International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 10 Issue 10 - October 2022 - Pages 29-40

QUANTITATIVE ANALYSIS OF TOP GITHUB REPOSITORIES: EXPLORING ACTIVITY, EFFORT, AND PROJECT SUCCESS

Dr. C N Ravi, Meenu Paulose, Prema Mani, Fayisa M K, Thasni K M
Department of Computer Science and Engineering (CSE)
Indira Gandhi Institute of Engineering and Technology
Nellikuzhi P.O, Kothamangalam, Ernakulam (Dist), Kerala, Pincode 68669, India

Abstract:

As of the year 2023, GitHub has compiled a list of the 10 repositories that have the most number of contributors. The authors of the current research believed that it was important to examine these repositories based on certain quantitative factors in order to get some understanding as to the reasons why these initiatives have been successful. Several innovative criteria, including 'activity' and 'effort' of various types, have been attempted to be defined during the course of this article. Additionally, these parameters have been computed for the 10 projects that performed the best. Based on this data, several conclusions have been drawn about the relationship between the quantitative analysis of software projects and the quality of these projects. Additionally, there is the potential for some future work to be done in this subject. Both the relevance of the research and its explanation have been effectively stated, making it suitably fascinating.

Keywords: *GitHub, repositories, activity, and effort*

.Introduction

In the event that we are interested in gaining a quantitative understanding of the expansion of software projects, GitHub is an essential location to search for data. In this particular subject, the repositories on GitHub continue to be an intriguing location to investigate and make conclusions from. When it comes to the field of software development, the year 2020 has been a year that has seen amazing advancements. In light of the fact that governments are encouraging their people to stay at home and work from home, GitHub has seen a significant increase in activity, which has inspired developers from all over the world to work together and participate in creative endeavours. These endeavours include the search for connections and the investigation of potential solutions to issues. Each year, the GitHub team releases a report called "State of Octoverse," which provides an overview of several fascinating statistics and numbers. According to the paper that was published for the year 2020[1,] there are more than fifty-six million developers, and there have been more than sixty million repositories established in the year 2020 alone.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

There is a great deal of information on these repositories on GitHub. Through the use of appropriate data gathering techniques, we are able to retrieve both historical and present data. In addition, there is the Google Big Query website, which makes a significant amount of information on GitHub repositories available to the general public for academics to investigate and draw conclusions from. But for the time being, this paper contains some information on some of the most popular projects on GitHub, which are ranked according to the number of people who have contributed to the project. The study that was conducted on all of those data will serve as the foundation for some future work. Now, if we are able to talk about the data that we can discover regarding these initiatives, then the answer to this issue is that there is a great deal of data that can be found. We are able to determine when a certain project was started, for how long it was operational, how it expanded, how individuals working on the project engaged with one another, and how lines of code were included into it. The process by which forks, pull requests, and commits were created. Investigating the data on GitHub may reveal a great deal of information, including how various projects' branches came into being and how they were included into the master branch. In addition, there are many more things that can be discovered. And how exactly do we go about doing it? GitHub, on the other hand, provides us with a wealth of documentation [2] that explains not only how we may search for data on this platform but also the different kinds of data that we can seek for. The writers of this research have focused mostly on a few ratios throughout their work. There has been an effort made to define some new terminology that pertain to the activity of a specific project as well as the work that is involved in producing a specific project.
These two keywords have been used in different contexts by a number of writers, and there is a good chance that some more appropriate words may be conceived of.

The authors of the current research are of the opinion that there is a high probability that in the future, peers will come up with new terms that are more significant than those that have already been used. There are two categories of activities that have been specified in this article. These categories are the issue activity and the pull request activity. Now, these values are essentially ratios, which may seem to be the same for both tiny data and huge data as well. There are several instances in which ratios might be deceptive statistics due to the fact that they may display the same value for different amounts. This article thus offers some suggestions for avoiding falling into this trap. Then there is also the attempt to describe three different types of activities that are engaged in software projects. These are the commit effort, the fork effort, and the branch effort.

