3. Kaplan-Meier estimates of survival probability were computed with the Surv and survfit functions of the Survival package in R. To compute predicted survival probabilities using the final model, the predictSurvProb function was used from the pec package. This function uses the Cox model and a vector of times to compute predictions. Data used for this analysis can be found here(Mendeley citation). Cited data includes all variables used in the model. Code for variable setup and analysis is available upon request.

5. Results and Discussion

Of 8,025 questions in the full data, 7,760 were in English (97% of the full data). Of those questions, 4,951 (63.8%) received an answer by the download date. The shortest response time was 0.5 hours. The longest was 2,159 hours (90 days). Figure 1 shows the distribution of response times for all questions analyzed.

Figure 2 shows the Kaplan-Meier estimated survival probability curve for all questions in the data. The curve indicates that if a question does not receive an answer soon after it is posted, the likelihood of it receiving an answer in the future is low. The Kaplan-Meier estimated mean survival time, or the average time until a question received its first answer was 776 hours, or 32.3 days. The median survival time, the time at which 50% of the questions in the data received an answer, was 9.2 hours.

Each training set contained 6,208 questions; each test set contained 1,552 questions. Univariate analysis performed on one training set indicated that all categorical predictors were significant with partial likelihood ratio p-values of

less than 0.001. Continuous predictors, excluding device length and average tag length, also resulted in univariate p-values of less than 0.001. Applying either square root or log transformations to all continuous predictors resulted in univariate p-values of less than 0.001. For variables that contained zeros, one was added to each observation before applying the log transformation. The following transformations were applied to continuous variables: square root transformation of average tag frequency, log plus one transformation of average tag length, log plus one transformation of device name length, square root transformation of line break to text length ratio, and log transformation of text length. All categorical predictors and all continuous predictors with necessary transformations were retained for the full model.

Predictor	K	AIC
$\sqrt{\text{average tag frequency}}$	0*	65890.75
	5	65891.69
	4	65891.75
	3	65892.35
$\log\left(\text{average tag length} + 1\right)$	0	65920.24
	5	65911.20
	4*	65910.81
	3	65914.98
$\log\left(\text{device length} + 1\right)$	0	65918.06
	5*	65847.34
	4	65880.89
	3	65898.00
$\sqrt{\text{line breaks / text length}}$	0	65890.28
	5	65884.70
	4	65882.98
	3*	65881.93
$\log\left(\text{text length}\right)$	0	65862.83
	5*	65862.02
	4	65863.35
	3	65862.08

Table 2: The model with the lowest AIC was used to determine the optimal k number of splines for each continuous predictor.

Table 2 shows the results of determining the optimal number of knots for each continuous predictor. Those marked with an asterisk (*) were included in the final model.

Average performance metrics for test and training sets are in Table 3. Partial likelihood ratios and p-values indicate that the model as a whole is significantly associated with response time. However, R^2 statistics and discrimination indexes are considerably low. Metrics for the final model's performance on the full data, also found in Table 3, are consistent with metrics found in cross-validation and indicate low predictive accuracy. Results for training, test, and the full data did

	HR	LR	p-value	R^2	Dxy	Concordance
Training Sets	2.03	937.39	<0.0001	0.14	0.27	0.63
Test Sets	1.99	220.83	<0.0001	0.14	0.26	0.63
Full Data	2.03	1165.03	<0.0001	0.14	0.28	0.63

Table 3: Performance metrics of final model. (HR: Hazard Ratio, LR: Partial Likelihood Ratio)

not change significantly, indicating that the model was not over fit.

Variable		**Coefficient (SE)**	**p-value**
Device Category	Apple Product	0.93 (0.048)	<0.0001
	Camera	-0.32 (0.090)	
	Electronics	-0.01 (0.078)	
	Game Console	0.24 (0.083)	
	Home	0.34 (0.070)	
	Other	-0.13 (0.056)	
	PC	0.28 (0.060)	
	Tablet	-0.16 (0.081)	
	Vehicle	0.40 (0.069)	
	Android/Other Phone (reference)	—	
Weekend		-0.13 (0.033)	<0.0001
Text contains end punctuation		0.03 (0.050)	0.613
Text is in all lower case		-0.18 (0.064)	0.006
Title contains terms considered frequently-used among answered questions		0.05 (0.042)	0.260
Title contains terms considered frequently-used among unanswered questions		-0.28 (0.034)	<0.0001
Title ends in a question mark		0.26 (0.033)	<0.0001
User edited or added to the question's text after posting it		0.30 (0.086)	0.001
User was a member for less than one day before posting		-0.11 (0.036)	0.003
User made an effort to solve the problem prior to asking the question		-0.07 (0.036)	0.045
Square root of the average tag frequency		2.23 (0.720)	0.002

Table 4: Coefficients for predictors in the final model. (Continuous predictors fit with restricted cubic splines are not shown)

Assessing the proportional hazards assumption indicated that several levels of the device categorization variable, specifically Apple Product, Camera, Game Console, and Other, violated the assumption.

