Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 13, Issue. 1, January 2024, pg.1 – 35

Team Size Dynamics: A Study on Factors Influencing Pull Request Acceptance in Open-Source Software Repositories

Miriam Al Arbaji Al Sayegh[1]; Nasser A Nasser[2]

[1]Faculty of Informatics Engineering, Department of Software Engineering and Information Systems, Tishreen University, Syria

[2]Faculty of Informatics Engineering, Department of Software Engineering and Information Systems, Tishreen University, Syria

[1] miriam.arbaji@tishreen.edu.sy; [2] n.nasser@tishreen.edu.sy

DOI: https://doi.org/10.47760/ijcsmc.2024.v13i01.001

Abstract— In contemporary software development, Pull-Based Development has become the prevalent methodology, prompting extensive research into the factors influencing the success of contributions. Our study delves into numerous factors while specifically investigating the pivotal role of team size in the evaluation and acceptance of pull requests. The significance of team size is underscored by its historical importance in guiding the choice of software methodologies. Our findings reveal that team size impacts the decision-making process concerning pull requests. The factors influencing these decisions are affected by team size, emphasizing the necessity of considering this variable in modern software projects. Our research illuminates the complex dynamics of pull-based development, providing valuable insights for project owners and developers alike.

Keywords— pull-based development, pull requests, software engineering, non-parametric tests, Mann-Whitney U test, Chi-Square test, practical significance, machine learning, classification.

1. Introduction

Pull-Based Development, as a widely adopted methodology, offers a wealth of insights into collaborative software development. Efforts to understand the complex process of reviewing and deciding on specific pull requests have been extensive, driven by the need to optimize time for both reviewers and contributors [1]. Platforms like GitHub, serving as vast repositories of open-source projects, provide invaluable data for scholarly exploration. Leveraging one of the largest and most recent datasets available [2], researchers have delved into the decision-making dynamics surrounding pull requests [3]. These studies have revealed that influencing factors vary across diverse scenarios and contexts [4], emphasizing the need for fresh perspectives.

In this context, our research introduces a novel viewpoint by examining the impact of team size on the pull request handling process. Team size has been demonstrated to significantly influence both traditional plan-based and agile software processes [5]. Smaller teams enjoy streamlined communication channels, fostering direct and informal interactions that facilitate swift and well-informed decisions regarding pull requests. These teams often share a cohesive vision and coding standards, simplifying the evaluation of pull requests against project

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

objectives. Conversely, larger teams navigate complex communication structures, necessitating multi-layered reviews involving specialized individuals or teams. The resultant coordination challenges can prolong decision-making processes, exacerbated by diverse opinions and coding styles that prompt extensive discussions and revisions. In light of these observations, our study seeks to address the following questions:

RQ1: Does team size influence the decision-making process for a specific pull request?

RQ2: What are the key factors affected by team size, and how do they differ among large, medium, and small teams?

By delving into these questions, our research aims to clarify the delicate relationship between team size and pull request acceptance. Ultimately, we aspire to propose comprehensive guidelines that enhance the core process of pull-based development, informed by the unique dynamics of team size in software development contexts.

2. Dataset Handling

The selected dataset [2] required adjustments to align with the objectives of our study. Initially comprising 96 features, we conducted essential modifications, including the removal of redundant elements and thoughtful imputation of missing values. To facilitate a nuanced analysis based on team size, we segmented the dataset into three distinct subsets: large teams, medium teams, and small teams. Furthermore, to ensure impartiality in our subsequent analyses, we meticulously balanced the classes—merged and not merged—within each subset. In the forthcoming section, we will demonstrate the sequential steps undertaken during this dataset handling process.

2.1. Elimination of Redundant Features

We will be enumerating the eliminated features along with the rationale behind their removal:

- open_diff/extra_diff/cons_diff/neur_diff/agree_diff: These five features correspond to the well-established OCEAN model, extensively studied for its profound impact on various aspects of an individual's professional life, including careers and, notably, on developers and how they react to pull requests [6]. While the dataset initially included ten features related to personality traits, with five assigned to contributors and five to integrators, the introduced difference attributes merely calculate the variance between integrator and contributor for each trait. However, these features lack specificity, as they do not indicate whether the difference is in the positive or negative direction. For instance, a zero difference in openness does not distinguish between integrator and contributor having compatible high scores or low scores for openness. Given that these nuances can be inferred from the original ten features—inte_open, inte_extra, inte_neur, inte_agree, inte_cons, and contrib_open, contrib_neur, contrib_extra, contrib_agree, contrib_cons—we have opted to remove the redundant five difference features for a more streamlined and interpretable dataset.
- files_changed: this feature represents the cumulative count of changes made in a pull request and can be derived from three distinct features: files_deleted, files_modified, and files_added. Consequently, to streamline our dataset and avoid redundancy, we have opted to eliminate the files_changed feature.
- num_comments: this feature is an aggregate metric derived from three individual features: comments_commit_num, pr_comment_num, and num_issue_comments. To enhance the efficiency of our dataset and mitigate redundancy, we have chosen to exclude the num_comments feature.
- num_commit_comments/pr_comment_num: We have identified redundancy in the features num_commit_comments and pr_comment_num, both of which quantify technical comments. In light of this, we have opted for a more streamlined representation, consolidating these two features into a singular variable named num_technical_comments. Consequently, our dataset will now distinguish between comments of a technical nature and general comments (num_issue_comments).
- part_num_commit/part_num_pr: In a similar way, we will merge the features part_num_commit and part_num_pr into a consolidated attribute named code_num_part, reflecting the number of contributors engaging in technical discussions. It's important to note that with this adjustment, our dataset now includes separate counts for contributors commenting on the contributed code (captured by the new feature, code_num_part) and those participating in general issue discussions (indicated by the existing feature, part_num_issue)
- num_participants: representing the total count of participants engaging in comments (both general and code-related), is found to be redundant, providing no additional or unique information. Consequently, we have decided to remove this feature from the dataset.
- perc_neg_emo/perc_pos_emo/perc_neu_emo: while crucial for gauging emotions in the comments and discussions surrounding a resolved issue in a pull request, lack the specificity to discern whether these emotions stem from the contributor or the integrator. Given that our dataset already encompasses features measuring positive, negative, and neutral emotions for both the contributor and integrator, we find these three attributes redundant. Consequently, we have opted to eliminate them.
- lifetime_minutes/mergetime_minutes: are features that can only be computed after the decision on a Pull Request has been finalized. While these attributes may offer valuable insights for other aspects of studying pull requests, they are not directly relevant to our research focus on the attributes influencing the decision-making

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

process of merging or not merging a pull request. As such, features related to post-decision stages are deemed extraneous to our current study.

- asserts_per_kloc: exhibits a positive correlation with two other attributes: test_cases_per_kloc and test_lines_per_kloc. This correlation is inherent since asserts typically exist within test cases, and the lengthier the test cases or lines, the higher the count of asserts. Recognizing this inherent relationship, we have decided to remove asserts_per_kloc from the dataset.

Consequently, 17 features have been removed, and two new features have been incorporated, resulting in a revised feature set totalling 81.

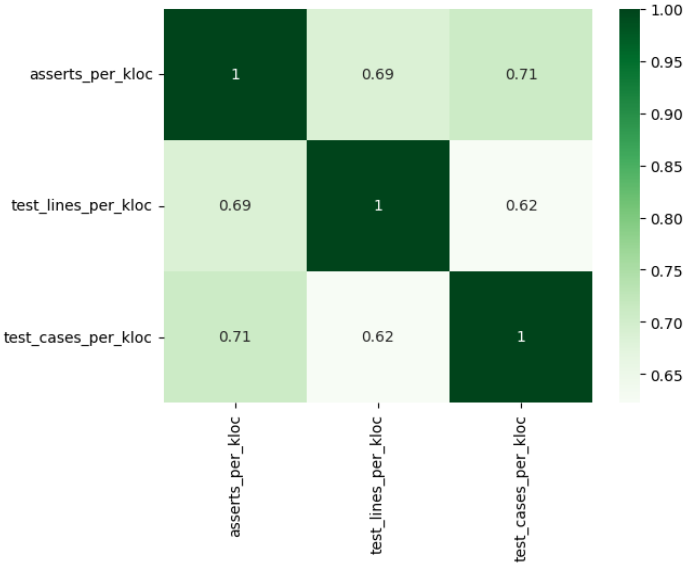

Fig. 1 Correlation Matrix of Testing Metrics: Asserts, Test Cases, and Test Lines per KLOC

2.2. Analysis of Missing Values

The presence of missing values in a dataset can significantly impact the performance of algorithms, as certain models struggle to process incomplete information effectively. Recognizing the importance of addressing these gaps in our dataset, we embark on a comprehensive exploration of missing values in each feature and column. Our aim is to generate a clear visual representation through plots, shedding light on the distribution of missing values. This analysis not only aids in understanding the extent of missing data but also lays the groundwork for informed strategies to handle these gaps effectively.

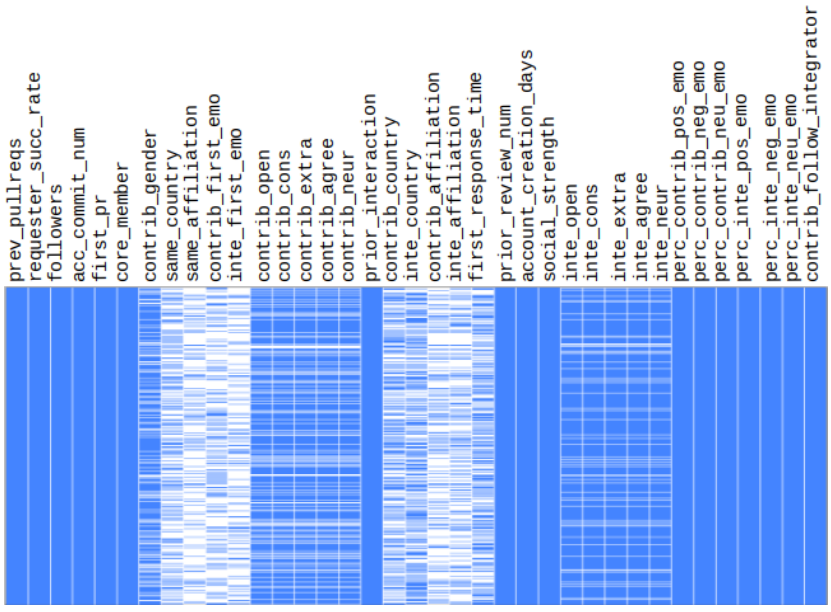

Fig. 2 Visualizing Missing Values: Human Factors

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

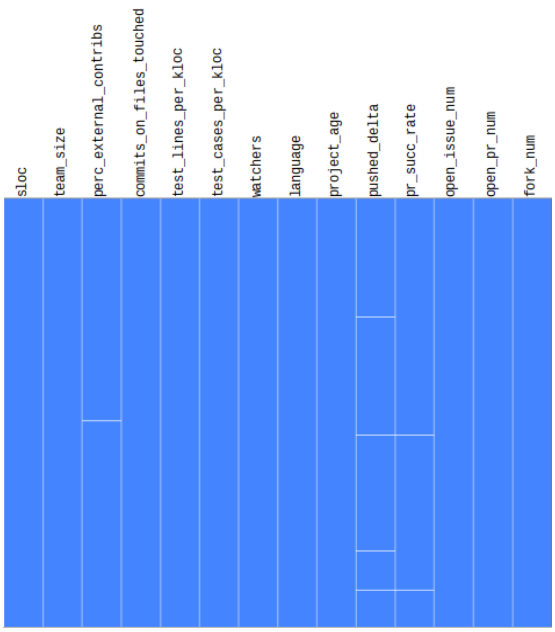

Fig. 3 Visualizing Missing Values: Project Factors

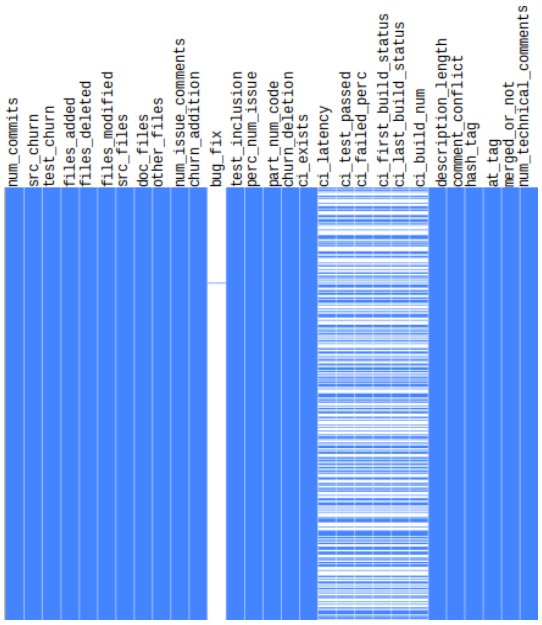

Fig. 4 Visualizing Missing Values: Pull Request Factors

In Figures 2 and 4, illustrating Human factors and features related to Pull Requests, a conspicuous pattern emerges with an abundance of missing values evident as white spaces within each bar. In contrast, the project features exhibit a comparatively lower occurrence of missing values. Recognizing the significance of quantifying these missing values, we present a detailed breakdown in Table 1, providing insights into the extent of gaps in our dataset across various features.