The question of what precautions need to be taken in order to extract the greatest amount of value from these factors is another topic that is discussed in this study. The article not only specifies these values, but it also calculates these values for the top software projects on GitHub. This is done in order to obtain a sense of what these values would be for top projects. After conducting an impartial investigation, the authors have made an effort to compare these numbers for the software projects that are presented below and have drawn conclusions based on their findings. One thing that has been seen is that practically all of the projects have patterns that are comparable to one another, and this fact makes the current study even more intriguing. Additionally, this work emphasises the necessity of

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

such a research and provides a hint at the prospect of establishing a connection between the qualitative analysis of software projects and the quantitative study of software development. Specifically, it provides some observations on the potential for engineering approaches to be improved by the use of such a research. The article covers the future scope of this study and points out the areas that might assist to deepen the knowledge of the criteria that are specified here. This discussion comes after the conclusions have been reached throughout the text.

2 Literature Review

The topic of software repositories on GitHub has been the subject of writing by a number of different writers. Many people have brought attention to the enormous potential that this platform has. The initial purpose for which this platform was built was for collaborative development. Later on, version control was also added, and this resulted in GitHub being incredibly popular and beneficial, particularly for the creation of open-source software. There has been a debate on the involvement in the development process of open-source software projects in the research that was conducted by Padhey et al. [3]. This study is the most important work that has to be mentioned. This article has not only shown that such data is readily accessible on Github, but it has also offered the fundamental impetus to investigate projects hosted on an online platform. It has been argued that the number of forks and pulls may serve as the foundation for quantitative study of the growth of software. This work has reviewed the fork and pull model and has recommended that this model can be used. Having historical data on forks and pulls, which may lead to a possible analysis of the various projects in terms of their increase in number, is the only element that has the potential to make the research more comprehensive. The incentive to investigate the ratio between the number of open problems and the number of closed issues was supplied by this publication, which played a role in providing the motivation.

There is still another parameter that has been explored in this study, and it is also influenced by Padhey et al. [3]. This parameter is about the ratio of the pull requests that are open to the pull requests that are closed. Within the scope of this article, we will be discussing the many ways in which this data might be understood. An additional characteristic that has been examined in this context is the average pace at which pull requests are sent and eventually closed. It is important to note that the primary concept that can be attributed to the use of pull requests on Github is something that has been taken from the work that Padhey et al. have done. [3] despite the fact that the authors of the current study have presented an interpretation that is quite unusual and seeks to bring the existing state of the art up to date.

It goes without saying that understanding the behavior of the community and the involvement of people in the projects can help understand the health of the project and that has been acknowledged in the paper published by Hata et. al. [4] which stresses on the fact that there are studies that show that ‘sustainable open-source communities’ have to be explored in order to better understand the organi9sational structure of open source software projects to drive them forward.

Tamburi et al. [5] has showed that there is the potential for a significant amount of research to be conducted in the areas of best practices for organisation, research on social networks, and also by

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

taking into consideration various sorts of models, various types of theories, and the features of various open source software projects. In addition, there has been an emphasis placed on the "social aspects" of software projects, such as the participation of members (Gamaleilsson et al. [6] and Schweik et al. [7]). The existence of these works demonstrates that there was a period of time when individuals had a sense that there had to be a means of comprehending and evaluating this activity. It should come as no surprise why this is the case. The authors of the present article have often and confidently asserted that the expansion of software projects, and more specifically, the growth of open-source software projects, must be considered as connected to the quality of the projects. This is the fundamental notion that was built up, and it was what the authors of the present study have claimed.

According to the opinions of a great number of engineers and analysts, this socialisation did not place a significant amount of significance on the overall quality of the software project. However, this concept has faced significant obstacles due to the gradual collapse of open-source software projects like Softonic. This has occurred as a result of a lack of increase in the number of volunteers who could steer the project. Perhaps this could be determined by measuring the activity, which is something that the authors have attempted to do in the present paper.

An effort has been made by Tamburi et al. [8] to propose an automated tool that is known as the Yoshi. Yoshi is an acronym that stands for delivering open-source health insurance information. What is often referred to as the open-source community health status may be measured with the use of this instrument. In addition to this, it "associates a community pattern of organisational structure types" [8]. However, the current state of affairs does not adequately highlight the activities of the organisation. The authors of the present paper are of the opinion that without the quantitative measurement of the activities of those involved and the analysis of those activities, very little can be accomplished in terms of actually realising in the direction of quantitative growth. As a consequence of this, the current article concentrates on the submission of patches and the number of commits as a measurement of one kind of activity. It does so with more depth. Nevertheless, there are more forms, such as the documentation and the correction of existing bugs. In addition, these activities do not help in any manner, shape, or form to the increase of the quality of the software project. As a result, the authors have made the decision to make these criteria topics of future study and research.