As a whole, the final model was significant with a partial likelihood ratio of 1265.29 (p-value <0.0001). Its R^2 statistic was 0.15, and Somers' Dxy was 0.27.

The following are interpretations of select hazard coefficients in Table 4. Hazard is approximately equivalent to the conditional probability that a question will receive an answer within the next moment in time, given that it has not

already received an answer.

- The estimated hazard of receiving an answer is 154% higher (95% CI (132%, 179%)) for questions pertaining to Apple products than the hazard for questions about Android and Other phones, controlling for all other predictors.

- The estimated hazard of receiving an answer is 25% lower (95% CI (19%, 29%)) for questions with titles that contain at least one word considered to be frequently-used among unanswered questions, than the hazard for questions with titles that do not, controlling for all other predictors.

- The estimated hazard of receiving an answer for questions posted on a weekend is 13% lower (95% CI (7%, 18%)) than the hazard for questions posted on a weekday, controlling for all other predictors.

Results and coefficients of the final model reveal how users can ask questions in order to decrease response time. Findings suggest that users should phrase the title in the form of a question, use correct grammar (e.g. punctuation and capitalization), include specific and concise tags, and post the question on a weekday.

The model's weak predictive accuracy, which may be explained by limitations in the data, lead to suggestions for changes in CQA design. Many users incorrectly specified the device name (e.g. a user asking a question about a Turtle Beach Ear Force XO ONE headset defined the device name as, "Turtle Beach Ear Force Xmy grandson chewed through the wire while he was playing it's brand-new is there anyway I can have it fixed0 One"), or did not include a device name. Many users also incorrectly used the tagging system by including ambiguous and lengthy tags like "someone sat on it :(" or "help me please!!!!!" (tags are generally a couple key words that allow the answering community to quickly discern the question's topic). It is likely that these inconsistencies contributed to the final model's low predictive power. This, along with the results of the final model, reveal some of the ways the CQA can be structured to potentially decrease response times. As findings indicate that questions with correctly defined tags and device names may lead to quicker response times, the CQA can provide a stricter framework for asking questions. Allowing users to enter any device name or tag leaves room for user error, shown by the examples described above. Instead, the CQA can restrict the tags or devices that users can include to a drop-down list. The CQA can also include more tips to guide users asking questions. Implementing a more structured framework for asking questions can help users create understandable and clear questions and in turn decrease response time, as well as create a set of consistent questions for improved analysis.

6. Conclusion

This study developed a Cox proportional hazards model to predict the probability that a question posted on iFixit's *Answers* forum receives an answer

before a certain time. Predictors found to be significant included: device category, if the question was posted on a weekend or a weekday, whether or not the question's title contained at least one word considered frequently-used among unanswered questions, and whether or not the question's title ends in a question mark. While overall the model was significant, its predictive performance was considerably low.

7. Vitae

Lisa Oshita is a senior undergraduate statistics major at the California Polytechnic State University in San Luis Obispo. Shannon Pileggi is an Assistant Professor in the Statistics Department at California Polytechnic State University, San Luis Obispo. Anthony Pileggi is a Data Scientist at iFixit. Declarations of interest: none.

8. Author Contributions

All authors contributed to the analysis featured in this paper and have edited and approved of the final article.

9. Acknowledgements

10. Appendix

10.1. Posted questions

Table 5 provides observed and derived variable values from an example of an answered question (regarding the iPhone 6) and unanswered question (regarding android tablet). All fields in device, title, text, and tags are shown as the users entered it.

10.2. Variable Derivation

10.2.1. Average tag frequency

This variable was created to investigate if including popular or widely-used tags in a question lead to faster response times. This is computed as the average frequency, or proportion of times, a question's tags appear in all of the data set. Questions without tags were assigned a value of 0. If a question has a higher average, then at least one of its tags are considered frequently-used. This variable was created based off of findings from Bhat et al. [2].

