UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Karl-Martin Uiga

What happens to all these hackathon projects?

Master's Thesis (30 ECTS)

Supervisor: Alexander Nolte

Tartu 2019

What happens to all these hackathon projects?

Abstract:

Hackathons are growing in popularity over the past few years drawing the attention of researchers in order to learn more about them. Research before has focused on hackathons organizational side, on how to get more participants to the events and the outcomes of a hackathon has been researched on a very small basis of up to 10 projects per study. This paper focuses on collegiate hackathons where the participants are students by looking at what happens to the hackathons and what could be the factors that can influence a project's continuation after the hackathon on a large basis with over 60 000 projects. It is identified that not all the projects produce something tangible after the hackathon but there are projects that produce value to the community. The importance of winning is the only factor found to significantly influence project's continuation after the hackathon.

Keywords:

Hackathon, Team event, Software engineering, Project Sustainability, Innovation

CERCS: P170 - Computer science, numerical analysis, systems, control

Mis saab projektidest pärast häkatonide lõppu?

Lühikokkuvõte:

Häkatonid on järkjärgult populaarsust kogunud ning seeläbi sattunud teadlaste huviorbiiti, kes soovivad uurida häkatone lähemalt. Varasemalt on uurimistööd keskendunud häkatonide korralduslikule küljele ning sellele, kuidas kaasata rohkem huvilisi. Uurimistööd, mis on keskendunud häkatonide tulemustele, on tehtud väikese ulatusega kaasates kuni 10 projekti. Antud uurimistöö keskendub rohkem kui 60 000 projektile, mis on alguse saanud häkatonidel, kus osalejad on õpilased või äsja lõpetanud vilistlased ning selgitatakse välja, mis juhtub projektidega pärast häkatone ja millised tegurid võivad mõjutada projekti edukust pärast häkatoni. Uurimistöö käigus leiti, et mitte kõik projektid ei ole tulemuslikud, kuid on meeskondi, kes loovad tehnoloogia valdkonda väärtust. Häkatoni võitmine on ainus tegur, mis antud uurimustöö põhjal mõjutab projekti edukust pärast häkatoni lõppu.

Võtmesõnad:

Häkaton, Meeskonnaüritus, Tarkvaraarendus, Projekti jätkusuutlikkus, Innovatsioon

CERCS: P170 - Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Table of Contents

1 Introduction

1.1 What is a hackathon?

Fast-paced events such as hackathons have grown in popularity in recent years as the events started out from focusing on young developers to expand their skillsets and work on projects for fun to being events where participants from different fields come together to network, work on novel ideas and advance in their career path [1], [2].

The goal of a hackathon in every form is to, in a short time (usually 24-48 hours), solve a given problem, whether the idea is hackathon specific or brought by the participants themselves [2], [3]. A hackathon's goal is not always to create a successful product but to find new perspectives and grow as a person [2], [3].

1.2 Objectives of the research

There are many hackathons taking place all around the world, in which organisations spend time and money on, looking to get a positive outcome out of the event. The work so far has focused on the organizational side of a hackathon of how teams organize themselves [4] and on the social interactions [4], [5]. Few studies have focused on the outcomes of the projects by looking at the learning outcomes [6], the sustainability of a project [5], [7] and a detailed research on what contributes to project's continuation after the hackathon [1]. The attention so far has been on a very small scope, looking at a specific event with 10 projects at maximum.

This research focuses on a larger dataset of 2 000 hackathons with 60 478 teams each consisting a separate project. The majority of hackathons are collegiate events including 103 026 unique participants who are students or graduates from a year before. The data was in majority gathered from Devpost [8] – a website holding metadata about hackathons. To see what happens to the projects after the hackathon has ended the following research question is asked:

> **RQ1:** What happens to all these hackathon projects?

Hackathon is an event to expand skillset, kick-start an idea, network with similar-minded people, advance in career path or to win prizes. It is found before that teams with preparation were more successful at being continued after the hackathon [1] but the research has been done on a very small scope. To see what are the possible factors that could promote or hinder a project's continuation the following research question is asked:

> **RQ2**: What are the potential factors that can promote or hinder a project's (dis-)continuation?

In order to answer the research questions the initial dataset has to be cleaned to remove projects that contain invalid information and afterwards an additional data gathering is conducted via GitHub API to collect extra data about each project. To answer the first research question, the gathered data is thoroughly analysed and investigated. Answering the second question starts with operationalizing the factors that are measurable in the given dataset in order to make them comparable, which is followed by a regression analysis to find the predictors that have significant impact on predicting project's survivability after the hackathon.

The results from this research provide an understanding of what happens to the projects after the hackathons end by exploring the changes between a few days versus a couple of months after the hackathon and examining a couple of example outcomes. In addition to the analysis, this research suggests what factors to consider in order to have a better chance to continue with the project after the hackathon.