Upon reviewing Table 1, it becomes evident that specific features exhibit a substantial proportion of missing data, surpassing the 50% threshold across more than half of the rows. Notably, these missing values seem arbitrary in nature. An illustrative example is the disclosure of country or affiliation on GitHub profiles, a choice frequently omitted by many developers, contributing to the elevated incidence of missing values in the corresponding country/affiliation attributes. Further investigation includes assessing the correlation of these features with the label. Consequently, any feature demonstrating both a low correlation with the label and a percentage of missing values exceeding 50% will be removed, with the exception of the first_response_time feature. Despite its 50% missing values, this feature boasts a high correlation with the label, rendering it indispensable. This decision results in the removal of an additional 15 features from our dataset, which are:

- same_country
- same_affiliation
- contrib_country
- inte_country
- contrib_affiliation
- inte_affiliation
- bug_fix
- ci_latency
- ci_test_passed
- ci_failed_perc
- ci_first_build_status
- ci_last_build_status
- ci_build_num
- contrib_first_emo
- inte_first_emo

Hence, the final set of attributes utilized in our study comprises 66 features along with the label

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Attribute	Type	Number of Missing Values	Percentage of Missing Values
contrib_gender	Human Factor	694396	20%
same_country	Human Factor	2238549	66%
same_affiliation	Human Factor	2707224	80%
contrib_first_emo	Human Factor	2259030	67%
inte_first_emo	Human Factor	2646705	79%
contrib_open/cons/neur/extra_agree	Human Factor	634738	18%
contrib_country	Human Factor	1874929	56%
inte_country	Human Factor	1767681	52%
contrib_affiliation	Human Factor	2496936	74%
inte_affiliation	Human Factor	2398544	71%
first_response_time	Human Factor	1700048	50%
social_strength	Human Factor	250	0.0007%
inte_open/cons/neur/extra/agree	Human Factor	340196	10%
perc_contrib_neg/pos/neu	Human Factor	11344	0.33%
perc_inte_neg/pos/neu	Human Factor	3800	0.11%
bug_fix	Pull Request	3324411	99%
ci_exists	Pull Request	2683	0.08%
ci_latency	Pull Request	1800437	53%
ci_test_passed	Pull Request	1800437	53%
ci_failed_perc	Pull Request	1800437	53%
ci_first_build_status	Pull Request	1800437	53%
ci_last_build_status	Pull Request	1800437	53%
ci_build_num	Pull Request	1800437	53%
sloc	Project Factor	2056	0.06%
perc_external_contribs	Project Factor	4071	0.12%
test_lines_per_kloc	Project Factor	2056	0.06%
test_cases_per_kloc	Project Factor	2123	0.06%
pushed_delta	Project Factor	20542	0.6%
pr_succ_rate	Project Factor	10203	0.3%

Table. 1 Missing Values Percentage Across Features

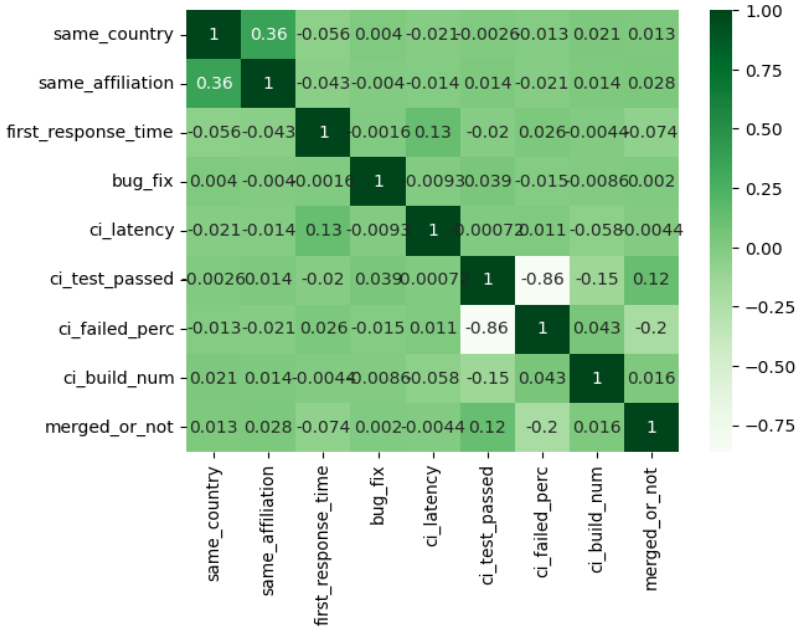

Fig. 5 Correlation of Features with High Missing Values with the Label

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

2.3. Addressing Rows with Incorrect Values

Several rows in the dataset were found to contain incorrect negative values, attributed to limitations in the data provided by GitHub [2]. Specifically, these discrepancies were observed in the features listed in Table 2. Consequently, we have opted to eliminate any rows containing such inaccuracies in one of the three specified attributes. Following this corrective action, the final count of rows in the dataset stands at 3,330,915.

Feature	Number of Negative Values
project_age	13177
account_creation_days	341
first_response_time	3516

Table.2 Quantifying the Occurrence of Negative Values in the Dataset

2.4. Dataset Splitting

We previously highlighted the presence of the 'team_size' factor in our dataset, categorized as small, large, and medium teams. Given the acknowledged impact of team size on choosing suitable software engineering methodologies, we sought to investigate whether this factor influences the dynamics of pull-based development systems under our study. Consequently, we partitioned the dataset into three distinct subsets, each corresponding to a specific team size. In Table 3, we present the detailed statistics for each of these datasets. Additionally, Tables 4, 5, and 6 provide comprehensive insights into the statistics of missing values within each respective dataset.

Dataset	No. rows	merged	merged %	not merged	not merged%
Large Projects	275459	182837	66%	92622	34%
Medium Projects	852376	639233	74%	213143	26%
Small Projects	2478539	2059213	83%	419326	17%

Table. 3 Datasets Statistics After Splitting

Attribute	Type	Number of Missing Values
contrib_gender	Human Factor	535277
contrib_open/cons/neur/extra/agree	Human Factor	541988
first_response_time	Human Factor	1385004
social_strength	Human Factor	247
inte_open/cons/neur/extra/agree	Human Factor	258706
perc_contrib_neg/pos/neu	Human Factor	5809
perc_inte_neg/pos/neu	Human Factor	2014
ci_exists	Pull Request Factor	1338
Sloc	Project Factor	1530
perc_external_contribs	Project Factor	4020
test_lines_per_kloc	Project Factor	1530
test_cases_per_kloc	Project Factor	1591
pushed_delta	Project Factor	20229
pr_succ_rate	Project Factor	10044

Table. 4 Missing values across features for the small projects' dataset

Attribute	Type	Number of Missing Values
contrib_gender	Human Factor	51121
contrib_open/cons/neur/extra/agree	Human Factor	30468
first_response_time	Human Factor	69324
inte_open/cons/neur/extra/agree	Human Factor	44219
perc_contrib_neg/pos/neu	Human Factor	2672
perc_inte_neg/pos/neu	Human Factor	1015
ci_exists	Pull Request Factor	640
Sloc	Project Factor	3
test_lines_per_kloc	Project Factor	3
test_cases_per_kloc	Project Factor	5
pushed_delta	Project Factor	16
pr_succ_rate	Project Factor	10

Table. 5 Missing values across features for the large projects' dataset

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Attribute	Type	Number of Missing Values
contrib_gender	Human Factor	154870
contrib_open/cons/neur/extra/agree	Human Factor	88990
first_response_time	Human Factor	313011
inte_open/cons/neur/extra/agree	Human Factor	79763
perc_contrib_neg/pos/neu	Human Factor	5438
perc_inte_neg/pos/neu	Human Factor	1740
ci_exists	Pull Request Factor	1324
Sloc	Project Factor	517
test_lines_per_kloc	Project Factor	517
test_cases_per_kloc	Project Factor	523
pushed_delta	Project Factor	156
pr_succ_rate	Project Factor	77

Table .6 Missing values across features for the medium projects' dataset

2.5. Datasets Balancing

We observed from Table 3 that the distribution between the two classes, 'merged' and 'not_merged,' is highly imbalanced, potentially introducing bias into the dataset. While certain datasets might naturally exhibit such imbalances, as seen in those focusing on rare diseases where the number of cases is inherently limited, our dataset does not fall under such circumstances. To address this imbalance, we have implemented the undersampling technique, ensuring a more equitable representation. Specifically, we aimed for a balanced distribution with 60% for the dominant class and 40% for the other class, mitigating potential bias and enhancing the dataset's suitability for analysis.

2.6. Datasets imputation

Given the presence of numerous outliers in our dataset, as indicated in figure 6 as an example through the Interquartile Range (IQR) method, traditional imputation methods may not yield optimal results. Utilizing an inappropriate method could introduce bias to the dataset and impact the relationships between different attributes. To address this, we have opted for a more advanced imputation algorithm known as MissForest, which relies on Random Forests. This method has demonstrated superior performance compared to alternatives like K Nearest Neighbor (KNN) and Multivariate Imputation using Chained Equations (MICE) [7]. The selection of Miss Forest aligns with the unique characteristics of our data. Following the application of this algorithm, our datasets will be devoid of any missing values.

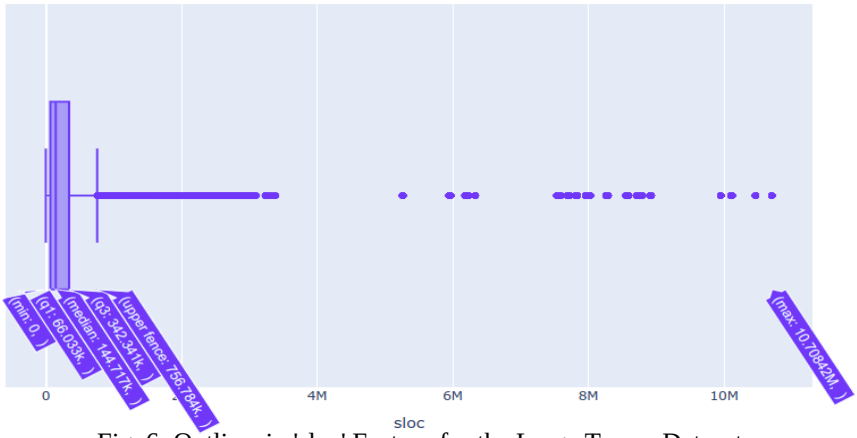

Fig. 6: Outliers in 'sloc' Feature for the Large Teams Dataset

3. Extracting Statistically Significant Features

To determine the appropriate statistical tests for feature comparison, we need to consider the types of features and their distributions. Our features can be categorized into numerical (continuous, discrete) and categorical (ordinal, nominal) data types [8]. Out of the 66 features remaining in our dataset, 10 are Nominal features:

1. first_pr
2. core_member
3. test_inclusion

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

4. ci_exists
5. at_tag
6. comment_conflict
7. hash_tag
8. contrib_gender
9. contrib_follow_integrator
10. language

For comparing Nominal features across the three datasets, we will employ the Chi-Square Test [9]. On the other hand, for Numerical features, we will use the Mann-Whitney U test [10]. It's noteworthy that from statistical tests, we derive two values: Statistical Significance and Practical Significance. The latter holds more importance for us, and to derive it from Statistical Significance, we employ what is known as "effect size."
For the Chi-Square Test, to calculate practical significance, we will use the measure Phi [11] and the measure Odds Ratio [12] when the chi-square table is of size 2x2; otherwise, we will use Cramer's V [11]. For the Mann-Whitney test, we will use Cliff's Delta [13] to derive practical significance, similar to the method used to evaluate features between completed and abandoned pull requests [1].
It's crucial to note that the Mann-Whitney U test is deemed suitable for our data distribution, as examined through the Shapiro-Wilk test [14] and by visually inspecting Histogram diagrams, revealing that the distribution of our datasets is not normal.

Effect Size	**Values**
Negligible	Phi < 0.1
Small	Phi >= 0.1 and Phi < 0.3
Medium	Phi >= 0.3 and Phi <= 0.5
Large	Phi > 0.5

Table.7 Effect Size using Phi measure for Chi-Square test

df*	**Negligible**	**Small**	**medium**	**large**
1	V < 0.1	V >= 0.1 and V < 0.3	V >= 0.3 and V < 0.5	V >= 0.50
2	V < 0.07	V >= 0.07 and V < 0.21	V >= 0.21 and V < 0.35	V >= 0.35
3	V < 0.06	V >= 0.06 and V < 0.17	V >= 0.17 and V < 0.29	V >= 0.29
4	V < 0.05	V >= 0.05 and V < 0.15	V >= 0.15 and V < 0.25	V >= 0.25
5	V < 0.04	V >= 0.04 and V < 0.13	V >= 0.13 and V < 0.22	V >= 0.22

Table.8 Effect Size using Cramer's V measure for Chi-Square test

Delta's Value (d)	**Practical Significance**
|d| <= 0.147	Negligible
0.147 < |d| <= 0.33	Small
0.33 < |d| <= 0.474	Medium
0.474 < |d| <= 1	Large

Table.9 Effect Size using Cliff's Delta measure for Mann-Whiteny U test

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

3.1. Implementing Statistical Tests on the Large Teams Dataset

In this section, we will provide a detailed presentation of the results obtained from the statistical tests using the large team's dataset, followed by an analysis and interpretation of these results.