3 Methodology

An analysis of the document titled "The State of the Octoverse," which was released by GitHub [9], reveals a great deal of fascinating information that can be used for the purpose of analysing the quantitative growth of software projects that are stored in GitHub repositories. On an annual basis, this paper is published by GitHub. As we go through the document that was produced in 2019, we have compiled a list of the twenty software projects that have received the greatest contributions in terms of the total number of contributors. Ten projects are found to have more than twenty thousand contributors, according to our findings. It is evident from this that the amount of people that

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

contribute to software projects has a significant impact on the level of success that such projects achieve. The writers made the decision to take this list into consideration and investigate the GitHub website in order to uncover more information that is intriguing about these software repositories.

The website GitHub itself is a rich source of data on a variety of repositories, and there are strategies that have been well described that can be used to get all of this data from the hosting site. As a result, the documentation that GitHub provides served as a source of inspiration for more data collecting and additional research.

The authors proceeded to collect data about the ten software projects that were mentioned in the document titled State of Octoverse that was published for the year 2019. They did this by working according to the methods that are available on the documentation of GitHub, which are about how data can be collected about various software repositories.

During the course of the investigation of these software projects, information on the following criteria was gathered: open and closed problems, open and closed pull requests, the total number of months that the software project has been working for, the total number of commits, the total number of forks, the total number of branches, and so on. On the basis of all of this information that was gathered, the following ratios were formed and derived:

1. The proportion of a project's issue count that has been resolved in comparison to the overall number of problems for that project
2. The proportion of pull requests that have been terminated in comparison to the total number of pull requests for each project
3. The proportion of the total number of individual contributors to the total number of forks in each project
4. The proportion of the total number of contributors to the total number of individuals working on each project
5. The proportion of the total number of contributors to the total number of commits in each project

In order to fulfil this requirement, it is essential to examine the interpretation about the various amounts that were discussed before. The ratios may be understood in the following way, according to the following:

1. The ratio of the number of problems that have been resolved to the total number of issues that have been presented provides insight into the functioning of the software project in relation to the participants who will ultimately utilise it. As far as we are aware, the majority of the problems that have been reported are caused by the end users of the product. It is not the case that the developers do not present problems of their own. However, the most of the time, the developers would go about working on the code itself and submitting modifications. As a result, it would not be an extremely incorrect assumption to suppose that the topic about difficulties that are filed in a software project is substantially tied to the activity of the end users. Now it is quite usual that a really active project

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

would have a high number of end users and it would also be a very excellent factor behind the popularity of the project. On the other hand, the greater the number of users, the greater the quantity of proposals that are received. It goes without saying that not all problems are bugs, and a significant number of problems would be resolved without the need for further investigation. Nevertheless, it goes without saying that a successful project would have individuals working on it who would promptly attend to all of the complaints that were raised and answer to them as quickly as possible. As a result, we are able to assert that this ratio is, in a sense, an indicator of the activity that the software project is experiencing in connection to its end customers. Possibly, we could refer to this parameter as the problem activity.

2. The following ratio includes the number of pull requests that have been closed in comparison to the total number of pull requests that have been filed. From this point forward, the pull requests are not something that end users are likely to be engaged in very much. Instead, pull requests are more or less concerned with the creators of the applications. One of the ways in which the developer or the contributor in the project actually asks for code, patches, documentation, or anything else to be merged into the project is via the use of pull requests. The workflow is designed in such a way that contributors open pull requests, and individuals within the project who are authorised to review and merge these pull requests will go through the specifics of the content contained within the pull request. After conducting an adequate review, they will decide whether to merge the pull request, reject it, or communicate with the concerned developer for further clarifications. Pull requests are the activity that indicates the activity of development work or contributions to the project. We may say that pull requests are the core of the workflow that GitHub provides on its platform. Because of this, this ratio is distinct from the first one and constitutes a distinctive indicator of the situation. There is also activity about the project, but this activity is not tied to the end users; rather, it is more or less related to the contributors of the project. This percentage represents activity regarding the project. On account of this, we may refer to this attribute as the pull request activity.