2 Literature selection criteria

Background: Hackathons have experienced a high rise in popularity over the past few years drawing the attention of researchers interested in learning more about the field of hacking [1] which results in several papers released each year. In order to identify the outcomes of a hackathon and predictors that have influenced project's continuation, a systematic literature review was conducted by following the rules described below [9] and results described in chapter 3. The results are used for the research, where the outcomes are used in analysing the data and predictors are used to identify factors that can influence project's continuation.

Research questions:

- **RQ1**: What happens to all these hackathon projects?
- **RQ2**: What are the potential factors that can promote or hinder a project's (dis-)continuation?

Search terms and strategy:

- Hackathon
- Results of a hackathon
- Project continuation
- Students
- Enterprise
- Hacking
- Learning platform
- Influencing factors

Terms are used in the search resource's search form by combining different terms with operators like AND or OR, for example *hackathon* AND *results of hackathon*.

Search resources:

- Digital library (Google scholar)

Study selection criteria and quality assessment:

To be included, the study must satisfy all the following criteria's:

- The study provides results of a hackathon.
- The study provides the actions made by the teams before the hackathon.
- The study is written in English or Estonian.
- The study is peer-reviewed.

Data extraction strategy:

Publication details	Date of data extraction	RQ1	RQ2	Additional notes

- Details about the publication: authors, creation date, found from, title.
- Date of data extraction: when was the publication found/last opened?
- RQ1: Data/information to answer the first research question.
- RQ2: Data/information to answer the second research question.
- Additional notes: Any additional information that might be useful.

3 Background

Hackathon is a time-bounded and fast-paced event where participants demonstrate their knowledge in small teams by creating working software or hardware from an idea within a small timeframe [6]. The events started out as a place for students to hang around with their friends, code and be provided with free food and drinks which are usually a must at a hackathon but the events have grown in popularity in recent years resulting in being more competitive, attracting a lot of non-computer science related people and more often making an impact for the world [1], [4], [6].

Hackathons have a variety of objectives they aim to achieve, therefore each hackathon is classified by one of the following classification [10]:

- Communal – collaboration within a community to develop resources, infrastructures and practices for the community (internal hackathons within an organization to innovate new ideas within the organization/community).
- Contributive – adding value to an already existing project by breaking work into smaller tasks that can be completed in these intense sessions. The organizers give ideas to the participants.
- Catalytic – a more competitive scene where participants are trying to demonstrate the utility of a dataset, an approach or use of a technology to show their ideas.

Hackathons used in this research cover all the classifications stated above.

The research that has been conducted before this thesis has mainly focused on the hackathon itself and its organizational side [1] by looking at how to make a hackathon more effective and how a hackathon can become a learning platform or an opportunity for career advancement [6]. There has also been research done on hackathon projects and actions taken before or after the hackathon but those researches have mainly focused on looking at a specific hackathon types with a much smaller scope, usually referring to 4-5 projects [1], [3], [6], [11]. This thesis will be focusing on what happens to hackathon projects after their completion and what could be the factors of a (un-)successful hackathon project on a much larger basis, including 60 487 projects.

3.1 Outcomes

In this section outcomes that have been found to influence project's continuation before are listed and described.

3.1.1 Activity after hackathons

A hackathon project does not have to win the hackathon; neither become an enterprise to be called successful. If it contributes to a given field, it could already be a great success. A common thing to happen is that when the hackathon ends some work will be done on the project to polish it enough to be published [1], [7] and in some cases it grows to an open source project that everyone is open to use and contribute to [12] or an influential project that could grow into something bigger and potentially save lives [13]. On the other hand, some projects are left to a dormant state for a long period and then picked up to be worked on yet again. In these cases, it could be that the idea they had at that point just did not fit into the company's business case [3].

A hackathon is not meant to be an event where revolutions are made but to make more evolutions of an idea [1] and open source projects are a good example of it as the ideas might be refined and new perspectives found by somebody else [14]. There are several examples where hackathon projects have been taken over by someone else after the hackathon, in case the original contributors did not have any more motivation to continue, lacked ideas or, in case of organizational hackathons, they were taken over by the company as a new business case [1].

Out of all the hackathons there are also many projects with the potential of attracting a large amount of people but due to several reasons these projects are archived before someone else could pick them up. In some cases, it could be that the organization is not ready for a business case [1], [3] as they do not have the resources for it but for some other cases it is the motivation that turns the project to dormant state before it could rise [5].

3.1.2 Hackathon as a learning platform

Software engineering is considered to be the fastest growing field that needs to adapt based on new technologies every year and a hackathon is considered to be a great place to learn and adapt to something new [2] as competitiveness in a hackathon enforces imaginative thinking [6].

Hackathons provide a great learning platform because while enforcing a competitive out-of-box thinking, there are also professional mentors who help participants and teach specific details about the given hackathon topic [5], [6] which creates a perfect environment to learn new skills.