3.1.1. Chi-Square Test Results on the Large Teams Dataset

First Pull Request Attribute

$X^2(1, N = 232622) = 3758.37, p = .0000$

Label attribute	Not Merged	Merged	Total
No	77453	128630	206083
Yes	15169	11370	26539
Total	92622	140000	232622

Table. 10 Observed values for first_pr attribute for large teams

Label attribute	Not Merged	Merged	Total
No	82055.092064	124027.907936	206083
Yes	10566.907936	15972.092064	26539
Total	92622	140000	232622

Table. 11 Expected values for first_pr attribute for large teams

Core Member Attribute

$X^2(1, N = 232622) = 5961.96, p = .0000$

Label attribute	Not Merged	Merged	Total
No	33205	29838	63043
Yes	59417	110162	169579
Total	92622	140000	232622

Table.12 Observed values for core_member attribute for large teams

Label attribute	Not Merged	Merged	Total
No	25101.532727	37941.467273	63043
Yes	67520.467273	102058.532727	169579
Total	92622	140000	232622

Table. 13 Expected values for core_member attribute for large teams

If the Contributor Follows the Integrator Attribute

$X^2 (1, N = 232622) = 884.29, p = .0000$

Label attribute	Not Merged	Merged	Total
No	90172	132775	222947
Yes	2450	7225	9675
Total	92622	140000	232622

Table. 14 Observed values for contrib_follow_integrator attribute for large teams

Label attribute	Not Merged	Merged	Total
No	88769.751	134177.248927	222947
Yes	3852.248927	3852.248927	9675
Total	92622	140000	232622

Table. 15 Expected values for contrib_follow_integrator attribute for large teams

If the Pull Request Contains Tests Attribute

$X^2 (1, N = 232622) = 884.29, p = .0000$

Label attribute	Not Merged	Merged	Total
No	76366	107717	184083
Yes	16256	32283	48539
Total	92622	140000	232622

Table. 16 Observed values for test_inclusion attribute for large teams

Label attribute	Not Merged	Merged	Total
No	73295.45626	110787.54374	184083
Yes	19326.54374	29212.45626	48539
Total	92622	140000	232622

Table. 17 Expected values for test_inclusion attribute for large teams

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

If the Pull Request Contains a link to another Pull Request Attribute

X^2 (1, N = 232622) = 90.15, p = .0000

Label attribute	Not Merged	Merged	Total
No	67427	99380	166807
Yes	25195	40620	65815
Total	92622	140000	232622

Table. 18 Observed values for hash_tag attribute for large teams

Label attribute	Not Merged	Merged	Total
No	66416.753162	100390.246838	166807
Yes	26205.246838	39609.753162	65815
Total	92622	140000	232622

Table. 19 Expected values for hash_tag attribute for large teams

If the Pull Request Uses Continuous Integration Tools Attribute

X^2 (1, N = 232622) = 16657.07, p = .0000

Label attribute	Not Merged	Merged	Total
No	33932	19167	53099
Yes	58690	120833	179523
Total	92622	140000	232622

Table. 20 Observed values for ci_exists attribute for large teams

Label attribute	Not Merged	Merged	Total
No	21142.177344	31956.822656	53099
Yes	71479.822656	108043.177344	179523
Total	92622	140000	232622

Table. 21 Expected values for ci_exists attribute for large teams

If the Pull Request Contains Mentions Attribute

X^2 (1, N = 232622) = 15.45, p = .0000

Label attribute	Not Merged	Merged	Total
No	49283	73327	122610
Yes	43339	66673	110012
Total	92622	140000	232622

Table. 22 Observed values for at_tag attribute for large teams

Label attribute	Not Merged	Merged	Total
No	48819.042997	73790.957003	122610
Yes	43802.957003	66209.042997	110012
Total	92622	140000	232622

Table. 23 Expected values for at_tag attribute for large teams

If the Pull Request Results a Conflict Attribute

X^2 (1, N = 232622) = 112.93, p = .0003

Label attribute	Not Merged	Merged	Total
No	90421	136990	227411
Yes	2201	3010	5211
Total	92622	140000	232622

Table. 24 Observed values for comment-conflict attribute for large teams

Label attribute	Not Merged	Merged	Total
No	90547.16081	136863.83919	227411
Yes	2074.83919	3136.16081	5211
Total	92622	140000	232622

Table. 25 Expected values for comment-conflict attribute for large teams

The Contributor's Gender Attribute

X^2 (1, N = 232622) = 121.61, p = .0000

Label attribute	Not Merged	Merged	Total
Male	84325	125514	209839
Female	8297	14486	22783
Total	92622	140000	232622

Table. 26 Observed values for contrib_gender attribute for large teams

Label attribute	Not Merged	Merged	Total
Male	83550.600794	126288.399206	209839
Female	9071.399206	13711.600794	22783
Total	92622	140000	232622

Table. 27 Expected values for contrib_gender attribute for large teams

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

The Pull Request's Language Attribute

X^2 (4, N = 232622) = 20098.78, p = .0000

Label attribute	Not Merged	Merged	Total
Go	7260	23129	30389
Java	13686	24417	38103
Javascript	12881	9930	22811
Python	31671	57360	89031
Ruby	10086	20103	30189
Scala	17038	5061	22099
Total	92622	140000	232622

Table. 28 Observed values for contrib_gender attribute for large teams

Label attribute	Not Merged	Merged	Total
Go	12099.844202	18289.16	30389
Java	15171.29	22931.708953	38103
Javascript	9082.55	13728.452167	22811
Python	35449.05	53581.948397	89031
Ruby	12020.21	18168.788851	30189
Scala	8799.054165	13299.95	22099
Total	92622	140000	232622

Table. 29 Expected values for contrib_gender attribute for large teams

In the upcoming table, we will present the effect sizes derived from the previous results:

Attribute	Criteria	Chi-Square X^2	Effect Size	Odds Ratio	Practical Significance
first_pr	Phi	3758.37	0.13	0.45	Small
core_member	Phi	5961.96	0.16	2.06	Small
contrib_follow_integrator	Phi	884.29	0.06	2	Negligible
test_inclusion	Phi	1023.98	0.06	1.04	Negligible
hash_tag	Phi	90.15	0.01	1.09	Negligible
ci_exists	Phi	16657.07	0.26	3.46	Small
at_tag	Phi	15.45	0	1.03	Negligible
comment_conflict	Phi	112.93	0.02	0.9	Negligible
contrib_gender	Phi	121.61	0.02	1.17	Negligible
language	Cramer's V	20098.78	0.14	-	Small

Table. 30 Effect Size of the Nominal Features in the Large Teams Dataset

3.1.2. Mann-Whitney U Test Results on the Large Teams Dataset

We won't provide detailed Histogram diagrams and Shapiro-Wilk results for our dataset; we will only state that we observed a non-normal data distribution.

The next table will consolidate the results of the Mann-Whitney U test, along with the corresponding effect sizes and practical significance.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Attribute	U value	p-value	Delta	Significance
prev_pullreqs	5096854327.5	p-value < 0.0001 ***	-0.21	Small
requester_succ_rate	5859085833	p-value < 0.0001 ***	-0.09	Negligible
followers	6561710142	p-value < 0.0001 ***	0.01	Negligible
acc_commit_num	5309279559	p-value < 0.0001 ***	-0.18	Small
contrib_open	6241400956	p-value < 0.0001 ***	-0.04	Negligible
contrib_cons	6875339097	p-value < 0.0001 ***	0.06	Negligible
contrib_extra	7029868937	p-value < 0.0001 ***	0.08	Negligible
contrib_agree	7670967057	p-value < 0.0001 ***	0.18	Small
contrib_neur	6055435196	p-value < 0.0001 ***	-0.06	Negligible
prior_interaction	5328761247.5	p-value < 0.0001 ***	-0.17	Small
first_response_time	6823967635.5	p-value < 0.0001 ***	0.05	Negligible
prior_review_num	5184066570	p-value < 0.0001 ***	-0.2	Small
account_creation_days	5604444576	p-value < 0.0001 ***	-0.13	Negligible
social_strength	5151581398.5	p-value < 0.0001 ***	-0.2	Small
inte_open	5020088482	p-value < 0.0001 ***	-0.22	Small
inte_cons	6033079364.5	p-value < 0.0001 ***	-0.06	Negligible
inte_extra	7079265629	p-value < 0.0001 ***	0.09	Negligible
inte_agree	8095565565	p-value < 0.0001 ***	0.24	Small
inte_neur	6547683450.5	p-value < 0.0001 ***	0	Negligible
perc_contrib_pos_emo	6745418354	p-value < 0.0001 ***	0.04	Negligible
perc_contrib_neg_emo	6672698756.5	p-value < 0.0001 ***	0.02	Negligible
perc_contrib_neu_emo	6615292926	p-value < 0.0001 ***	0.02	Negligible
perc_inte_pos_emo	6701803703.5	p-value < 0.0001 ***	0.03	Negligible
perc_inte_neg_emo	6411696934.5	p-value < 0.0001 ***	-0.01	Negligible
perc_inte_neu_emo	6393473345	p-value < 0.0001 ***	-0.01	Negligible
num_commits	6523075424.5	p-value < 0.0001 ***	0	Negligible
src_churn	6684042601	p-value < 0.0001 ***	0.03	Negligible
test_churn	6545368411.5	p-value < 0.0001 ***	0	Negligible
files_added	6516888291	p-value < 0.0001 ***	0	Negligible

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

files_deleted	6520771222	p-value < 0.0001 ***	0	Negligible
files_modified	6337237822.5	p-value < 0.0001 ***	-0.02	Negligible
src_files	6743597633	p-value < 0.0001 ***	0.04	Negligible
doc_files	6539772179	p-value < 0.0001 ***	0	Negligible
other_files	6220635361.5	p-value < 0.0001 ***	-0.04	Negligible
num_issue_comments	7393163005.5	p-value < 0.0001 ***	0.14	Negligible
churn_addition	6712738539.5	p-value < 0.0001 ***	0.03	Negligible
part_num_issue	7464112585	p-value < 0.0001 ***	0.15	Small
part_num_code	6516530505.5	p-value < 0.05 *	0	Negligible
churn_deletion	6592708711.5	p-value < 0.0001 ***	0.01	Negligible
description_length	6689682075	p-value < 0.0001 ***	0.03	Negligible
sloc	6409605668.5	p-value < 0.0001 ***	-0.01	Negligible
team_size	7364703488	p-value < 0.0001 ***	0.13	Negligible
perc_external_contribs	7184248929.5	p-value < 0.0001 ***	0.1	Negligible
commits_on_files_touched	6379399571.5	p-value < 0.0001 ***	-0.01	Negligible
test_lines_per_kloc	7174735321.5	p-value < 0.0001 ***	0.1	Negligible
test_cases_per_kloc	5432822424	p-value < 0.0001 ***	-0.16	Small
watchers	7161848089	p-value < 0.0001 ***	0.1	Negligible
project_age	5073826376	p-value < 0.0001 ***	-0.21	Small
pushed_delta	6607881406.5	p-value < 0.0001 ***	0.01	Negligible
pr_succ_rate	4883108771.5	p-value < 0.0001 ***	-0.24	Small
open_issue_num	5639609068.5	p-value < 0.0001 ***	-0.13	Negligible
open_pr_num	7244094593.5	p-value < 0.0001 ***	0.11	Negligible
fork_num	7173673285	p-value < 0.0001 ***	0.1	Negligible
num_technical_comments	6511207897	p-value < 0.05 *	0	Negligible

Table. 31 Summary of Mann-Whitney U Test on Large Teams Dataset

3.1.3. Interpreting Chi-Square Test Results for the Large Teams Dataset

From Table 30, it becomes evident that features exhibiting practical significance include:

1. **First Pull Request for the Integrator (first_pr):** Notably, the acceptance percentage for the first pull request is lower than that for subsequent pull requests. This observation suggests that experience significantly influences the acceptance rate, with the percentage of acceptance almost doubling for developers with more experience. It's noteworthy that previous studies have found higher chances of acceptance for less experienced developers when comparing identical pull requests [15]. However, in our general case, increased developer experience correlates with a higher acceptance percentage.
2. **The presence of continuous integration tools (ci_exists):** significantly impacts the likelihood of accepting a pull request, nearly tripling the chances. This is attributed to the confidence it instills in project owners that the submitted pull request can seamlessly integrate with the project, enhancing trust and favoring the overall quality of the pull request. Previous studies have emphasized the pivotal role of continuous integration tools as a crucial and necessary element in the pull-based development process [16].