3. The third ratio provides an estimate of the amount of effort that is generally contributed by an average contributor. In point of fact, the writers of the recently published research have not discriminated between the various categories of individuals who have contributed to the project. On the other hand, there might be contributions who are not officially recognised. Therefore, it is necessary to evaluate the total amount of contributions from a variety of perspectives. The first one is the proportion of users who have committed to a project in comparison to the overall number of people who have contributed to the initiatives. When a contributor believes that a portion of the work has been completed, they will commit the project, which is similar to a timestamp for the project. These are the minuscule pieces of work that are being done, and they may provide an indication of the amount of effort that is being put into the project as a whole.

There is a possibility that the number of commits made by each contributor is not a particularly reliable source of authenticated data; however, it does provide an idea about the activity of the project as a whole with regard to each contributor in the project. This is true regardless of whether the

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

contributor is a person working within the project or a person working outside of the project. This parameter might be referred to as the project's commit effort, given the circumstances.

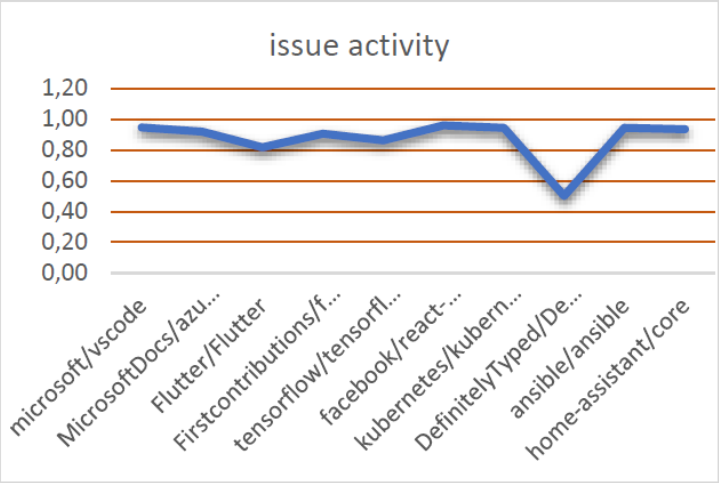
4. In the same vein, we have taken into consideration the proportion of the total number of contributors to the total number of forks.

Currently, forking is a pretty significant operation. As a result of a broad variety of factors, an individual may decide to split off a certain project. It may be for personal or independent use, it could be to study the code and the project itself, or it could be to work on it, hack through it, and contribute to it. All of these things are possible but not guaranteed. It doesn't matter what the circumstances are; the majority of forking is carried out by individuals who are not formally affiliated with the project in any manner. One may argue that forking does not necessarily result in something of great significance, and there are numerous instances in which forks do not have a significant impact. However, there is one thing that can be said with absolute certainty. One would only fork a project if they were in some way or another tied to some interest with the project. This is the only circumstance in which a person would fork a project. Because of this, forking is an indication that individuals who are not as inexperienced with the project are generally interested in it. As a matter of fact, we may state that this ratio itself serves as a benchmark of interests for the project. It is possible that we will refer to this attribute as the fork effort of the project thus.

Finally, we have the ratio that represents the proportion of the total number of contributors to the project to the number of branches that are part of the project. Most of the work that is done on a branch is carried out by individuals who are formally associated with the project. This is the primary reason why the number of forks and the number of branches need to be handled in a distinct manner. It is possible to determine the level of activity of the project's internal contributors by counting the number of branches they have. If there is a master branch, then everytime there is anything new that has to be tested, it is not done in the master branch. Instead, a new branch is formed, and all of the modifications are done in that branch. This is the process. It is strongly recommended by the workflow of GitHub that you should not work directly on the main branch or the master branch, even while working on forks. It is possible that the number of branches divided by the total number of contributors to the project might be referred to as a measure of the branch effort that was put into the project.
Now that the following procedures have been carried out and the parameters have been addressed in great detail, it is time to examine the data that has been made accessible via this exercise and make an effort to arrive at some appropriate conclusions.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