By working together intensively for a weekend, participants share their experience and knowledge and a hackathon is believed to be a good learning source [5] as some students have stated that hackathons allow them to accomplish more in twenty-four hours than they had done in a year at school [6]. A hackathon provides a fun and engaging format that is believed to have a good effect on students GPA [6] as they learn to perform in stressful situations so that they are better prepared for similar situations at university.

In case students, who participate in a hackathon, do not know yet what they would like to specialize on, the hackathon provides a perfect environment to try new fields, obtain new skills [1] and maybe find a better suitable field for them.

3.1.3 Hackathon as a networking platform

Networking is not merely the exchange of information with others — and it is certainly not about begging for favours. Networking is about establishing and nurturing long-term, mutually beneficial relationships with people you meet. Nowadays having a well-established network has become an important aspect in our lives to be more successful. To be 'networking' is not attending dedicated networking events as it is happening every day around us while we are communicating with each other.

Hackathons are a great place to practice networking, as the attending people are somewhat interested in the same fields and thinking along the same paths. This creates a good platform to build relationships and chemistry between the participants which is important if participants wish to build something together even after the hackathon ends [2], [6], [13]. Moments of physical contact (e.g. handshakes or pats on the back) or eating and drinking together are fundamental ways to connect with each other [4] – everything that is inevitable at a hackathon.

Networking is important to kick-start something with companions, to get recruited or just to develop as a person. In some cases, attendees, who might have a good connection base but are lacking connections in the field they are interested in, might find companions to work with on their ideas [2]. Apart from growing personal connections, participants stated that taking part of a hackathon is very important for their career advancements and to enhance their CV [7]. For example, an analysis revealed that in two out of four cases the leaders of hackathon projects got promoted to different leading roles due to showing creativity and capability by performing well in the hackathon [1]. In addition to promotions there were findings that participating in hackathons perceived positive impact when having their annual reports at work [1].

3.2 Predictors

A dependent variable is what is being studied and measured in research [15], in this thesis it is project's continuation. In this section a set of factors that are found from literature related to hackathons, which could promote or hinder a project's (dis-)continuation, are listed and explained.

3.2.1 Activities prior hackathon

Hackathons as events have evolved significantly during the last few years and so has the preparation that participants do before a hackathon. While at the beginning attendees would just come to an event and would then start to think about what to do at the hackathon, nowadays a weekend long hackathon may even need weeks of preparation before the event [1], [4], [5]. In some cases, teams meet-up once a week for at least three weeks before a hackathon to prepare for the event and set up everything they need for the hackathon. This is proven to have a positive effect on the end result [5].

The preparation before the event also includes approving an idea with all the teammates, talking about the idea with different people to get the scope of the project [1], [6]. To benefit the continuation of a hackathon project it is also recommended to make up an initial plan with tasks being shared between team members that have the abilities to complete certain tasks. This also applies to setting up the team in first place – it is important to include different roles (Product Manager, UX Designer …) with a large set of skills – to have a skill diversity within the team [1], [5]. Skill diversity gives the team a multi-dimensional view of the idea and helps to avoid trivial mistakes during initial planning.

During the planning phase some of the teams managed to already find a home (find a team/organization to take it over after the hackathon) [1], [5] for the project after the hackathon ends which was identified to have benefits on the continuation of the project as it will set clear goals for the team to achieve before the end of the hackathon [5]. It was also identified that planning before the hackathon allows projects to be finished before the event ends and leaves the teams with time to polish the produced code [1].

3.2.2 Winning, team size and outside help

Participants state that the main reasons for attending a hackathon are to win prizes, learn from the experienced ones and to socialise with similar-minded people [2].

Having a great idea alone is not enough to be successful, a project needs to be backed and have funding from outside of the team to be sustainable and a hackathon is a great place to catch eyes of possible funders [2]. Some hackathons are backed by corporations who offer prizes such as the possibility to develop the idea under their wing and get mentored by experienced developers [2], [16].

At the same time mentorship at a hackathon is growing popularity as the results show that participants tend to try new technologies and have a better scope of the project in overall due to seniors being at the event to help them out [6]. Events are investing more heavily and focusing on the problem of getting enough seniors to an event to cope with student's problems [2], [4], [7].

Having mentors at a hackathon tends to have a good influence on students finishing their project's or even figuring out a way to continue their work after the hackathon because students feel that they are not left alone with their issues [5].

Team size, on the other hand, can be varying depending on the hackathon format, usually ranging from 2-4 members per team or in some cases even up to 10 members. In cases where teams consisted of more than 4 members per team it has been found that although it felt overwhelming and not so effective as people started to step on each other's toes with taking other's tasks, the project would not have been continued without having many members in the team due to lack of motivation from some participants. [5]

3.2.3 Hackathon size and location

Hackathon is an event where information has to be exchanged rapidly and efficiently. For that to happen it requires trust between the participants and trust develops more easily by being face to face interactions [17]. Face-to-face contact is thought to be an essential feature of most innovation behaviour due to the fact that [4], [17]:

- It is a way for efficient communication;
- Builds trust and incentives in relationships;
- It is a motivation for participants;
- Facilitates screening and socialising;
- Richest form of social interaction;

The events that are co-located are usually limited by how many teams may attend but the ones that are held online are without boundaries with some hackathons having over 900 participating projects [7].