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

3. **Pull requests submitted by project owners (core_member):** there is a notable bias as they are accepted with double the chance compared to external contributors, confirming the results presented in previous study [22]. This bias can be attributed to a fundamental aspect of human nature: a greater inclination to trust individuals we are familiar with than new acquaintances. This emphasizes the importance for developers to cultivate a healthy and strong network of relationships within the programmatic environment. Technical proficiency alone may not suffice, as the development environment is inherently shaped by human nature and traits. One such categorization is the "protective environment," which evaluates the integrity and trustworthiness of developers, in addition to their technical efficiency [17].
4. **The language of the project (language)**: The effect of the used programming language in the project has been discussed in previous studies, one of them encouraged the continuous of investigating this feature, because they suffice in presenting what they've seen in chosen projects without putting the results in a mathematical or statistical formula [18]. A subsequent study examined the identical factor, formulating it in a mathematical framework and establishing a connection with the likelihood of accepting a pull request [19], and they shared their results mentioning that three of them increases the acceptance of a pull request which are Go, Scala, Ruby and another three of them decreases the chance of acceptance which are Python, Java and JavaScript and throughout our dataset we can sort the languages according to their acceptance rate with the following: Java(0.22), Javascript(0.55), Python(0.9), Ruby(1.11), Scala(1.35), Go(1.51) but in our datasets back to projects in 2020, whereas the mentioned research has its data from 2015, and by visiting Github we can see that the popularity of languages changes from year to year as well as the number of projects using these languages, which in our opinion may make the language factor on itself, not an accurate factor, and my give a preference to one language in a false way. Thus, we won't be discussing it any further, and we assure that the dynamics of languages over the years need careful study on its own.
   While our results did not reveal practical significance regarding the contributor's gender, it is noteworthy to discuss the broader context and findings from other studies. Previous research [20] has suggested that pull requests submitted by women have a higher acceptance rate than those submitted by men, with a 1.7 times higher chance of acceptance when gender is not disclosed. Interestingly, the original dataset we utilized in our research [2] mentioned a similar observation: if a contributor's gender is revealed as a woman, it may decrease the chances of accepting the pull request.
   Upon closer examination, we discovered that the referenced research [20] relied on the social media platform Google+ to identify the gender of contributors. They categorized pull requests based on contributor gender and found that women had higher chances of acceptance. In our methodology, the original dataset had a significant amount of missing information about the gender of contributors (694,396 missing values, approximately 20% of the entire gender data). We addressed this by imputing these missing values using random forests. In the large dataset, there were 44,357 missing values. Deleting rows with missing gender data would have altered the Chi-square results and potentially decreased the chances of women. Therefore, our imputation of gender data, where it was not revealed, is comparable to the approach used in the referenced research, confirming that if women reveal their identity, they may have lower chances of acceptance.

Male	Not Merged	Merged	Total
Female	62846	104896	167742
Total	7363	13160	20523
Male	70209	118056	188265

Table. 32 Observed values for contrib_gender attribute for large teams without imputation

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

3.1.4. Interpreting Mann-Whitney U Test Results for the Large Teams Dataset

From Table 31, we identified 12 numerical features that exhibit practical significance. It's important to note that the calculations for delta were derived from the vector of declined pull requests:

1. **Number of previously accepted pull requests (prev_pullreqs):** This feature, analogous to the first_pr attribute, exhibits correlation with contributor experience. Higher acceptance chances align with increased contributor experience.
2. **Number of previously accepted commits (acc_commit_num):** Given that each pull request involves at least one commit, the practical significance of this feature is evident, especially when considered alongside prev_pullreqs.
3. **Agreeableness trait of contributor's personality (contrib_agree):** Contributor agreeableness may foster a cooperative atmosphere within the team. A lower percentage of this trait in contributors is associated with a higher likelihood of pull request acceptance [21].
4. **Number of interactions between contributor and integrator in the past three months (prior_interaction):** Emphasizing the importance of building a social network, higher prior interactions correlate with increased acceptance chances, indicating gained trust from the integrator.
5. **Number of reviews processes the integrator was part of (prior_review_num):** This feature underscores the significance of integrator experience. More experience in evaluating pull requests leads to confident decision-making, affecting acceptance probabilities.
6. **Openness trait in the integrator's personality (inte_open):** Openness correlates with the integrator's acceptance of new ideas or code enhancements. Higher openness increases the likelihood of pull request acceptance.
7. **Number of people that contacted any project team member in the past three months (social_strength):** Increased contact with project team members indicates an active and interactive team, positively influencing pull request acceptance chances.
8. **Agreeableness trait in the contributor's personality (contrib_agree):** Similar to inte_agree, a lower agreeableness trait in contributors is associated with a higher likelihood of pull request acceptance.
9. **Number of commenters on the issue that the submitted pull request tried to solve (part_num_issue):** A lower number of commenters on an issue simplifies the reviewing decision, increasing the chances of pull request acceptance.
10. **Number of test cases per 1000 lines of code (test_cases_per_kloc):** More tests indicate a well-documented and tested project, facilitating pull request evaluation and increasing acceptance chances.
11. **Project Age (project_age):** Older projects, combined with ongoing activity, exhibit higher pull request acceptance probabilities. Project age is a positive indicator, especially when combined with continuous project activity.
12. **The rate of accepting previous pull requests for a project (pr_succ_rate):** Higher acceptance rates for previous pull requests in a project correspond to an increased likelihood of accepting new pull requests submitted to the same project.

3.1.5. A Summary of Findings form the Large Teams Dataset

As previously outlined, each feature can be classified into one of three categories: those related to pull requests, those related to the project, and those related to human factors. It is noteworthy that the most impactful features, constituting 10 out of 16, are associated with human factors. In contrast, four features are linked to the project, and two are specifically tied to the pull request itself. This suggests valuable advice for contributors engaging with large projects and extensive teams. Building a robust social network with team members, acquiring experience before submitting a pull request, monitoring project activity and age, and considering the number of previously merged pull requests are key factors. Moreover, maintaining a flexible mindset, avoiding an overly authoritative or closed approach, and being receptive to change and suggestions for enhancing pull requests emerge as crucial strategies. Additionally, contributors are encouraged to select issues with minimal complications and limited diverse opinions to optimize the chances of acceptance, thus conserving time and energy."

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

3.2. Implementing Statistical Tests on the Medium Teams Dataset

In a similar way to what we did with the large teams' dataset, we will apply the same work but this time on the medium teams' dataset.

3.2.1. Chi-Square Test Results on the Medium Teams Dataset

First Pull Request Attribute

$X^2(1, N = 538143) = 9665.86, p = .0000$

Label attribute	Not Merged	Merged	Total
No	179120	300789	479909
Yes	34023	24211	58234
Total	213143	325000	538143

Table. 33 Observed values for first_pr attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	190078.183656	289830.816344	479909
Yes	23064.816344	35169.183656	58234
Total	213143	325000	538143

Table. 34 Expected values for first_pr attribute for medium teams

Core Member Attribute

$X^2(1, N = 538143) = 17158.10, p = .0000$

Label attribute	Not Merged	Merged	Total
No	77197	65349	142546
Yes	135946	259651	395597
Total	213143	325000	538143

Table. 35 Observed values for core_member attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	56458.380167	86087.619833	142546
Yes	156684.619833	238912.380167	395597
Total	213143	325000	538143

Table. 36 Expected values for core_member attribute for medium teams

If the Contributor Follows the Integrator Attribute

X^2 (1, N = 538143) = 2015.97, p = .0000

Label attribute	Not Merged	Merged	Total
No	204659	302603	507262
Yes	8484	22397	30881
Total	213143	325000	538143

Table. 37 Observed values for contrib_follow_integrator attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	200911.922047	306350.077953	507262
Yes	12231.077953	18649.922047	30881
Total	213143	325000	538143

Table. 38 Expected values for contrib_follow_integrator attribute for medium teams

If the Pull Request Contains Tests Attribute

X^2 (1, N = 538143) = 235.42, p = .0000

Label attribute	Not Merged	Merged	Total
No	169911	253384	423295
Yes	43232	71616	114848
Total	213143	325000	538143

Table. 39 Observed values for test_inclusion attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	167655.002825	255639.997175	423295
Yes	45487.997175	69360.002825	114848
Total	213143	325000	538143

Table. 40 Expected values for test_ inclusion attribute for medium teams

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

If the Pull Request Contains a link to another Pull Request Attribute

X^2 (1, N = 538143) = 626.64, p = .0000

Label attribute	Not Merged	Merged	Total
No	163987	240242	404229
Yes	49156	40620	133914
Total	213143	325000	538143

Table. 41 Observed values for hash_tag attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	160103.507334	244125.492666	404229
Yes	53039.492666	80874.507334	133914
Total	213143	325000	538143

Table. 42 Expected values for hash_tag attribute for medium teams

If the Pull Request Uses Continuous Integration Tools Attribute

X^2 (1, N = 538143) = 14228.09, p = .0000

Label attribute	Not Merged	Merged	Total
No	59826	47975	107801
Yes	153317	277025	430342
Total	213143	325000	538143

Table. 43 Observed values for ci_exists attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	42696.882693	65104.117307	107801
Yes	170446.117307	259895.882693	430342
Total	213143	325000	538143

Table .44 Expected values for ci_exists attribute for medium teams

If the Pull Request Uses Contains Mentions Attribute

X^2 (1, N = 538143) = 1286.23, p = .0000

Label attribute	Not Merged	Merged	Total
No	135006	221226	356232
Yes	78137	103774	181911
Total	213143	325000	538143

Table. 45 Observed values for at_tag attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	141093.27293	215138.727067	356232
Yes	72049.727067	109861.272933	181911
Total	213143	325000	538143

Table. 46 Expected values for at_tag attribute for medium teams

If the Pull Request Results a Conflict Attribute

X^2 (1, N = 538143) = 70.92, p = .0000

Label attribute	Not Merged	Merged	Total
No	208718	319291	528009
Yes	4425	5709	10134
Total	213143	325000	538143

Table. 47 Observed values for comment_conflict attribute for medium teams

Label attribute	Not Merged	Merged	Total
No	209129.213401	318879.786599	528009
Yes	4013.786599	6120.213401	10134
Total	213143	325000	538143

Table. 48 Expected values for comment_conflict attribute for medium teams

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

The Contributor's Gender Attribute
X^2 (1, N = 538143) = 270.86, p = .0000

Label attribute	Not Merged	Merged	Total
Male	194372	291980	486352
Female	18771	33020	51791
Total	213143	325000	538143

Table.49 Observed values for at_tag attribute for medium teams

Label attribute	Not Merged	Merged	Total
Male	192630.07107	293721.92893	486352
Female	20512.92893	31278.07107	51791
Total	213143	325000	538143

Table.50 Expected values for at_tag attribute for medium teams

The Pull Request Language
X^2 (4, N = 538143) = 20098.78, p = .0000

Label attribute	Not Merged	Merged	Total
Go	15897	40193	56090
Java	51711	69143	120854
Javascript	38488	65586	104074
Python	60911	99737	160648
Ruby	23390	38208	61598
Scala	22746	12133	34879
Total	213143	325000	538143

Table.51 Observed values for language attribute for medium teams

Label attribute	Not Merged	Merged	Total
Go	22215.639468	33874.36	56090
Java	47866.8	72987.198570	120854
Javascript	41220.724941	62853.275059	104074
Python	63628.06	97019.937080	160648
Ruby	24397.200213	37200.799787	61598
Scala	13814.571029	21064.43	34879
Total	213143	325000	538143

Table.52 Expected values for language attribute for medium teams

Now, similar to our approach with the large teams' dataset, we will compute the practical significance using the Phi measure for the initial nine attributes and Cramer's V for the language attribute. The results will be presented in the subsequent table:

Attribute	Criteria	Chi-Square X^2	Effect Size	Odds Ratio	Practical Significance
first_pr	Phi	9665.86	0.13	0.42	Small
core_member	Phi	17158.1	0.17	2.26	Small
contrib_follow_integrator	Phi	2015.97	0.06	1.79	Negligible
test_inclusion	Phi	235.42	0.02	1.11	Negligible
hash_tag	Phi	626.64	0.03	0.56	Negligible
ci_exists	Phi	14228.09	0.16	2.25	Small
at_tag	Phi	1286.23	0.04	0.81	Negligible
comment_conflict	Phi	70.92	0.01	0.84	Negligible
contrib_gender	Phi	270.86	0.02	1.17	Negligible
language	Cramer's V	13607.35	0.07	-	Negligible

Table. 53 Effect Size of the Nominal Features in the Medium Teams Dataset

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

3.2.2. Mann-Whitney U Test Results on the Medium Teams Dataset

For the medium teams' dataset, we observed a non-normal data distribution similar to the large teams' dataset. We will proceed directly to presenting the results of the Mann-Whitney U test.