Projects	issues		issue activity
	closed	open	
microsoft/vscode	91403	5222	0.95
MicrosoftDocs/azure-docs	39041	3410	0.92
Flutter/Flutter	34948	7765	0.82
Firstcontributions/first-contributions	215	22	0.91
tensorflow/tensorflow	23710	3751	0.86
facebook/react-native	19256	793	0.96
kubernetes/kubernetes	33958	2025	0.94
DefinitelyTyped/DefinitelyTyped	3423	3400	0.50
ansible/ansible	26470	1531	0.95
home-assistant/core	15059	1016	0.94

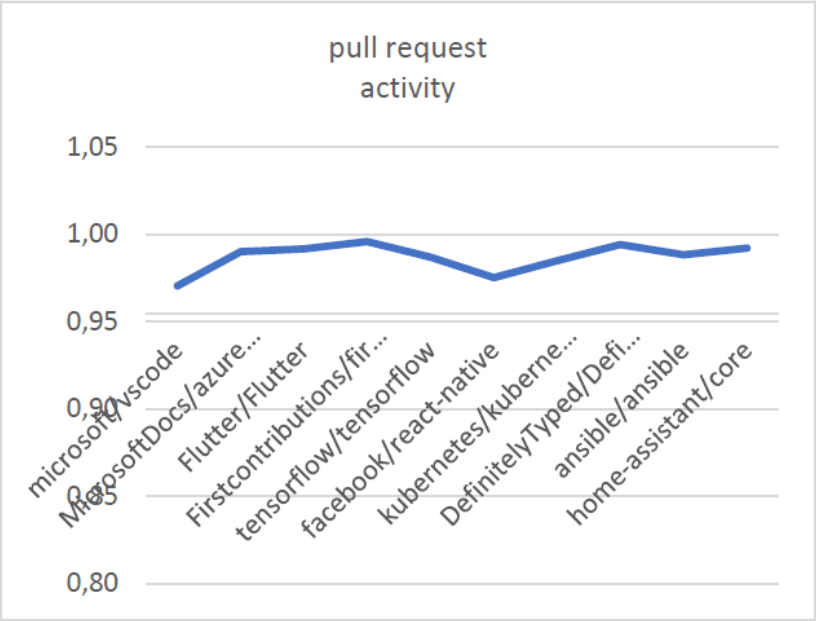


For the majority of the projects, the issue activity demonstrates that a threshold value of 0.8 is obviously present. There is just one project in which the issue activity is known to be below the threshold value, and that project is undoubtedly typed. As far as this specific project is concerned, we can observe that the value is around 0.5. It's possible that this has anything to do with the nature of this software project. Since it incorporates TypeScript type definitions, is there sufficient room for problems to manifest themselves?

Projects	Pull Requests	Pull Request Activity	Closed	Open
microsoft/vscode	7620	0.96	235	2
MicrosoftDocs/azure-docs	1940	0.98	188	7
Flutter/Flutter	2225	0.98	180	5

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

Firstcontributions/firstcontributions	3005	0.99	120	1
tensorflow/tensorflow	1515	0.98	195	6
facebook/react-native	9360	0.97	240	3
kubernetes/kubernetes	5770	0.97	875	8
DefinitelyTyped/DefinitelyTyped	4025	0.98	240	5
ansible/ansible	4300	0.98	505	10
home-assistant/core	2325	0.98	180	6



The pull request activity displays a threshold value that ranges from 0.95 to 1 for the majority of the projects available. This is a really interesting trend since it suggests that excellent projects need to have comparable tendencies in pull requests. This is a very intriguing trend. The fact that pull requests are an essential part of the collaborative software development process, in addition to the fact that they are almost entirely of the open source kind, makes this information very crucial.

Projects	**Commits**	**Contributors**	**Commit Effort (in 1000s)**
microsoft/vscode	70350	19.4	3.70
MicrosoftDocs/azure-docs	62900	15	45.00
Flutter/Flutter	20400	13.5	1.60
Firstcontributions/firstcontributions	6850	12	0.60
tensorflow/tensorflow	94950	10	9.60
facebook/react-native	21000	9.3	2.35

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

kubernetes/kubernetes	94000	7	13.65
DefinitelyTyped/DefinitelyTyped	71050	7	10.30
ansible/ansible	50700	7	7.50
home-assistant/core	29300	6.5	4.70