A hackathon with lots of participators may attract more sponsors which in turn may attract more skilful participants but on the other hand having a lot of competition may be demotivating for some of the teams as they feel that they do not have a chance to win at all [5], [7].

3.2.4 Motivation

The motivation is the willingness of an individual to make an effort on a project which impacts on it being successful or not. When team leaders are motivated to get things done the projects are more successful than others as the planning was though thoroughly and because of that the team motivation was higher [1], [6]. A team does not have to make revolution to be successful but be motivated to finish what they had initially planned [1].

Table 1. Predictors with sources

Predictor	**Source**
Winning	[2]
Outside help	[2], [4], [6], [7]
Skills diversity within the team	[1]
Project’s skills coverage	[1]
Location	[4], [5], [7], [17]
Team familiarity	[5]
Motivation	[1], [6]
Hackathon size	[4], [5], [7], [17]
Team size	[5]
Team preparation	[1]

Usage of skills learned at a hackathon	[6]

Table 1 contains all the predictors found to have an impact on project's continuation in other researches along with the source the predictor is found from.

4 Method

To answer the research questions stated in the introduction, data analysis and regression modelling are applied following a data gathering to collect information from GitHub about each project. The context, methods for data collection and means of analysis are described in the following sections.

4.1 Setting

Hackathons investigated in this research are mostly collegiate hackathons where the participants are students or graduates from a year before. Organization to bring hackathon organizers and students together is called the Major League Hacking (MLH) – an official student hackathon league whose goal is to assist organizers by providing necessary resources and help to organize the event [18]. More than 65 000 students attend various events organized by different organizations each year and MLH is a community that helps organizers and participants to come together [18]. MLH encourages hackathon organizers to register their events under Devpost – an independent website holding metadata about hackathons [8].

Large amount of the initial data used in the thesis is gathered from Devpost's last five years, although Devpost does not only contain hackathons covered by MLH as it includes many other events majority of the events are covered by MLH, including:

- 60 478 projects;
- 103 026 unique participants;

The initial data with the structure visible on Figure 1 was gathered on 10th of May 2018.

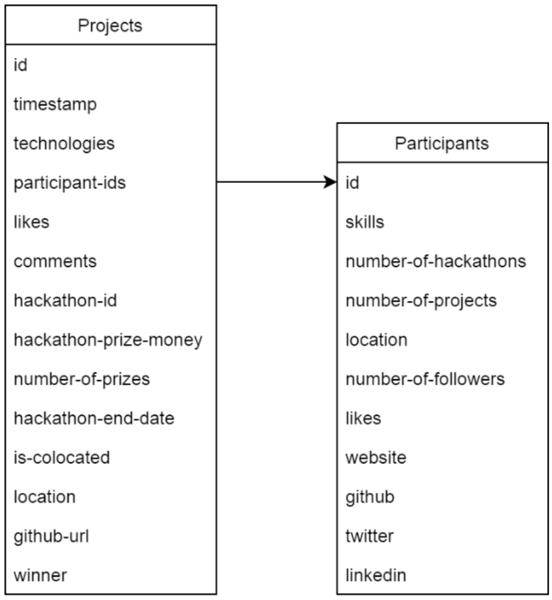


Figure 1. Structure of the initial gathered data.

Gathered data contains knowledge about the hackathon related to the project and basic information about the project, including its GitHub URL and participants ids linked to another table containing data of each individual participant.

4.2 Additional data gathering

To gather more in depth data about each project an additional data gathering was conducted using Python [19]. Each project in the gathered data has a GitHub URL, therefore data had to be gathered from GitHub which has a request limit of 5 000 requests per hour and no possibility to gather information about private repositories. To gather additional data, over 200 000 requests had to be made, which led to discover GHTorrent – an open source project that monitors GitHub public event time line and stores the data into a database. This database can be used via an SSH tunnel and the request limits are way higher (requests that will take longer than 100 seconds are denied).

However, it was found that GHTorrent has issues with its database. Some of the projects are missing while others have false data in the database. To avoid issues with the gathered data only GitHub API is used.

The initial data has the issue that the majority of the projects (38 811) do not contain a repository link. The number of projects without necessary information is relatively high because inserting information to Devpost is not obligatory, so the participants might not want to enter it. Therefore, a GitHub link was generated for each project that has it missing by using the following formula:

- $GitHub\ API\ URL\ path + team\ leader's\ username + project\ name$

By generating a link for each project that has it missing increased the number of projects with some information after the hackathon by 2 500 projects.