Attribute	U value	p-value	Delta	Practical Significance
prev_pullreqs	27449839208	p-value < 0.0001 ***	-0.2	Small
requester_succ_rate	31684871506	p-value < 0.0001 ***	-0.08	Negligible
followers	34336494424.5	p-value < 0.0001 ***	-0.008	Negligible
acc_commit_num	26958746202	p-value < 0.0001 ***	-0.22	Small
contrib_open	32656094021	p-value < 0.0001 ***	-0.05	Negligible
contrib_cons	35618490249.5	p-value < 0.0001 ***	0.02	Negligible
contrib_extra	36630345660	p-value < 0.0001 ***	0.05	Negligible
contrib_agree	38079874371.5	p-value < 0.0001 ***	0.09	Negligible
contrib_neur	34066092203.5	p-value < 0.0001 ***	-0.01	Negligible
prior_interaction	27563254128.5	p-value < 0.0001 ***	-0.2	Small
first_response_time	36557850063.5	p-value < 0.0001 ***	0.05	Negligible
prior_review_num	36557850063.5	p-value < 0.0001 ***	-0.16	Small
account_creation_days	29605628691	p-value < 0.0001 ***	-0.14	Negligible
social_strength	26914019026.5	p-value < 0.0001 ***	-0.22	Small
inte_open	28641195544	p-value < 0.0001 ***	-0.17	Small
inte_cons	31975809362.0	p-value < 0.0001 ***	0.07	Negligible
inte_extra	36013814415	p-value < 0.0001 ***	0.03	Negligible
inte_agree	39460647860.5	p-value < 0.0001 ***	0.13	Small
inte_neur	35559092508.5	p-value < 0.0001 ***	0.02	Negligible
perc_contrib_pos_emo	35607169149.5	p-value < 0.0001 ***	0.02	Negligible
perc_contrib_neg_emo	35363564622.0	p-value < 0.0001 ***	0.02	Negligible
perc_contrib_neu_emo	34545816341	p-value < 0.0001 ***	0.05	Negligible
perc_inte_pos_emo	35147506546.5	p-value < 0.0001 ***	0.014	Negligible
perc_inte_neg_emo	34408275849.5	p-value < 0.0001 ***	-0.006	Negligible
perc_inte_neu_emo	34545816341	p-value < 0.0001 ***	-0.002	Negligible
num_commits	33965374600	p-value < 0.0001 ***	-0.01	Negligible
src_churn	35531362734.5	p-value < 0.0001 ***	0.02	Negligible
test_churn	35245598314	p-value < 0.0001 ***	0.01	Negligible
files_added	35410682122	p-value < 0.0001 ***	0.02	Negligible
files_deleted	34713307262	p-value < 0.0001 ***	0.002	Negligible

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

files_modified	33760350535	p-value < 0.0001 ***	-0.02	Negligible
src_files	35316523094	p-value < 0.0001 ***	0.01	Negligible
doc_files	34430459261	p-value < 0.0001 ***	-0.005	Negligible
other_files	33826592374.5	p-value < 0.0001 ***	-0.02	Negligible
num_issue_comments	42326968141	p-value < 0.0001 ***	0.2	Small
churn_addition	35792280313	p-value < 0.0001 ***	0.03	Negligible
part_num_issue	41980100465	p-value < 0.0001 ***	0.21	Small
part_num_code	34248389777	p-value < 0.0001 ***	-0.01	Negligible
churn_deletion	34652294316	p-value < 0.0001 ***	0.0004	Negligible
description_length	35962722078.5	p-value < 0.0001 ***	0.03	Negligible
sloc	38050992591.5	p-value < 0.0001 ***	0.09	Negligible
team_size	40955785671.5	p-value < 0.0001 ***	0.18	Small
perc_external_contribs	39502523571	p-value < 0.0001 ***	0.14	Negligible
commits_onfiles_touche	32103977700.5	p-value < 0.0001 ***	-0.07	Negligible
test_lines_per_kloc	38258098065.5	p-value < 0.0001 ***	0.1	Negligible
test_cases_per_kloc	32901091697	p-value < 0.0001 ***	-0.05	Negligible
watchers	39999838332	p-value < 0.0001 ***	0.15	Small
project_age	32733759670.5	p-value < 0.0001 ***	-0.05	Negligible
pushed_delta	34182961648.5	p-value < 0.0001 ***	-0.01	Negligible
pr_succ_rate	26596452118	p-value < 0.0001 ***	-0.23	Small
open_issue_num	31602659623.5	p-value < 0.0001 ***	-0.08	Negligible
open_pr_num	41176340812.5	p-value < 0.0001 ***	0.18	Small
fork_num	41016094644	p-value < 0.0001 ***	0.18	Small
num_technical_comments	34271820467.5	p-value < 0.0001 ***	-0.01	Negligible

Table. 54 Summary of Mann-Whitney U Test on Medium Teams Dataset

3.2.3. Interpreting Chi-Square Test Results for the Medium Teams Dataset

From Table 53, it is evident that the features demonstrating practical significance are:

1. **First Pull Request for the Integrator (first_pr).**
2. **The presence of continuous integration tools (ci_exists).**
3. **Pull requests submitted by project owners (core_member).**

The preceding three features are consistent with those identified in the large teams' dataset, with the exception of the language feature, which exhibited practical significance for large teams but not here. Now, let's delve into the statistics for the 'gender' attribute. In the medium teams' dataset, we encountered 101,993 missing values, constituting approximately 18% of the total samples. These missing values were imputed with 97,683 values assigned as male and 4,310 values as female. If we were to remove these empty rows instead of imputing them, the chi-square table for gender would be as it's in Table. 51, with the odds ratio approaching 1. This reaffirms our earlier findings regarding this feature.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Male	Not Merged	Merged	Total
Female	62846	104896	167742
Total	7363	13160	20523
Male	70209	118056	188265

Table. 55 Observed values for contrib_gender attribute for medium teams without imputation

3.2.4. Interpreting Mann-Whitney U Test Results for the Medium Teams Dataset

From Table 54, we identified 14 numerical features that exhibit practical significance. It's important to note that the calculations for delta were derived from the vector of declined pull requests:

1. **Number of previously accepted pull requests (prev_pullreqs).**
2. **Number of previously accepted commits (acc_commit_num).**
3. **The rate of accepting previous pull requests for a project (pr_succ_rate)**
4. **Number of interactions between contributor and integrator in the past three months (prior_interaction).**
5. **Number of reviews processes the integrator was part of (prior_review_num).**
6. **Openness trait in the integrator's personality (inte_open).**
7. **Number of people that contacted any project team member in the past three months (social_strength).**
8. **Number of commenters on the issue that the submitted pull request tried to solve (part_num_issue)**

The preceding eight features are shared with the large teams' dataset, unlike the following:

1. **Agreeableness of the Integrator (inte_agree):** Similar to the contrib_agree feature, which demonstrated practical significance in the large teams' dataset, this personality trait is influential in the medium teams' dataset, but from the integrator's perspective. A lower level of agreeableness correlates with a higher likelihood of accepting the pull request.
2. **Number of Comments on the Issue the Pull Request Is Trying to Resolve (num_issue_comments):** The fewer the comments, the higher the chance of acceptance. This finding suggests that a high number of comments, while indicating popularity and demand within the coding community, also increases the complexity of the problem. Pull requests with higher comment counts were found to be rejected more frequently This finding was also mentioned in [16].
3. **Team Size (team_size):** Declined pull requests tend to have a higher number of team members. In medium-sized teams, where the number of members ranges from 13 to 30, each increment in team size leads to an increase in communication channels. The communication channels can be represented as a complete graph, where nodes are team members and edges are communication channels. The higher the team size, the more communication channels, potentially leading to increased latency in decision-making and erroneous decisions if not managed carefully.
4. **Number of Users Following the Project on Github (watchers):** More followers correspond to a lower chance of accepting a pull request. This indicates increased pressure on the project and heightened competition in terms of pull request submissions, making it more challenging to get a pull request accepted in projects with a larger following.
5. **Number of Copies Taken from the Project (fork_num):** Similar to the "watchers" feature, a higher number of forks correlates with a lower chance of accepting pull requests. Taking a fork from the project is the initial step in contributing, and a higher number of forks implies more developers willing to contribute, increasing competition.
6. **Number of Opened Pull Requests in the Project (open_pr_num):** A higher number of opened pull requests in the project increases pressure on project members, making it more challenging to discuss pull requests related to an issue and propose edits. Consequently, this may lead to a direct rejection of pull requests that are otherwise ready for merging. The results show that this feature decreases the chance of accepting a pull request.

The features that exhibit practical significance in the large teams' dataset but not in the medium teams' dataset are contrib_agree, test_cases_per_kloc, project_age. Conversely, the features demonstrating practical significance in the medium teams' dataset but not in the large teams' dataset are num_issue_comments, team_size, watchers, inte_agree, open_pr_num, and fork_num.

3.2.5. A Summary of Findings form the Medium Teams Dataset

The most influential factors in the medium teams' dataset were human factors (9 out of 17), which is three less than the large teams' dataset. There is a notable shift in the project factors, with 5 significant factors in this dataset compared to four in the previous one. Moreover, these two sets of project factors are entirely different, sharing only one feature, pr_succ_rate. Additionally, there are 3 significant pull request factors.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Our recommendation for contributors to projects with a medium team size is to prioritize building a social network with team members and gaining experience. Moreover, it's advisable to steer clear of projects with high competition and a substantial number of open pull requests.

3.3. Implementing Statistical Tests on the small Teams Dataset

In this section, we will provide a detailed presentation of the results obtained from the statistical tests using the small teams' dataset, followed by an analysis and interpretation of these results.

3.3.1. Chi-Square Test Results on the small Teams Dataset

First Pull Request Attribute

$X^2(1, N = 1048575) = 39693.93, p = .0000$

Label attribute	Not Merged	Merged	Total
No	297876	546050	843926
Yes	121450	121450	204649
Total	419326	629249	1048575

Table. 56 Observed values for first_pr attribute for small teams

Label attribute	Not Merged	Merged	Total
No	337486.697543	506439.302457	843926
Yes	81839.302457	122809.697543	204649
Total	419326	629249	1048575

Table. 57 Expected values for first_pr attribute for small teams

Core Member Attribute

$X^2(1, N = 1048575) = 66364.76, p = .0000$

Label attribute	Not Merged	Merged	Total
No	234288	192802	427090
Yes	185038	436447	621485
Total	419326	629249	1048575

Table. 58 Observed values for core_member attribute for small teams

Label attribute	Not Merged	Merged	Total
No	170793.640264	256296.359736	427090
Yes	248532.359736	372952.640264	621485
Total	419326	629249	1048575

Table. 59 Expected values for core_member small for medium teams

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

If the Contributor Follows the Integrator Attribute
X^2 (1, N = 1048575) = 2441.60, p = .0000

Label attribute	Not Merged	Merged	Total
No	396758	579695	976453
Yes	22568	49554	72122
Total	419326	629249	1048575

Table. 60 Observed values for contrib_follow_integrator attribute for small teams

Label attribute	Not Merged	Merged	Total
No	390484.35322	585968.64678	976453
Yes	28841.64678	43280.35322	72122
Total	419326	629249	1048575

Table. 61 Expected values for contrib_follow_integrator attribute for small teams

If the Pull Request Contains Tests Attribute
X^2 (1, N = 1048575) = 289.37, p = .0000

Label attribute	Not Merged	Merged	Total
No	338738	498776	837514
Yes	80588	130473	211061
Total	419326	629249	1048575

Table. 62 Observed values for test_inclusion attribute for small teams

Label attribute	Not Merged	Merged	Total
No	334922.533499	502591.466501	837514
Yes	84403.466501	126657.533499	211061
Total	419326	629249	1048575

Table. 63 Expected values for test_inclustion attribute for small teams

If the Pull Request Contains a link to another Pull Request Attribute
X^2 (1, N = 1048575) = 359.75, p = .0000

Label attribute	Not Merged	Merged	Total
No	127234	161386	288620
Yes	29292	467863	759955
Total	41326	629249	1048575

Table. 64 Observed values for hash_tag attribute for small teams

Label attribute	Not Merged	Merged	Total
No	115419.374027	173200.625973	288620
Yes	303906.625973	456048.376402	759955
Total	419326	629249	1048575

Table.65 Expected values for hash_tag attribute for small teams

The Contributor's Gender Attribute
X^2 (1, N = 1048575) = 482.17, p = .0000

Label attribute	Not Merged	Merged	Total
Male	390562	578805	969367
Female	28764	50444	79208
Total	419326	629249	1048575

Table. 66 Observed values for comment_conflict attribute for small teams

Label attribute	Not Merged	Merged	Total
Male	387650.656026	581716.343974	969367
Female	31675.343974	47532.656026	79208
Total	419326	629249	1048575

Table. 67 Observed values for comment_conflict attribute for small teams

If the Pull Request Results a Conflict Attribute
X^2 (1, N = 1048575) = 516.02, p = .0000

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Label attribute	Not Merged	Merged	Total
No	413682	623708	1037390
Yes	5644	5541	11185
Total	419326	629249	1048575

Table. 68 Observed values for at_tag attribute for small teams

Label attribute	Not Merged	Merged	Total
No	414853.109353	622536.890647	1037390
Yes	4472.890647	6712.109353	11185
Total	419326	629249	1048575

Table. 69 Observed values for at_tag attribute for small teams

The Pull Request Language: X^2 (4, N = 1048575) = 2645.13, p = .0000

Label attribute	Not Merged	Merged	Total
Go	25592	48662	74254
Java	86799	114557	201356
Javascript	138250	208910	347160
Python	110354	168882	279236
Ruby	48478	67753	116231
Scala	9853	20485	30338
Total	419326	629249	1048575

Table.70 Observed values for language attribute for small teams

Label attribute	Not Merged	Merged	Total
Go	29694.235323	44559.76	74254
Java	80522.429064	120833.570936	201356
Javascript	138829.567899	208330.432101	347160
Python	111666.7	167569.295247	279236
Ruby	46480.871951	69750.128049	116231
Scala	12132.191010	18205.81	30338
Total	419326	629249	1048575

Table.71 Expected values for language attribute for small teams

Now, similar to our approach with the previous two datasets, we will compute the practical significance using the Phi measure for the initial nine attributes and Cramer's V for the language attribute. The results will be presented in the subsequent table:

Attribute	Criteria	Chi-Square X^2	Effect Size	Odds Ratio	Significance
first_pr	Phi	39693.93	0.19	0.55	Small
core_member	Phi	66364.76	0.25	2.87	Small
contrib_follow_integrator	Phi	2441.6	0.04	1.5	Negligible
test_inclusion	Phi	289.37	0.01	1.09	Negligible
hash_tag	Phi	359.75	0.01	1.1	Negligible
ci_exists	Phi	2780.43	0.05	1.26	Negligible
at_tag	Phi	4450.34	0.06	0.71	Negligible
comment_conflict	Phi	516.02	0.02	0.65	Negligible
contrib_gender	Phi	482.17	0.02	1.18	Negligible
language	Cramer's V	2645.13	0.02	-	Negligible