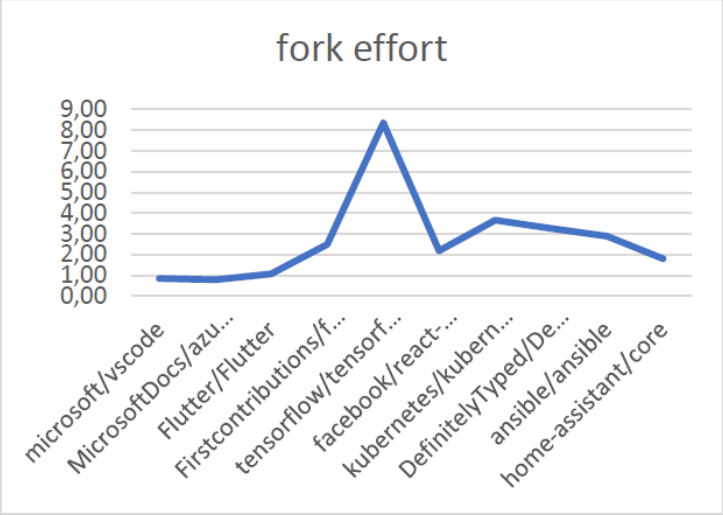
It is possible that we will use a rough threshold value of about 1.5 for the remaining projects if we eliminate the second item from the list, which is a project that centres on documentation in its truest sense. The fact that documentations are updated more often than the project itself means that the amount of time that is committed to the documentations is almost certainly going to be higher.

Projects	**Forks (in 1000s)**	**Contributors (in 1000s)**	**Fork Effort (in 1000s)**
microsoft/vscode	16.4	19.3	0.84
MicrosoftDocs/azure-docs	11.2	14.2	0.78
Flutter/Flutter	14.1	13.2	1.06
Firstcontributions/firstcontributions	29.2	11.8	2.47
tensorflow/tensorflow	83.0	10.0	8.30
facebook/react-native	20.1	9.3	2.16
kubernetes/kubernetes	25.5	7.0	3.64
DefinitelyTyped/DefinitelyTyped	22.8	7.1	3.23
ansible/ansible	19.8	6.9	2.85
home-assistant/core	11.6	6.5	1.80

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 10 Issue 10 - October 2022 - Pages 29-40



With the exception of one project, the threshold value may be considered to be very close to unity. With a few exceptions, it is slightly less than one (about 0.8). As in earlier instances, it is somewhat more than one. A dramatic increase in the increasing trend can be seen in the data for Tensorflow. Its value is a distinctively different proposition. On the other hand, this is not surprising given that the primary purpose of this project is machine learning. Based on the most current developments in the field of machine learning, we have discovered that individuals have chosen to fork this project in order to include it into the main project.

4 Conclusions

In light of the facts presented above, what are some of the implications that we may take away from it? It is probable that we will go over the potential findings in detail, as they have been explained in the previous section. Prior to carrying out this action, it is essential to draw attention to certain regions that have the potential to cause issues if careful attention is not paid when using such parameters. An illustration would be helpful in gaining a deeper comprehension of this particular issue. Let us assume that we have a software project with the name A that has a total of 80,000 pull requests, and that sixty percent of these pull requests have been closed. At this point, the pull request activity would be determined by calculating the ratio of the number of pull requests that were closed to the total number of pull requests. The pull request activity for this project, which is denoted by the letter A, would be 0.75. Consider the following scenario: we have another software project that we will refer to as B. Within that project, we have a total of ten pull requests, and eight of those requests have been closed. 0.8 would be the value of the pull request activity for this project that is referred to as B. Based only on these facts, which indicate that the pull request activity of project A is 0.75 and the pull request activity of project B is 0.8, one could be tempted to assume that project B is in a better situation than project A. However, the pull request activity ratio for B has been computed using a much lower number of pull requests in comparison to A. As a result, the conclusion that B is doing better than A is an incorrect assessment of the situation. For the purpose of avoiding issues of this kind, it would be

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

preferable to examine a project and conduct an analysis of its pull request activity or issue activity only when it has reached a certain threshold for the number of pull requests or issues. Within the scope of this research, the authors have addressed projects that have a significant number of pull requests and problems. More than ten thousand problems or ten thousand pull requests have been submitted for the majority of the projects that are being discussed here. A handful of them do not satisfy these conditions, but there are a few. Despite this, they have been brought up for discussion due to the fact that they are included on the list of the top ten software projects on GitHub for 2019.