The additional data was gathered on 11th of December 2018 and the structure of the data can be seen on Appendix I. When gathering the data, it is taken into consideration that each project's activity timeframe needs to be the equal to be comparable afterwards. Therefore, a timeframe of 6 months after the hackathon is chosen to be the frame where the activity is observed. Initial results on 60 478 hackathon projects are following:

- 8 785 projects with work after the hackathon
- 11 893 projects with no work after the hackathon
    - Total of 20 678 projects with some information after the hackathon
- 3 870 projects with a possibility of a private repository
- 35 930 projects without necessary information

GitHub has limitation of private repositories, which are not reachable via the API. Because of this for 3 870 projects there is no information available due to being private repositories.

4.2.1 Data cleaning

To make a dataset comparable it needs to be consistent and therefore cleaning the data of incomplete rows is the next step to be taken in order to make the data comparable. Data of 20 678 projects was cleaned by following these requirements:

- A project has to have at least one technology stated that is used for producing this project: a project is made out of something and when a project has no technologies stated, it is impossible to calculate accurate values therefore making it incomplete and incomparable.
- A project has to have at least two members in the team to be considered as a valid one: a hackathon team consists of at least two participants [2].
- At least two participants of one team are required to state their skills: a skill diversity within the team and project's technology coverage cannot be calculated without having proper data of the team therefore making the values incomparable.

- A hackathon has to consist of at least four projects to be considered as a valid hackathon.

The dataset was reduced to 12,707 projects after cleaning it using Python [19] by following the rules stated. In the following analysis, the cleaned dataset is used.

4.3 Analysis strategy

In order to identify what happens to all these hackathon projects and to investigate possible factors that could promote or hinder project's continuation a dormancy property is to be set at first. Setting a dormancy property avoids considering projects with minimal activity as active ones.

Variables are operationalized in order to make the data comparable, as the gathered data is not in a comparable form. To answer the first research question – what happens to all these hackathon projects (RQ1) – an analysis is conducted on the gathered data in order to explore the events after the hackathons and to see if outcomes match with the ones found in the literature before.

To answer the second research question – finding factors that could promote or hinder a project's continuation after the hackathon (RQ2) – a regression model was considered as a valid way to investigate the relationship between the variables and the outcome of a project.

4.3.1 Setting a dormancy formula

To analyse the hackathon data used in this research there is a need to set a requirement by which to consider a project dormant after a hackathon. A project is considered to be dormant if it is no longer maintained for a period after a hackathon. One approach would be to use the latest commit timestamp as the dormant date but this might now be always accurate as there are projects that after a few months of having no activity are being revived by a few commits to either clean up the already completed project or to indicate that the project has been deprecated [20].

To build a more robust requirement to determine dormant projects and using knowledge from papers that has covered dormancy before [20], then a project is considered to be dormant after a hackathon if it has less than one commit per 6 months on average after the hackathon had ended.

4.3.2 Operationalization of factors

Several factors, listed in Table 1, could influence a project's continuation after a hackathon but some of the predictors in the table are not measurable with the data available for this research. For example, it is not possible to measure the experience or motivation of the participants, which is important to promote project's continuation. Therefore, the predictors that have no measurable information in the gathered data are not used in this research.

At the same time, there are factors measurable in the data available but they are not on the same level with other factors, for example hackathon size and location. These are hackathon related factors therefore they cannot be used together with team related predictors, for example every team in the same hackathon has the same values for hackathon size and location but different values in other factors, making it incomparable.

Table 2 contains predictors that turned out to be influencing projects continuation before and are measurable with the available data but for some of the factors the data available is not comparable, therefore a comparable value was calculated with the formula shown in the operationalization column.

Table 2. Predictors with sources operationalized.

Predictor	Operationalization	Source
Team size	Available in the dataset	[5]
Importance of winning	$1 - \frac{amount\ of\ winners}{amount\ of\ projects}$	[2]
Skill diversity	$\frac{Significance\ rate\ *\ amount\ of\ technology\ known}{amount\ of\ team\ members}$	[1]
Technology coverage	$\frac{Significance\ rate\ *\ amount\ of\ technology\ known}{amount\ of\ team\ members}$	[1]
Team familiarity	$\frac{ratio * amount\ of\ pair\ working\ together}{amount\ of\ team\ members}$	[5]

In the following chapters measurable factors are described and calculation formulas explained.

4.3.2.1 Importance of winning

It has been found that over a quarter of participants state that one of the reasons for coming to a hackathon is to win prizes, whether the prize is something small (e.g. a subscription to a magazine) or something way bigger like funding from a company or being backed by a huge corporation [2], [16].