Table. 72 Effect Size of the Nominal Features in the Small Teams Dataset

3.3.2. Mann-Whitney U Test Results on the Small Teams Dataset

We won't provide detailed Histogram diagrams and Shapiro-Wilk results for the small teams' dataset, as we observed a non-normal data distribution similar to the large and medium teams' datasets. We will proceed directly to presenting the results of the Mann-Whitney U test.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Attribute	U value	p-value	Delta	Practical Significance
prev_pullreqs	102150431292	p-value < 0.0001 ***	-0.22	Small
requester_succ_rate	113460150084	p-value < 0.0001 ***	-0.13	Negligible
followers	127205488722	p-value < 0.0001 ***	-0.03	Negligible
acc_commit_num	93380074635.5	p-value < 0.0001 ***	-0.29	Small
contrib_open	126653570039	p-value < 0.0001 ***	-0.03	Negligible
contrib_cons	133419170479	p-value < 0.0001 ***	0.01	Negligible
contrib_extra	136143806700	p-value < 0.0001 ***	0.03	Negligible
contrib_agree	139926716244.5	p-value < 0.0001 ***	0.06	Negligible
contrib_neur	132399519867	p-value < 0.0001 ***	0.003	Negligible
prior_interaction	96951468765	p-value < 0.0001 ***	-0.26	Small
first_response_time	167477032939	p-value < 0.0001 ***	0.26	Small
prior_review_num	107009324015.5	p-value < 0.0001 ***	-0.18	Small
account_creation_days	104058707135	p-value < 0.0001 ***	-0.21	Small
social_strength	99569724099.5	p-value < 0.0001 ***	-0.24	Small
inte_open	126217220136	p-value < 0.0001 ***	-0.04	Negligible
inte_cons	134600342259.0	p-value < 0.0001 ***	0.02	Negligible
inte_extra	136122139975	p-value < 0.0001 ***	0.03	Negligible
inte_agree	141137812829	p-value < 0.0001 ***	0.06	Negligible
inte_neur	131571742210.5	p-value < 0.0001 ***	-0.002	Negligible
perc_contrib_pos_emo	133271548766.5	p-value < 0.0001 ***	0.01	Negligible
perc_contrib_neg_emo	134614713366.5	p-value < 0.0001 ***	0.02	Negligible
perc_contrib_neu_emo	137941233156.5	p-value < 0.0001 ***	0.04	Negligible
perc_inte_pos_emo	131769052155	p-value < 0.0001 ***	-0.001	Negligible
perc_inte_neg_emo	132449166535	p-value < 0.0001 ***	0.003	Negligible
perc_inte_neu_emo	137292350812.5	p-value < 0.0001 ***	0.04	Negligible
num_commits	125367186649	p-value < 0.0001 ***	0.04	Negligible
src_churn	125009160535.5	p-value < 0.0001 ***	-0.05	Negligible
test_churn	128880161208.5	p-value < 0.0001 ***	-0.02	Negligible
files_added	135588340268.5	p-value < 0.0001 ***	0.02	Negligible
files_deleted	131091415509	p-value < 0.0001 ***	-0.006	Negligible

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

files_modified	118391809413.5	p-value < 0.0001 ***	-0.10	Negligible
src_files	121881891264	p-value < 0.0001 ***	-0.07	Negligible
doc_files	129744681405.5	p-value < 0.0001 ***	-0.01	Negligible
other_files	131283532680	p-value < 0.0001 ***	-0.004	Negligible
num_issue_comments	167696885274	p-value < 0.0001 ***	0.27	Small
churn_addition	125810761924	p-value < 0.0001 ***	-0.04	Negligible
part_num_issue	165010610527.5	p-value < 0.0001 ***	0.25	Small
part_num_code	129161853839	p-value < 0.0001 ***	-0.02	Negligible
churn_deletion	120583989143.5	p-value < 0.0001 ***	-0.08	Negligible
description_length	147534566111	p-value < 0.0001 ***	0.11	Negligible
sloc	129409424846	p-value < 0.0001 ***	-0.01	Negligible
team_size	124935208917.5	p-value < 0.0001 ***	-0.05	Negligible
perc_external_contribs	153988408518	p-value < 0.0001 ***	0.16	Small
commits_onfiles_touche	112655241588	p-value < 0.0001 ***	-0.146	Negligible
test_lines_per_kloc	140417719172	p-value < 0.0001 ***	0.06	Negligible
test_cases_per_kloc	139070742017	p-value < 0.0001 ***	0.05	Negligible
watchers	148021422066.5	p-value < 0.0001 ***	0.12	Negligible
project_age	137659614069	p-value < 0.0001 ***	0.04	Negligible
pushed_delta	135144461469	p-value < 0.0001 ***	0.02	Negligible
pr_succ_rate	112552708559.5	p-value < 0.0001 ***	-0.146	Negligible
open_issue_num	134051864843.5	p-value < 0.0001 ***	0.01	Negligible
open_pr_num	153773924631.5	p-value < 0.0001 ***	0.16	Small
fork_num	154988057693.5	p-value < 0.0001 ***	0.17	Small
num_technical_comments	129304715496.5	p-value < 0.0001 ***	-0.01	Negligible

Table. 73 Summary of Mann-Whitney U Test on Small Teams Dataset

3.3.3. Interpreting Chi-Square Test Results for the Small Teams Dataset

From Table 72, it is evident that the features demonstrating practical significance are:

1. **Whether this is the first pull request for the contributor (first_pr).**
2. **If the contributor is a part of the team members (core_member).**

The previous two features are common with the previous two datasets, with the absence of one feature which was significant for both large and medium teams—continuous integration tools (ci_exists).

We will continue by sharing the alternative gender's chi-square test table, which had 612,425 missing values, equivalent to around 58% of the total number. Imputed values include 580,662 males and 31,727 females. Removing rows with missing gender values would result in the following table, with the odds ratio (OR) value approximately 1. Thus, we continue to confirm that the chance of accepting a pull request submitted by a female increases if the gender was not revealed.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Male	Not Merged	Merged	Total
Female	62846	104896	167742
Total	7363	13160	20523
Male	70209	118056	188265

Table. 74 Observed values for contrib_gender attribute without imputation for small teams

3.3.4. Interpreting Mann-Whitney U Test Results for the Small Teams Dataset

From Table 73, we identified 12 numerical features that exhibit practical significance. It's important to note that the calculations for delta were derived from the vector of declined pull requests:

1. **Number of previously accepted pull requests (prev_pullreqs).**
2. **Number of previously accepted commits (acc_commit_num).**
3. **Number of interactions between contributor and integrator in the past three months (prior_interaction).**
4. **Number of reviews processes the integrator was part of (prior_review_num).**
5. **Number of people that contacted any project team member in the past three months (social_strength).**
6. **Number of commenters on the issue that the submitted pull request tried to solve (part_num_issue).**
7. **Number of Comments on the Issue the Pull Request Is Trying to Resolve (num_issue_comments).**
8. **Number of Copies Taken from the Project (fork_num).**
9. **Number of Opened Pull Requests in the Project (open_pr_num).**

The previous nine features were in common with either one or both of the previous two datasets, unlike the following features:

1. **Minutes between the moment of opening the pull request and the interaction of the integrator with it (first_response_time)**: According to [16], they found that the closer the first interaction, the faster the decision to be taken towards a certain pull request. We found that this fast interaction increases the chances of the acceptance of the pull request because the contributor would have the chance to fix or edit anything needed and requested by the integrator rather than waiting for a long time, which may lead to neglecting the pull request. This feature may also be looked at as the availability of team members to review pull requests, so the more engaging and available the team is, the higher the chance of accepting a pull request.
2. **Days passed since opening the account of the contributor (account_creation_days)**: The greater the age of the account, the higher the chance of accepting a pull request, as this may indicate the experience of the developer on the GitHub platform, which may also indicate a broader examination of pull requests.
3. **Percentage of external contributions to the project (perc_external_contribs)**: The more the percentage increases, the lower the chance of accepting the pull request. This feature is close to the idea and result of the previously mentioned two features (open_pr_num, fork_num), which indicates pressure on the project owners. However, this feature emerged here additionally because small teams would fall behind due to pressure more than medium and large teams would.

3.3.5. A Summary of Findings form the Small Teams Dataset

The most effective features are Human Factors (9 out of 14) along with two pull request features and three project features. Based on the results shared, we recommend developers contributing to projects with small team members to pay close attention to every detail in their pull request, as they may not get a second chance or room for discussion and enhancements. Additionally, having significant experience on GitHub is preferable over being a newcomer. As observed in the previous two datasets, building a social network within the project environment remains advantageous for the acceptance of a pull request.

3.4. Comparative Analysis of Results Across the Three Datasets

In this section, we will provide a comprehensive overview of features that demonstrated practical significance across all three datasets. We will elucidate whether each feature positively or negatively impacts the likelihood of a pull request being accepted. By 'positive' and 'negative,' we refer to the increase or decrease in the probability of pull request acceptance corresponding to the rise or fall of a numerical feature. For nominal or boolean features, we will explore how their presence or absence influences the likelihood of pull request acceptance.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Attribute	Large Teams Dataset	Medium Teams Dataset	Small Teams Dataset	Type
perc_external_contribs	No Practical Significance	No Practical Significance	negative	Project
fork_num	No Practical Significance	negative	negative	Project
open_pr_num	No Practical Significance	negative	negative	Project
num_issue_comments	No Practical Significance	negative	negative	Pull Request
part_num_issue	negative	negative	negative	Pull Request
account_creation_days	No Practical Significance	No Practical Significance	positive	Human
social_strength	positive	positive	positive	Human
first_response_time	No Practical Significance	No Practical Significance	negative	Human
prior_review_num	positive	positive	positive	Human
prior_interaction	positive	positive	positive	Human
prev_pullreqs	positive	positive	positive	Human
acc_commit_num	positive	No Practical Significance	positive	Human
core_member	positive	positive	positive	Human
first_pr	negative	negative	negative	Human
int_open	positive	positive	No Practical Significance	Human
team_size	No Practical Significance	negative	No Practical Significance	Project
watchers	No Practical Significance	negative	No Practical Significance	Project
pr_succ_rate	positive	positive	No Practical Significance	Project
contrib_agree	negative	No Practical Significance	No Practical Significance	Human
inte_agree	negative	No Practical Significance	No Practical Significance	Human
test_cases_per_kloc	positive	No Practical Significance	No Practical Significance	Project
project_age	positive	No Practical Significance	No Practical Significance	Project
ci_exists	positive	positive	No Practical Significance	Pull Request

Table.75 Unified Analysis of Practical Significant Features Across Datasets

In light of the three conducted experiments, we offer the following recommendations:

1. **Participation in Projects with Popular Issues:** Contributing to projects with popular issues, where numerous developers are engaged in discussions, complicates the reviewing process. This complication is more pronounced in smaller teams, with a larger impact observed in teams of small size (effect size of 0.25), followed by medium (0.21) and large (0.15) teams.
2. **Building Connections with Team Members:** Building connections with project members is beneficial across all team sizes, with a heightened importance in smaller teams. The effect size for social_strength and prior_interaction increases as the team size decreases, suggesting a stronger impact in smaller teams (effect size: 0.24, 0.26 for small teams).
3. **Integrator Experience:** Repository owners should ensure integrators have prior experience, particularly in large teams where prior_review_num exhibited the highest effect size (0.20). Larger teams with more complex issues benefit from integrators with experience.
4. **Contributor Experience:** Contributor experience is crucial, especially in small teams, where features like prev_pullreqs and first_pr demonstrated the highest effect size (0.22, 0.19). Contributors to small teams should pay attention to detail from the moment of submitting a pull request.
5. **Continuous Integration and Testing:** Large projects should rely on continuous integration (CI) tools and testing. Features like test_cases_per_kloc and ci_exists are essential for evaluating pull requests in large projects. However, ci_exists may lack practical significance in small teams, possibly due to lower levels of experience or professionalism.
6. **Project Age:** Project maturity increases with age, particularly in large teams. Contributors participating in mature projects have a higher chance of acceptance.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

7. **Personality Traits:** Contributors in the technical community should embrace openness and avoid pedantic behavior. Features like contrib_agree, inte_agree, and inte_open emphasize the importance of being open to diverse opinions for the benefit of the project and contributors.
8. **Avoiding Pressure in Small and Medium Teams:** Contributors should steer clear of small and medium teams facing high pressure, numerous open pull requests, and comments. Features like watchers, fork_num, perc_external_contribs, num_issue_comments, and open_pr_num indicate potential challenges in such environments.
9. **Monitoring Previous Success Rate:** Contributors should monitor the rate of previously accepted pull requests (pr_success_rate) and the number of previously accepted commits (acc_commit_num) as indicators of their chances of acceptance.
10. **Quick Interaction in Small Projects:** Contributors are encouraged to contribute to small projects with active team members who tend to interact quickly with submitted pull requests (first_response_time).
11. **Optimal Team Size:** In projects with medium-sized teams, increasing team members should be carefully evaluated. Unless the increment transforms the team or has justifications, it may negatively impact team management.