1. The first data pertains to the activities of the problem. The problem activity is something that has to do with the users in general, as we have mentioned to some degree in the part that is devoted to the technique. One possible interpretation of this metric is that it serves as a measure of the degree to which ordinary people, both technical and non-technical, are becoming interested in the project in general. On the basis of the table and the graph that follows, we have seen that the values for issue activity for practically all of the projects, or more accurately, for all of the projects with the exception of one, are more than 0.8.
2. Considerable debate is warranted about the question of whether or not this figure may be considered a threshold value. In addition, the relevance of this parameter, which we will save for a discussion in the future, is something that should be actively debated.
3. The pull request activity is the second parameter that has been addressed in this article. Activities carried out by the developers are indicated by this. For each and every one of the projects that are shown here, the value of this parameter is pretty high. This is close to 0.95 and even higher. As a result, we are able to propose that the sign of a project that is quantitatively stable is one that has pull request activity that is close to a value of 0.95. Furthermore, we have already mentioned this topic, which is that the projects that are featured here are the greatest ones in terms of the amount of contributors that their respective projects have. It has already been established that the majority of the projects that have been listed here have surpassed the threshold of ten thousand pull requests that we have established over here. Additionally, there are several projects that have received a far larger number of pull requests than the ten thousand that were initially requested. Because of this, the value 0.95 is a credible number to have faith in.
4. The remaining characteristics, which are collectively referred to as effort parameters, are essentially items that are computed with respect to unit contributors. It is possible that these factors do not provide a significant challenge regardless of the size of the project. This is due to the fact that the work that the contributors put in cannot be minimised, regardless of how unimportant the project may be. Take, for instance, a certain project with the name X that has more than 5000 forks and 2500 collaborators. For the purposes of this project, the fork effort is the number of forks that are contributed by each unit, which would equal to around 2. Consider the following scenario: there is a different project called Y, and it has thirty forks and ten contributors. At this point in time, the fork effort for project Y is 3. Are we able to disregard the efforts that have been put forth by the contributors in Y just due to the fact that it has a lesser number of forks and a smaller number of contributors? During the process of assessing the effort parameters, the authors are of the opinion that the total number of contributors, as well as the number of commits, branches, and forks, may not be taken into

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

account. Nevertheless, in the current situation, we have chosen to focus on big projects. This was not because it was required from a sampling perspective; rather, it was just because these projects are the most popular ones on GitHub, and it was important to investigate the quantitative characteristics of these projects. Based on the information that we have gathered from the most popular repositories on GitHub, we are able to reach the following conclusions: As far as the development of the project's content is concerned, the commit effort will provide an indication of the degree to which the contributors are successfully becoming engaged in the project. This does not always mean that every single patch or other piece of material that is committed is for the purpose of fixing bugs or other issues. The addition of features or the enhancement of the project's current state of art will constitute a significant portion of the commits. Regardless of the circumstances, the purpose of commits is to contribute to the overall improvement of the quality of the software project. As a result, tracking the amount of work that the developer puts into contributions is an essential aspect of a specific project that should be monitored. We discover that this value is nearly one for all of the projects, and it is clearly more than 0.95. This is the conclusion that we get after looking at the data presented in this article on the commit efforts of the top 10 GitHub development projects of 2019. That there is at least one commit for each individual who has contributed to these projects is something that we can claim. It is possible that we will interpret this as an indication that the software project is in excellent health. Furthermore, we may interpret it in such a way that in order for us to become a software project that is highly rated, we will need to maintain this value at a minimum of one. At this time, we feel it necessary to provide a word of warning. To put it another way, a valuable project ought to have a commitment effort that is at least one value equal to one. Nevertheless, it is possible that the opposite is not true.

This means that even if a project has a commit-effort value that is equal to one, this does not always guarantee that the project will be a good one. To give you an example, let's say there is a project P that has two contributors and two contributions. In this case as well, the amount of work that was committed is one, but it alone does not qualify it for inclusion in the top ten projects that are discussed in this article. This is due to the fact that both the number of commits and the number of contributors are very low and hence negligible. The question of what the threshold value of the number of contributors or the number of commits would be that would be adequate to determine if a commit-effort of 1 is acceptable to assess a particular project is one that is up to debate, and we will leave it to the scope of study that will be conducted in the future.