As the prize for winning can be radically different for each hackathon and data that is used for this research contained hackathons that had several winners per each hackathon, as there were several prizes put out for the participants. Win's importance had to be calculated because two hackathons cannot be compared when in the first one there were in total of 3 winners out of 10 participating teams and the second hackathon had 3 winners out of 5 participating teams.

To see if winning influences project's continuation, a win's weight was calculated for each project following these guidelines:

- Weight has to be between 0.0-1.0 (0-100%).
- Formula for the calculation: $1 - \frac{amount\ of\ winners}{amount\ of\ projects}$
- If a project does not win, its weight is zero.
- If there are as many winners as there are projects, then following formula is used: $\frac{1}{amount\ of\ projects}$
- If a project is the only one to win something, then its weight is one.

4.3.2.2 Team size

Every hackathon has a format, which states for the team size either to be fixed (or at least a suggestion on size) ranging from 2-4 members per team or it's free of choice for the participants [2].

Team size is already available in the initial dataset therefore it was not required to do any extra work for this predictor.

4.3.2.3 Skill diversity and technology coverage

In research done before there was a consensus that hackathons should be promoted more heavily to increase the skill diversity among the participants because to just have a successful hackathon requires small amount of planning but to progress after the hackathon it has been found that initial planning determines heavily the accomplishments of a project [5], [7].

Therefore, it is important to have an advisor in the group to provide support throughout the hackathon. Some of the hackathons have people hired to help students with their planning [5] but in hackathons where you do not have any outside help it is important to have a wide skillset within your team [7].

In hackathons nowadays having a wider skillset does not always mean having people with different developing skills but also having participants who expertise on for example design or product management as they give new dimensions to the team [5]. Having too many participants in a team with the same skillset can turn out to be negative as they can start to step on each other's toes to try to complete other's tasks [5].

In case of having a wide skillset in the team, it still needs to be skilfully managed by the team leader. Tasks should be separated between the participants by the skills that each task requires and if a task demands an unknown ability the participant can study what he/she needs for the task beforehand to improve team efficiency [1].

To observe skill diversity's impact on project's continuation two values were calculated, one for team's diversity regarding technologies used for the project (technology coverage) and the other one for diversity within the team. Comparable values were calculated for each project following these guidelines:

- Each technology is given a significance rate.
    - For skill diversity within the team, technologies are regarded as all unique skills known by the team members.
    - For skill diversity regarding technologies used for the project, technologies are regarded as all skills used for the project.
        - For example, with five technologies each technology will have a significance rate of 0.2.
- A value is calculated for each technology/participant relationship with the following formula:

$$\frac{Significance\ rate * amount\ of\ which\ this\ technology\ is\ known}{amount\ of\ team\ members}$$

    - If a skill is known by at least 75% of project's participants, then it is regarded as everyone possess the skill.

- Skill diversity equals the sum of all the calculated values; the result is in a range of 0.0-1.0.

4.3.2.4 Team familiarity

A hackathon is a great place to build bridges or to endorse existing relationships between people with the same interest field [2], [3], [14]. It has been found that participants tend to work on a project with people that they know beforehand and if they were put randomly into a team, some of them would even refuse to join the event [7].

It is so because participants know each other's strengths and weaknesses beforehand which could also promote a project's continuation in the hackathon and avoid surprises at an event where time is limited and additional resource is nowhere to be found [2], [7], [14].

To evaluate team's familiarity, a formula is set to calculate a comparable weight (0.0-1.0) for each team following these guidelines:

- Research done before suggest that it is groups of pairs who elect to attend hackathons and groups often form from a pair of friends connecting with another group deciding to work together on an idea [5]. Therefore, the amount of participants working together before the observable event is calculated on pairs.
- Each pair is given a ratio with the formula: $\frac{1}{amount\ of\ team\ members}$
- Value is calculated for each pair with the formula:
$$\frac{ratio * amount\ of\ pair\ working\ together}{amount\ of\ team\ members}$$
- Team familiarity equals the sum of calculated values.
    - If the team familiarity value exceeds 1.0, it is automatically regarded as value of 2.0 – familiarity is exceptionally high.

4.3.3 Regression model

Regression modelling is a set of statistical processes to investigate functional relationship between factors [21]. Relationship between the variables is expressed in a form of model, explaining the connection between the dependent variable (dormancy) and one or more predictor variables [21]. In context of this thesis, it is to investigate if a variable has impact on project's continuation after the hackathon or not.

Data gathered for this thesis is left-skewed meaning that there are significantly more dormant projects than projects that are continued after hackathon, therefore Cox's proportional hazards regression method is chosen as it suits left-censored data [22]. Cox is a form of logistic regression also known as survival analysis, which investigates the effect of variables on the dependent variable while taking into consideration the time until an event of interest occurs [23].