Device	iPhone 6	android tablet
Title	iPhone water damage, touch screen issue	ccccaaaan you help me fix my touch screen
Text	So I dropped my iPhone in water 4 days ago. Have done the whole rice thing and seen huge difference in it. However, only one side of my screen works and it is the one side which I need to unlock the phone. What would be the best way forward?	touch screen not worki
Tags	touchscreen, water damage	NA
Response Time (hrs)	1.361	NA
Device Category	Apple Product	Tablet
Day Posted	Weekday	Weekday
Text length (characters)	241	81
Device name length (characters)	8	18
Average tag length (characters)	11.5	47
Average tag frequency	0.0097	0
Text contain end punctuation?	True	False
Text in all lower case?	False	True
Title end in a question mark?	False	False
Line break to text length ratio	0	0
User indicate prior effort?	False	False
User update question after posting?	False	False
Frequently-used answered words?	True	True
Frequently-used unanswered words?	False	False

Table 5: Examples of an answered question (left) and an unanswered question (right)

10.2.2. Average tag length

This variable was created to investigate if questions with correctly defined tags have faster response times than questions that do not include tags or incorrectly use the tagging system. It is hypothesized that questions with ambiguous or lengthy tags make it difficult for the answering community to discern the question's topic, and thus have slower response times or remain unanswered.

10.2.3. Device categorization

Original device categories defined by iFixit included: Apparel, Appliance, Camera, Car and Truck, Computer Hardware, Electronics, Game Console, Household, Mac, Media Player, PC, Phone, Skills, Tablet, Vehicle. Device titles were parsed and certain categories were combined or separated to create a new device categorization variable to use in the Cox model. Final categories created included: Android/Other Phone, Apple Product, Camera, Electronics, Game Console, Home, Other, PC, Tablet, Vehicle. Under the original device categorization, 1,954 questions (25.2% of 7,760) were categorized incorrectly and indicated an NA for its category. Missing values were a result of users creating questions for devices not already in the iFixit's database, or from the user incorrectly defining the device name. For questions with missing categories, key words were searched for in device titles to re-categorize accordingly. All other questions with ambiguous device titles were categorized as Other.

10.2.4. Device name length

This variable was created to investigate if questions with correctly defined device names have faster response times than questions that do not include a name or are incorrectly defined. Similar to the average tag length variable, it is hypothesized that questions with incorrectly defined device names (often lengthy) make it difficult for the answering community to discern the question's topic, and thus have slower response times.

10.2.5. If the question's text contains end punctuation

This variable was created to investigate if sentences with no end punctuation take longer to receive an answer.

10.2.6. If the question's title contains terms considered frequently-used among answered/unanswered questions

These variables were created to investigate if questions pertaining to popular topics receive answers faster than questions that do not. Similarly, another variable was created to investigate if questions pertaining to unpopular topics receive answers slower than questions that do not. This variable was created based off of methods from Correa and Sureka [21] and Ravi et al. [11].

To create the variables, the data was separated between answered and unanswered questions. For the data frame containing answered questions, text mining techniques were used to create a list of every word within each questions'

titles and the proportion of times those words occurred among all answered questions' titles, defined as the frequency. The same was performed for the data frame containing unanswered questions. "Frequently-used" words in answered questions were defined as those that appeared in 1% or more of all answered questions' titles, and appeared in more answered questions than in unanswered questions. To determine the latter, a ratio of frequencies was assessed. The ratio was calculated as the proportion of times a word occurred among answered questions, to the proportion of times a word occurred in unanswered questions. As an example, if "cracked" appeared in 2% of answered questions and 0.1% of unanswered questions, it would be considered "frequently-used" among answered questions as it occurs in more than 1% of answered questions' titles and occurs 20 times more in answered question than in unanswered questions (ratio = 0.02/0.001 = 20). Similarly, "frequently-used" words in unanswered questions must occur in 1% or more of all unanswered question's titles and occur in more unanswered questions than answered questions. As there was some overlap between "frequently-used" words in each list and levels of the device categorization variable, every word that matched a device name was removed from the lists. Table 6 contains the resulting terms for answered questions (111 words) and for unanswered questions (32 words).

Frequently-used terms among answered questions	screen, turn, working, replacement, power, work, replace, charging, charge, button, touch, black, turning, broken, start, stuck, new, LCD, upgrade, problem, change, port, replaced, card, open, boot, replacing, remove, reset, back, drive, error, cable, ssd, cracked, hard, one, dropped, logic, lines, white, keeps, pro, dead, now, front, damage, switch, parts, glass, still, charger, issue, sim, turns, digitizer, just, mode, model, backlight, usb, stopped, logo, starting, know, unresponsive, password, factory, call, use, damaged, find, sensor, possible, fixed, side, galaxy, data, ipod, problems, issue, slow, system, connector, without, overheating, code, ram, air, microphone, please, red, much, plugged, getting, booting, left, way, buy, plus, time, loose, lock, coming, got, shuts, says, install, key, door, top
Frequently-used terms among unanswered questions	sound, light, wifi, speaker, connect, picture, stay, noise, bluetooth, isnt, apps, going, rear, question, play, stop, service, take, hear, lights, showing, network, volume, come, keep, connection, flashing, shut, print, blue, buttons, edit

Table 6: Frequently-used terms among answered and unanswered questions' titles, ordered by decreasing frequency

10.2.7. If the question's title ends in a question mark

This variable was created to investigate if questions with titles in the form of questions receive answers faster than those otherwise phrased [2].