In summary, we identified 23 features with practical significance, with 8 related to projects, 3 to pull requests, and 12 to human factors. Human factors emerge as a dominant force in the pull-based development process, highlighting the significance of personalities and experiences. This section focused on the one-to-one relationship between features and the target. In the next section, we will explore the combined effect of these features on the target value.

4. Analyzing Feature Impact on Pull Request Decision-Making using Machine Learning

In this section, our focus shifts to the application of machine learning algorithms on the features identified with practical significance, as detailed in table 75. The primary aim is to evaluate the individual impact of each feature in predicting the label that indicates whether a pull request will be merged or not. We will employ three distinct machine learning algorithms—Support Vector Machines (SVM) [23], Naive Bayes [24], and Random Forests [25]. These algorithms will be applied to all three datasets to assess the accuracy of predictions using the selected 23 features. Furthermore, we will quantify the contribution of each feature in reaching the prediction outcome.

4.1. Implementation of SVM Algorithm

For each dataset, we determined the optimal C value through a tuning process, selecting the C value that resulted in the highest precision score. The identified values are as follows:

1. For the large teams' dataset: C = 0.1, achieving a precision score of 0.7431488633155823.
2. For the medium teams' dataset: C = 0.001, with a precision score of 0.7135344743728638.
3. For the small teams' dataset: C = 0.001, yielding a precision score of 0.7397226095199585

Next, we utilized SVM to derive the importance of features, represented by their coefficients, which can be tracked during the algorithm's execution. Instead of presenting the order and significance of features along with their weights in tables, we opted for a more visual approach. We will employ matplotlib [26] to create Bar plots for each dataset. These figures will provide a clearer and more accessible way to comprehend and compare.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

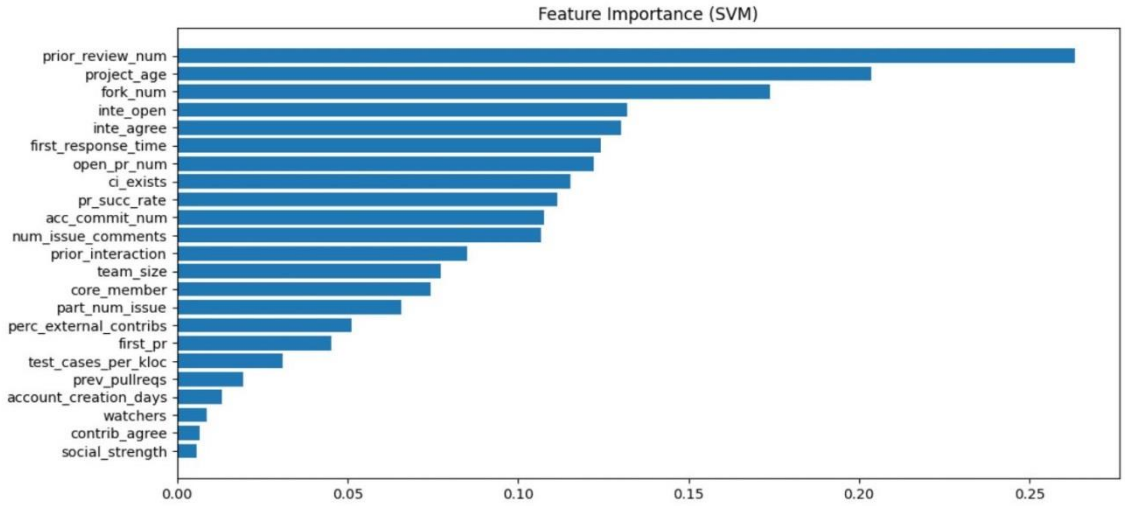

Fig. 7 Feature Importance for Large Teams Dataset as Determined by SVM

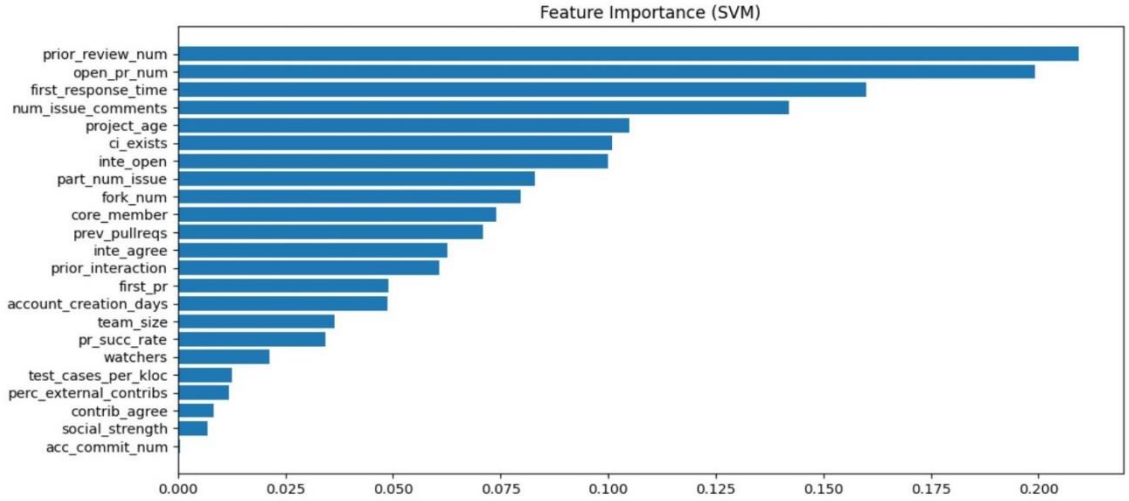

Fig. 8 Feature Importance for Medium Teams Dataset as Determined by SVM

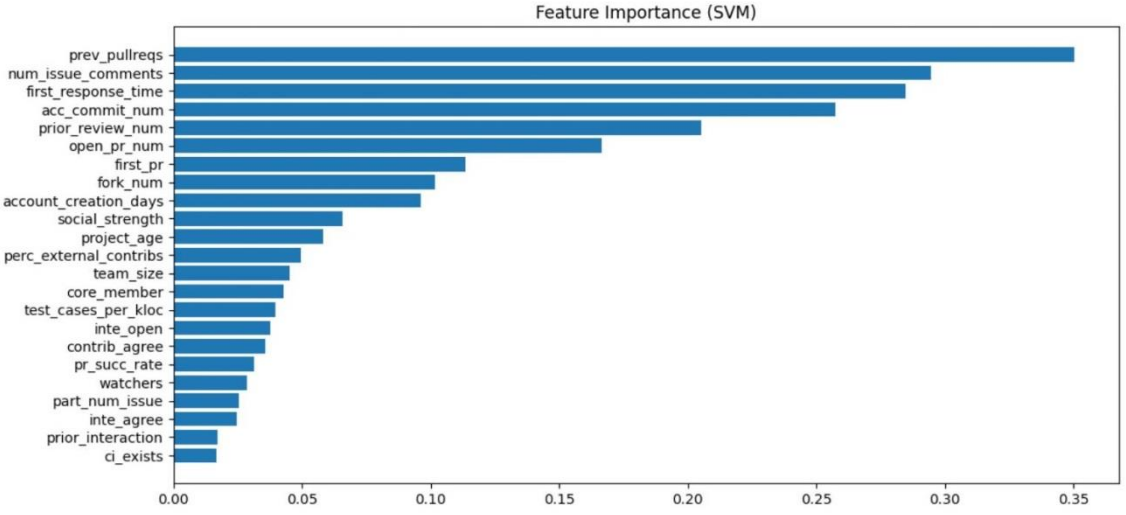

Fig. 9 Feature Importance for Small Teams Dataset as Determined by SVM

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

Noting that feature importance is derived from the coefficients assigned to each feature in the hyperplane equation. Larger absolute values of coefficients suggest a more influential role of the corresponding feature in the classification decision.

4.2. Implementation of Naive Bayes Algorithm

We will apply the Naive Bayes algorithm, specifically using the Multinomial Naive Bayes (MultinomialNB) variant, as it aligns well with the characteristics of our datasets. To ensure robust performance and avoid issues like zero frequencies leading to underfitting or overfitting, we will undergo a tuning process for the Alpha parameter. The optimal Alpha values, along with their corresponding scores for each dataset, are as follows:

1. Alpha = 10 with a score of 0.657667920472864 for the large teams' dataset.
2. Alpha = 1 with a score of 0.662395822687194 for the medium teams' dataset.
3. Alpha = 10 with a score of 0.6717165441633293 for the small teams' dataset.

Next, similar to our approach with SVM, we will extract the feature importance measured by examining the conditional probability of each feature given a specific class label. during the execution of the Naive Bayes algorithm. We will then visualize these results through plots for better understanding and comparison.

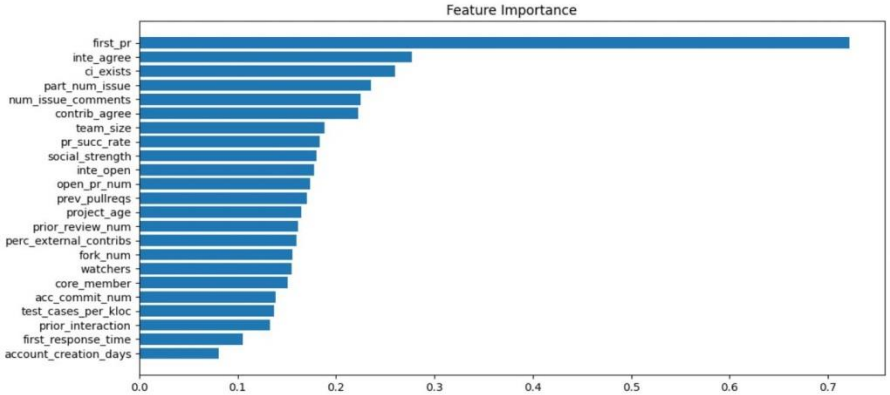

Fig. 10 Feature Importance for Large Teams Dataset as Determined by Naive Bayes

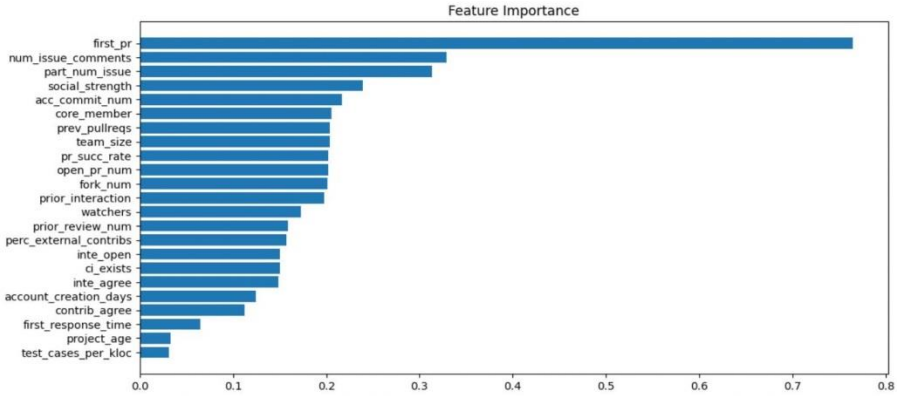

Fig. 11 Feature Importance for Medium Teams Dataset as Determined by Naive Bayes

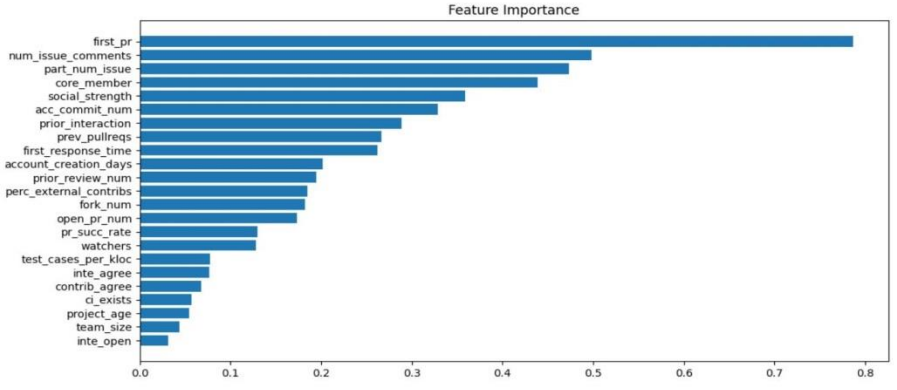

Fig. 12 Feature Importance for Small Teams Dataset as Determined by Naïve Bayes

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

4.3. Implementation of Random Forests Algorithm

For the Random Forests algorithm, we performed tuning for two parameters, namely n_estimators and max_features. Based on the highest score achieved through the combination of these two features, we determined the optimal settings for each dataset:

1. For the large teams' dataset, we found that setting n_estimators to 100 and max_features to log2 resulted in a score of 0.8745.
2. In the case of the medium teams' dataset, the optimal configuration was n_estimators = 100 and max_features = sqrt, yielding a score of 0.8291.
3. For the small teams' dataset, the best combination was n_estimators = 100 and max_features = sqrt, resulting in a score of 0.8148.

It's important to note that the tuning process considered the joint optimization of both parameters. Moving forward, we will proceed to calculate feature importance, which is determined by the mean decrease in impurity caused by each feature across all trees in the forest. Features contributing more significantly to reducing impurity are assigned higher importance.