The Cox model is expressed by hazard function *h (t)*, this shows the probability of an event to occur per time unit *t*, assuming that it has survived up to time *t* [22]. Hazard function is comparable to measure of risk, where having a greater hazard rate means greater risk of failure. It is calculated with the following formula:

$$h(t) = h_0(\mathrm{t}) \times \exp(b_1x_1 + b_2x_2 + \ldots + b_px_p),$$

where $h(t)$ is the hazard function at time t determined by the set of covariates $(x_1, x_2, \ldots, x_p)$ and coefficients $(b_1, b_2, \ldots, b_p)$ measuring the impact of the variables [23]. $h_0(t)$ is the baseline hazard, which is equivalent to the hazard if all the variables are equal to zero.

Perquisites to perform Cox's regression [24]:

- Both, time-dependent and fixed, types of covariates are allowed;
- Dependent variable must be binary; in this context the dependent variable is project's dormancy:
    - 1 in case of project being dormant;
    - 0 in case of project being not-dormant;
- Existence of time-to-event variable;
    - In the context of this thesis, the variable is *number of days being active after the hackathon;*
- Variable is statistically significant with p-value less than 0.05;
- Variables must be numeric values;

Interpreting a Cox's regression model:

- Hazard ratio (HR) is the measure of effect; it is an estimate of the hazard rate between the control group and the treated [23]. HR is calculated by $HR = \exp(b_p)$, where [23]:
    - HR = 1 means of no effect;
    - HR < 1 means reduction in the possibility of the project being dormant;

    - HR > 1 means increase in the possibility of project being dormant;
- Confidence interval of the hazard ratio (lower 95% - upper 95%) – a narrower confidence interval indicates for a precise hazard ratio while intervals including 1 (the null value) indicate variable's small contribution to the difference in hazard ratio.
- The coefficients in a Cox regression relate to hazard; a negative coefficient indicates a positive prognosis and a positive coefficient indicates a worse prognosis [23].

Covariates explained in chapter 4.3.2 are used in the Cox's regression model to investigate their impact on project's continuation alone and in a relationship with other variables.

5 Findings

The results of the analysis are organized along the two research questions stated in the introduction. The first chapter focuses on the first research question, which aims to find out what happens to all these hackathon projects after the event and the second chapter focuses on to find out the relationship between variables and project's dormancy, which is the second research question.

5.1 Activity after the hackathon

In order to identify what happens to the projects after the hackathons end (RQ1) an analysis was conducted on 12 707 projects available in the dataset, looking at the outcomes of them and identifying if any previously stated outcome from Section 3.1 appear in this study.

The majority (60%) of the projects were left inactive from day one after finishing the event based on the Figure 2. Based on Table 3 the first day after the hackathon was the most active day with 20 963 (22% of the commits) commits by 4 370 unique contributors, averaging 4 commits per project.