10.2.8. If the user made an effort to solve the problem prior to asking the question

This variable indicates true if the asker included words in the question's text that indicate that prior effort was made to find the solution to the question before posting. Words searched for include: tried, searched, researched, tested, replaced, used, checked, investigated, considered, measured, attempted, inspected, fitted. This variable was created based off of findings and methods from Bhat et al. [2] and Harper et al. [22].

10.2.9. Line break to text length ratio

This variable was created to investigate if questions that include line breaks in the text are easier to read than questions that do not include any, and thus have shorter response times. Questions with a ratio of 0 tend to include either short, ambiguous texts (e.g. "Start button not working.") or long and lengthy texts, and tend to remain unanswered or have slower response times. Questions with a ratio of greater than 0 tend to have faster response times.

References

[1] A. Rechavi, S. Rafaeli, Not all is gold that glitters response time & satisfaction rates in Yahoo! answers, Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011 (2011) 904–909doi:\let\@tempa\bibinfo@X@doi10.1109/PASSAT/SocialCom.2011.67.

[2] V. Bhat, A. Gokhale, R. Jadhav, R. Pudipeddi, L. Akoglu, Min (e) d Your Tags : Analysis of Question Response Time in StackOverflow, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (Asonam) (2014) 328–335.

[3] J. Mahmud, J. Chen, J. Nichols, When Will You Answer This? Estimating Response Time in Twitter, in: Seventh International AAAI Conference on Weblogs and Social Media, 697–700, URL http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewPDFInterstitial/6072/6328, 2013.

[4] M. Asaduzzaman, A. S. Mashiyat, C. K. Roy, K. A. Schneider, Answering questions about unanswered questions of Stack Overflow, 2013 10th Working Conference on Mining Software Repositories (MSR) (2013) 97–100ISSN 2160-1852, doi:\let\@tempa\bibinfo@X@doi10.1109/MSR.2013.6624015, URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6624015.

[5] M. Weimer, I. Gurevych, M. Mühlhäuser, Automatically assessing the post quality in online discussions on software, Proceedings of the ACL (June) (2007) 125–128, doi:\let\@tempa\bibinfo@X@doi10.3115/1557769.1557806, URL `http://atlas.tk.informatik.tu-darmstadt.de/Publications/2007/acl2007.pdf`.

[6] J. E. Blumenstock, Size matters: word count as a measure of quality on wikipedia, in: Proceedings of the 17th International Conference on World Wide Web, ACM, ISBN 9781605580852, ISSN 08963207, 1095–1096, doi:\let\@tempa\bibinfo@X@doi10.1145/1367497.1367673, URL `http://portal.acm.org/citation.cfm?id=1367673`, 2008.

[7] H. Fu, S. Wu, S. Oh, Evaluating answer quality across knowledge domains: Using textual and non-textual features in social Q&A, Proceedings of the Association for Information Science and Technology 52 (1) (2015) 1–5, ISSN 23739231, doi:\let\@tempa\bibinfo@X@doi10.1002/pra2.2015.145052010088.

[8] Y. Yao, H. Tong, T. Xie, L. Akoglu, F. Xu, J. Lu, Detecting high-quality posts in community question answering sites, INFORMATION SCIENCES 302 (2015) 70–82, ISSN 0020-0255, doi:\let\@tempa\bibinfo@X@doi10.1016/j.ins.2014.12.038, URL `http://dx.doi.org/10.1016/j.ins.2014.12.038`.

[9] H. Toba, Z.-y. Ming, M. Adriani, T.-s. Chua, Discovering high quality answers in community question answering archives using a hierarchy of classifiers, Information Sciences 261 (2014) 101–115, ISSN 0020-0255, doi:\let\@tempa\bibinfo@X@doi10.1016/j.ins.2013.10.030, URL `http://dx.doi.org/10.1016/j.ins.2013.10.030`.

[10] L. Ponzanelli, A. Mocci, A. Bacchelli, M. Lanza, Understanding and classifying the quality of technical forum questions, Proceedings - International Conference on Quality Software (2014) 343–352ISSN 15506002, doi:\let\@tempa\bibinfo@X@doi10.1109/QSIC.2014.27.