When we look into the statistics of fork-effort, we see that this is also virtually at the same level for the top 10 projects.

One project is the only one that has a different value. For the remaining nine, the value of this parameter is somewhere in the vicinity of and higher than 1. This parameter may be a clue regarding the level of interest that individuals have in the software project, as we have previously described before. In the event that the project is hosting some code, then this interest is unquestionably in the

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

code that the project is hosting. It is because forking it finally results in the creation of a clone that can be worked on. Many people are interested in forking the project just for the purpose of analysing the code. On the other hand, it is also a kind of interest. The writers are of the opinion that there need to be a criterion for determining the level of interest in the project, and in order to do this, the fork effort may be an alternative that can be taken into consideration. The fact that there is at least one fork for each contributor is shown by the forkeffort value of 1. In a manner that is analogous to the scenario of commit-effort, the fork-effort value must be close to 1 in order for a project to be considered successful; nonetheless, the fact that the fork-effort value is 1 does not necessarily indicate that the project is a successful one. In this regard as well, there must to be a minimum number of contributors in order to determine whether or not fork-effort can decide the degree of quantitative growth or even the quality of the project. The writers are of the opinion that this matter is not within the purview of this work, and as a result, they have made the decision to postpone examination of this matter to a later time.

All of the information that we have discussed up to this point is not the same as the data for the branch-effort. In order to work on anything new, branches are formed, which are official clones of the repositories that previously existed. Upon further investigation, we have discovered that the values of the branch-efforts of several projects are notably distinct from one another. The fact that this is the case shows that the branch-effort is highly reliant on the nature of the project, but the other metrics, notably the commit-effort and the fork-effort, are mostly independent of the nature of the project. It is not possible for us to use branch-effort as a yardstick to measure the quantitative growth of the software project or, for that matter, the quality of the software project. This is due to the fact that branch-effort is reliant on the nature of the project to a significant degree. This brings us to the next and most significant topic, which is about whether or not this parameter may possibly be of any benefit whatsoever. This parameter may be more significant and may give more information into the quantitative nature of the software project, according to the authors, if the software projects are categorised into categories. This is the opinion of the authors. This may also be brought up in a conversation at some point in the future.

References

1. Octoverse. (n.d.). Retrieved from https://octoverse.github.com/
2. GitHub Docs. (n.d.). Retrieved from https://docs.github.com/en/github
3. Padhye, R., Mani, S. K., & Sinha, V. (2014). A study of external community contribution to open-source projects on GitHub. https://doi.org/10.1145/2597073.2597113
4. Hata, H., Todo, T., Onoue, S., & Matsumoto, K. (2015). Characteristics of sustainable OSS projects: A theoretical and empirical study. In Proceedings of the 8th international workshop on cooperative and human aspects of software engineering (CHASE ’15) (pp. 15–21). IEEE Press, Piscataway.
5. Tamburri, D. A., Palomba, F., Serebrenik, A., et al. (2019). Discovering community patterns in open-source: A systematic approach and its evaluation. *Empirical Software Engineering, 24*, 1369–1417. https://doi.org/10.1007/s10664-018-9659-9

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 10 Issue 10 - October 2022 - Pages 29-40

6. Gamalielsson, J., & Lundell, B. (2013). Sustainability of open source software communities beyond a fork: How and why has the LibreOffice project evolved? *Journal of Systems and Software*, *3*(11), 128–145. https://doi.org/10.1016/j.jss.2013.11.1077
7. Schweik, C. M. (2013). Sustainability in open source software commons: Lessons learned from an empirical study of SourceForge projects. *Technology Innovation Management Review*, *3*, 13–19. Retrieved from http://timreview.ca/article/645
8. Tamburri, D. A., Lago, P., & van Vliet, H. (2013). Organizational social structures for software engineering. *ACM Computing Surveys*, *46*(1), 3.1–3.35. https://doi.org/10.1145/2522968.2522971
9. Octoverse. (2019). Retrieved from https://octoverse.github.com/2019/
10. Rashid, E. (n.d.). *International Journal of Computers*. Retrieved from http://www.iaras.org/iaras/journals/ijc