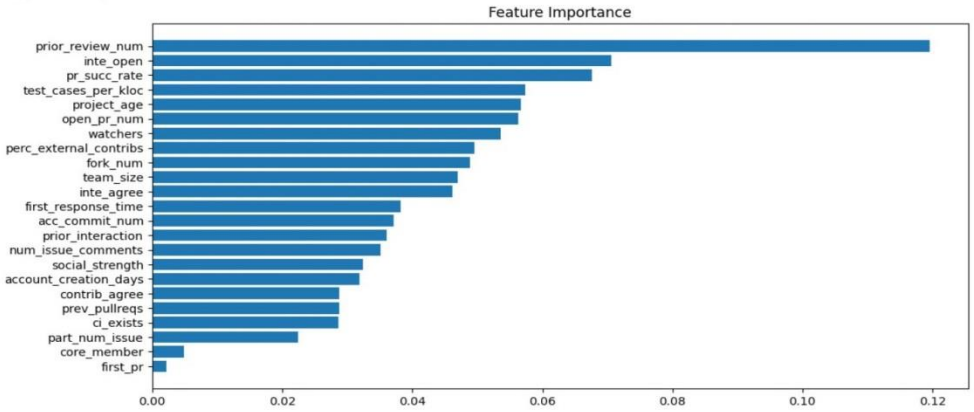

Fig. 13 Feature Importance for Large Teams Dataset as Determined by Random Forests

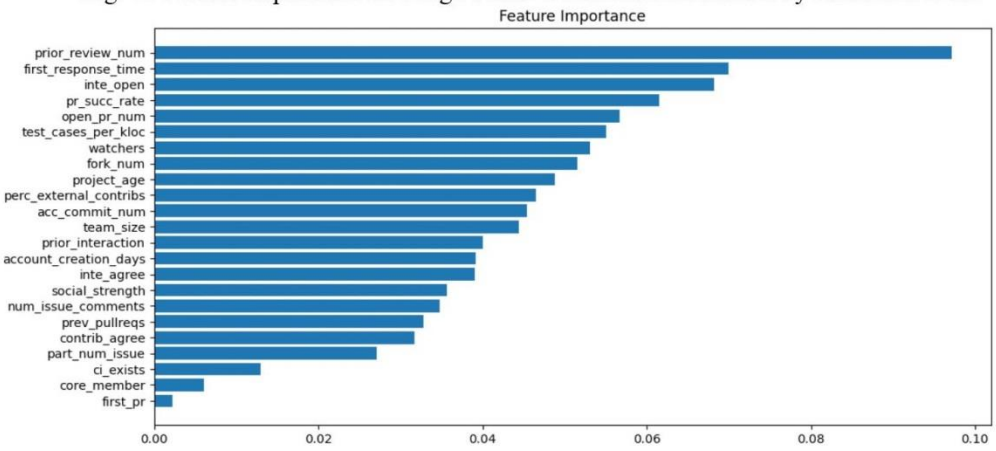

Fig. 14 Feature Importance for Medium Teams Dataset as Determined by Random Forests

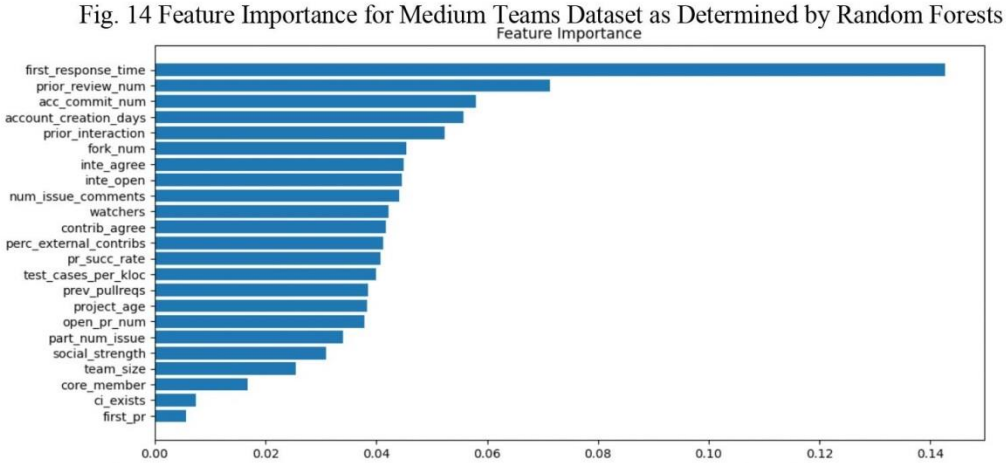

Fig. 15 Feature Importance for Small Teams Dataset as Determined by Random Forests

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

4.4. Synthesizing Feature Importance Insights

In this section, we present three tables—one for each dataset—each containing three columns representing the order of feature importance as determined by the three previously mentioned algorithms. This arrangement facilitates a comparative analysis of the results across the three experiments for each dataset.

Feature	Importance according to SVM	Importance according to Naive Bayes	Importance according to Random Forests
prev_pullreqs	19	12	19
acc_commit_num	10	19	13
contrib_agree	22	6	18
prior_interaction	12	21	14
social_strength	23	9	16
inte_open	4	10	2
inte_agree	5	2	11
part_num_issue	15	4	21
test_cases_per_kloc	18	20	4
project_age	2	13	5
pr_succ_rate	9	8	3
prior_review_num	1	14	1
num_issue_comments	11	5	15
team_size	13	7	10
watchers	21	17	7
open_pr_num	7	11	6
fork_num	3	16	9
first_response_time	6	22	12
account_creation_days	20	23	17
perc_external_contribs	16	15	8
first_pr	17	1	23
core_member	14	18	22
ci_exists	8	3	20

Table.76 Order of Importance for Large Teams Dataset

Feature	Importance according to SVM	Importance according to Naive Bayes	Importance according to Random Forests
prev_pullreqs	11	6	18
acc_commit_num	23	5	11
contrib_agree	21	20	19
prior_interaction	13	12	13
social_strength	22	4	16
inte_open	7	16	3
inte_agree	12	17	15
part_num_issue	8	3	20
test_cases_per_kloc	19	23	6
project_age	5	22	9
pr_succ_rate	17	9	4
prior_review_num	1	14	1
num_issue_comments	4	2	17
team_size	16	7	12
watchers	18	13	7
open_pr_num	2	10	5
fork_num	9	11	8
first_response_time	3	21	2
account_creation_days	15	19	14
perc_external_contribs	20	15	10

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

first_pr	14	1	23
core_member	10	8	22
ci_exists	6	18	21

Table.77 Order of Importance for Medium Teams Dataset

Feature	Importance according to SVM	Importance according to Naive Bayes	Importance according to Random Forests
prev_pullreqs	1	8	15
acc_commit_num	4	6	3
contrib_agree	17	19	11
prior_interaction	22	7	5
social_strength	10	5	19
inte_open	16	23	8
inte_agree	21	18	7
part_num_issue	20	3	18
test_cases_per_kloc	15	17	14
project_age	11	21	16
pr_succ_rate	18	15	13
prior_review_num	5	11	2
num_issue_comments	2	2	9
team_size	13	22	20
watchers	19	16	10
open_pr_num	6	14	17
fork_num	8	13	6
first_response_time	3	9	1
account_creation_days	9	10	4
perc_external_contribs	12	12	12
first_pr	7	1	23
core_member	14	4	21
ci_exists	23	20	22

Table.78 Order of Importance for Small Teams Dataset

Examining the figures presented in this section, we observe variations in the importance of features across datasets for each algorithm. Additionally, there are disparities within the same dataset when considering different algorithms.

Conclusion

This study marks an initial exploration into the impact of team size on pull-based development, specifically within the realm of pull requests in contemporary open-source repositories. Our findings reveal notable distinctions in how teams manage the review process and assess the quality of pull requests based on variations in team size. Just as team size has historically influenced traditional software engineering methodologies, our study affirms its enduring significance in the context of pull-based development projects. We emphasize the pivotal role of human factors in shaping the outcome of pull requests, underscoring the inherently human-driven nature of the pull request management process. This study serves as a foundational step, prompting further research into team sizes in future investigations. We recommend focused studies on specific feature subsets in diverse contexts, as our work aims at a comprehensive understanding. The broad scope of this study lays the groundwork for more targeted and nuanced examinations within the specific context of team size.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden

References

[1]. Khatoonabadi S., Costa E. D., Abdalkareem R., and Shihab E. "*On Wasted Contributions: Understanding the Dynamics of Contributor-Abandoned Pull Requests: A Mixed-Methods Study of 10 Large Open-Source Projects*". ACM Transactions on Software Engineering and Methodology, (2022).

[2]. X. Zhang, A. Rastogi and Y. Yu. "*On the Shoulders of Giants: A New Dataset for Pull-based Development Research*". In Proceedings of the 17th International Conference on Mining Software Repositories, MSR 547, (2020).

[3]. G. Gousios, Pinzger M., and A. van Deursen. "*An exploratory study of the pull-based software development model*". In Proceedings of the 36th International Conference on Software Engineering, ICSE. Association for Computing Machinery (ACM), New York, pp. 345–355, (2014).

[4]. X. Zhang, Y. Yu, G. Gousios and A.Rastogi. "*Pull request decision explained: An empirical overview*". IEEE Transactions on Software Engineering early access (2022).

[5]. N. Keshta and Y. Morgan. "*Comparison between traditional plan-based and agile software processes according to team size project domain*". In 8th IEEE Annual Information Technology. Electronics and Mobile Communication Conference pages 567–575, (2017).

[6]. R. N. Iyer, S. A. Yun, M. Nagappan and J. Hoey, "*Effects of personality traits on pull request acceptance*", IEEE Transactions on Software Engineering early access (2019)

[7]. S. Hong and H. S. Lynn, "*Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and inter-action*", BMC Medical Research Methodology vol. 20, no. 1, pages 1–12, (2020).

[8]. D. J. Rumsey. "Statistics For Dummies", 2nd Edition. John Wiley & Sons, (2016).

[9]. P. Karl. "*On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*". Philosophical Magazine. Series 5. 50 (302): 157–175, (1900).

[10]. H. B. Mann and D. R. Whitney. "*On a test of whether one of two random variables is stochastically larger than the other*". Annals of Mathematical Statistics (1947).

[11]. H. Kim. "*Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test*". Restor Dent Endod. 42(2):152-155. doi: 10.5395/rde.2017.42.2.152. Epub (2017).

[12]. Chinn, S. *"A simple method for converting an odds ratio to effect size for use in meta-analysis*" . Statistics in Medicine (2000).

[13]. M. R. Hess and J. D. Kromrey. "*Robust confidence intervals for effect sizes: A comparative study of Cohen's d and Cliff's delta under non-normality and heterogeneous variances*". Presented at the Annual Meeting of the American Educational Research Association (AERA), (2004)

[14]. S.S. Shapiro and M.B. Wilk. "*An analysis of variance test for normality (complete samples)*". Biometrika, 52(3–4), pp. 591–611, (1965)

[15]. Y. Yu, L. Zhixing, Y. Gang, T. Wang and W. Huaimin. "*A Dataset of Duplicate Pull-requests in GitHub*". In Proceedings of the 15th International Conference on Mining Software Repositories, MSR '18, pages 22–25 (2018).

[16]. Y. Yu, G. Yin, T. Wang, C. Yang, and H. Wang, "*Determinants of pull-based development in the context of continuous integration*". Science China Information Sciences vol. 59 (2016)

[17]. F. Calefato and F. Lanubile. "*Establishing Personal Trust-based Connections in Distributed Teams*". Internet Technology Letters. (2018).

[18]. M. M. Rahman and C. K. Roy. "*An insight into the pull requests of GitHub*". In Proceedings of the 11th Working Conference on Mining Software Repositories, MSR, pages 364–367 (2014).

[19]. D. M. Soares and M. L. de Lima Jánior, L. Murta, and A. Plastino, ''*Acceptance factors of pull requests in open-source projects*" in Proceedings of Symposium on Applied Computing (SAC), pp. 1541–1546. (2015).

[20]. J. Terrell, A. Kofink, J. Middleton, C. Rainear, E. Murphy-Hill, C. Parnin, and J. Stallings. "*Gender differences and bias in open source: Pull request acceptance of women versus men*" . PeerJ Computer Science 3, e111, (2017).

[21]. B. Antoncic, "*The entrepreneur's general personality traits and technological developments*". World Academy of Science, Engineering and Technology. (2009).

[22]. G. Pinto, L. F. Dias, and I. Steinmacher. "*Who Gets a Patch Accepted First? Comparing the Contributions of Employees and Volunteers*" In Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering (2018).

[23]. Y. Zhang. "Support Vector Machine Classification Algorithm and Its Application" in Information Computing and Applications, C. Liu, L. Wang, and A. Yang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 179–186, (2012).

[24]. I. Rish. "*An empirical study of the naive bayes classifier*". IJCAI 2001 workshop on empirical methods in artificial intelligence, 3 (22), 41–46, (2001).

[25]. Ho T. K. "*Random decision forests*". In Proceedings of the 3rd international conference on document analysis and recognition. p. 278–82. (1995)

[26]. J. D. Hunter. "*Matplotlib: A 2D Graphics Environment*", Computing in Science & Engineering, vol. 9, no. 3, pp.

© 2024, IJCSMC All Rights Reserved, ZAIN Publications, Fridhemsgatan 62, 112 46, Stockholm, Sweden