[11] S. Ravi, B. Pang, V. Rastagori, R. Kumar, Great Question ! Question Quality in Community Q & A, International AAAI Conference on Weblogs and Social Media (1) (2014) 426–435.

[12] D. Kleinbaum, M. Klein, Survival Analysis: A Self-Learning Text, Third Edition (Statistics for Biology and Health), in: M. Gail, K. Krickeberg, J. Samet, A. Tsiatis, W. Wong (Eds.), Biometrical Journal, chap. 1, Springer-Verlag New York, third edn., ISBN 1441966455, 715, doi:\let\@tempa\bibinfo@X@doi10.1016/B978-0-12-387667-6.00013-0, 2011.

[13] J. M. Bland, D. G. Altman, Statistics Notes: Survival probabilities (the Kaplan-Meier method), Bmj 317 (7172) (1998) 1572–1580, ISSN 0959-8138, doi:\let\@tempa\bibinfo@X@doi10.1136/bmj.317.7172.1572, URL `http://www.bmj.com/cgi/doi/10.1136/bmj.317.7172.1572`.

[14] D. Moore, Applied Survival Analysis Using R, March, chap. 5, Springer International Publishing, 1 edn., ISBN 9783319312439, 226, doi:\let\ @tempa\bibinfo@X@doi10.1007/978-3-319-31245-3, 2010.

[15] J. D. Rodríguez, A. Pérez, J. A. Lozano, Sensitivity analysis of kappa-fold cross validation in prediction error estimation, IEEE transactions on pattern analysis and machine intelligence 32 (3) (2010) 569–575, ISSN 1939-3539, doi:\let\@tempa\bibinfo@X@doi10.1109/TPAMI.2009.187.

[16] K. E. Hammermeister, T. A. DeRouen, H. T. Dodge, Variables predictive of survival in patients with coronary disease. Selection by univariate and multivariate analyses from the clinical, electrocardiographic, exercise, arteriographic, and quantitative angiographic evaluations., American Heart Association, Inc. 59 (3) (1979) 421–430, doi:\let\@tempa\bibinfo@X@ doi10.1161/01.CIR.59.3.421, URL http://dx.doi.org/10.1161/01.CIR.59.3.421.

[17] F. E. Harrell, Regression Modeling Strategies, vol. 64, ISBN 978-1-4419-2918-1, doi:\let\@tempa\bibinfo@ X@doi10.1007/978-1-4757-3462-1, URL papers:// 55069ee6-504c-4f60-bfa9-053c4dcabb39/Paper/p398{%}5Cnhttp: //link.springer.com/10.1007/978-3-319-19425-7{%}5Cnhttp:// www.ncbi.nlm.nih.gov/pubmed/21531065{%}5Cnhttp://linkinghub. elsevier.com/retrieve/pii/S0300893211003502{%}5Cnhttp: //link.springer.com/10.1, 2015.

[18] H.-C. Chen, R. L. Kodell, K. F. Cheng, J. J. Chen, Assessment of performance of survival prediction models for cancer prognosis, BMC Medical Research Methodology 12 (1) (2012) 102, doi:\let\ @tempa\bibinfo@X@doi10.1186/1471-2288-12-102, URL https://doi. org/10.1186/1471-2288-12-102.

[19] D. Oakes, Survival Times: Aspects of Partial Likelihood, International Statistical Review / Revue Internationale de Statistique 49 (3) (1981) 235–252, ISSN 0306-7734, doi:\let\@tempa\bibinfo@X@doi10. 2307/1402606, URL http://www.jstor.org/stable/1402606.

[20] P. Grambsch, T. Therneau, Proportional hazards tests and diagnostics based on weighted residuals, Biometrika 81 (3) (1994) 515–526, URL https://doi.org/10.1093/biomet/81.3.515.

[21] D. Correa, A. Sureka, Fit or unfit: Analysis and prediction of 'closed questions' on Stack Overflow, Proceedings of the first ACM Conference on Online Social Networks (2013) 201–212ISSN 1450320848, doi:\let\ @tempa\bibinfo@X@doi10.1145/2512938.2512954, URL http://dl.acm. org/citation.cfm?doid=2512938.2512954.

[22] F. M. Harper, D. Raban, S. Rafaeli, J. A. Konstan, Predictors of Answer Quality in Online Q&A Sites, Proceedings of the SIGCHI Conference

on Human Factors in Computing Systems (2008) 865–874doi:\let\
@tempa\bibinfo@X@doi10.1145/1357054.1357191, URL `http://dl.acm.
org/citation.cfm?id=1357054.1357191`.