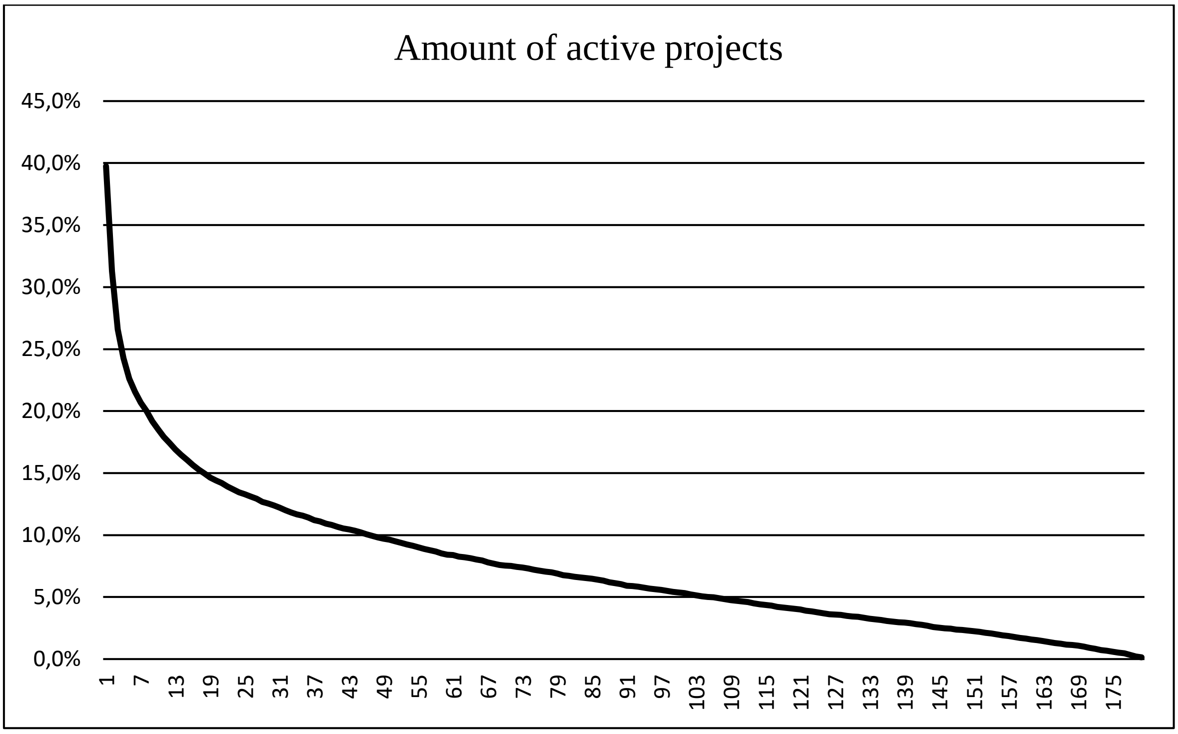


Figure 2. Percentage of projects (y) active after a number of days (x) after the hackathon.

The amount of inactive projects increased rapidly until the 5th day after the hackathon by 2,211 up to 9 835 projects averaging 442 projects per day, which was 43% of the active projects five days after the hackathon. During this time-period there were 15 467 (17% of the commits) commits by 2 700 extra contributors averaging 3 867 commits per day and 3 commits per project.

At 60-day mark after the hackathon, the amount of inactive projects had increased by 1 728, averaging 32 projects per day, which was 60% of the active projects between five to sixty days after the hackathon. The amount of commits made in this time-period was 29 168 (33% of the commits) averaging 530 commits per day and 10 commits per project.

The number of projects turning inactive raised by 635 projects during the time-period of 60 to 120 days after the hackathon, which was 44% of the active projects, on average 9 projects per day. There were 13 882 commits during this period (15% of the commits), averaging 231 commits per day and 12 commits per project.

During the next two months, the amount of projects inactive increased by 496 projects, averaging 8 projects per day, as 97% of the active projects were left behind. The active projects managed to contribute by 12 452 (14% of the commits) commits, which was on average 208 commits per day and 9 commits per project.

Table 3. Statistics of projects lifeline.

Days after hackathon	Amount of inactive projects (%)	Total contributors after the hackathon	Total contributions after the hackathon
1	7 624 (60%)	4 370	20 963
5	9 835 (77%)	7 099	36 430
60	11 563 (91%)	9 928	65 598
120	12 198 (96%)	10 793	79 480
180	12 694 (99%)	11 322	91 932

The results presented in Table 3 indicate that the amount of contributions made was high during the first five days after the hackathon, which can mean that the team had motivation to complete their goals set for the hackathon or to clean up the project to be presentable in the future. This implies the fact that there was no real intention to keep the project alive after the hackathon.

During the next 55 days after the hackathon, the contribution dropped with the amount of commits being slightly lower than during the first five days but on the other hand, the amount of projects still alive was also significantly lower. It implies that the projects may have been worked on, attempted to find a new perspective or the cleaning phase might have been taken on later than straight after the hackathon.

The period after two months after the hackathon show that commits per project increased as time went by with the highest ratio of 12 commits per project during the period of two to four months after the hackathon. This indicates that the more serious projects were worked on after the hackathon with each project having a relatively high contribution during that time.

The percentage of active projects turning inactive within a period did not decrease later on after the hackathon also indicating that the clean-up period for some of the projects was not during the first five days or the project did not seem reasonable.

In the context of what happened to 12 707 projects it is found that 471 (6%) projects that were left inactive after the hackathon were continued after the observed 6-month period which can at the same time mean that they were 'cleaned up' after the observed period.

Some of the projects grew out from hackathons to be open-source projects that are maintained up until now (10.03.2019) [25]. A particular case found is Bottender, which is a modern bot development tool that was created a few months before the hackathon, mainly developed at an hackathon and is now an open-source project that is contributed to monthly [25].

There are examples of an idea being refined at a hackathon and then a new perspective found which was then developed and in some cases maintained up until now [26], [27]. In addition, there are examples of projects that started out at a hackathon and are now available as a finished or occasionally updated product [28], [29].

5.2 Relationship between the factors

In order to analyse the relationship between variables and dormancy to see which could influence a project's continuation (RQ2) the outcome of variable operationalisation is described. It is followed by the regression analysis applied to each factor separately and together respectively against the dependent variable.

5.2.1 Factor operationalization

The statistics of factor operationalization are listed in Table 4 including standard deviation, minimal-, maximum- and mean value of the factors.

Table 4. Statistics of factor operationalization.

Factor	Minimal	Mean (Standard Deviation)	Maximum
Team size	2	3.4 (1.0)	10
Winners at hackathon	1	15.0 (9.0)	45
Times participants working together	0	2.3 (3.5)	11
Technologies used for a project	1	5.5 (3.3)	40
Skill's known per participant	1	4.7 (4.3)	116

Team size of up to four members in the team was the most popular setup based on Figure 3. Size of a team in the dataset ranged from 2-10 members per team, the mean size of a team was 3.4 members with a standard deviation of 1.0.

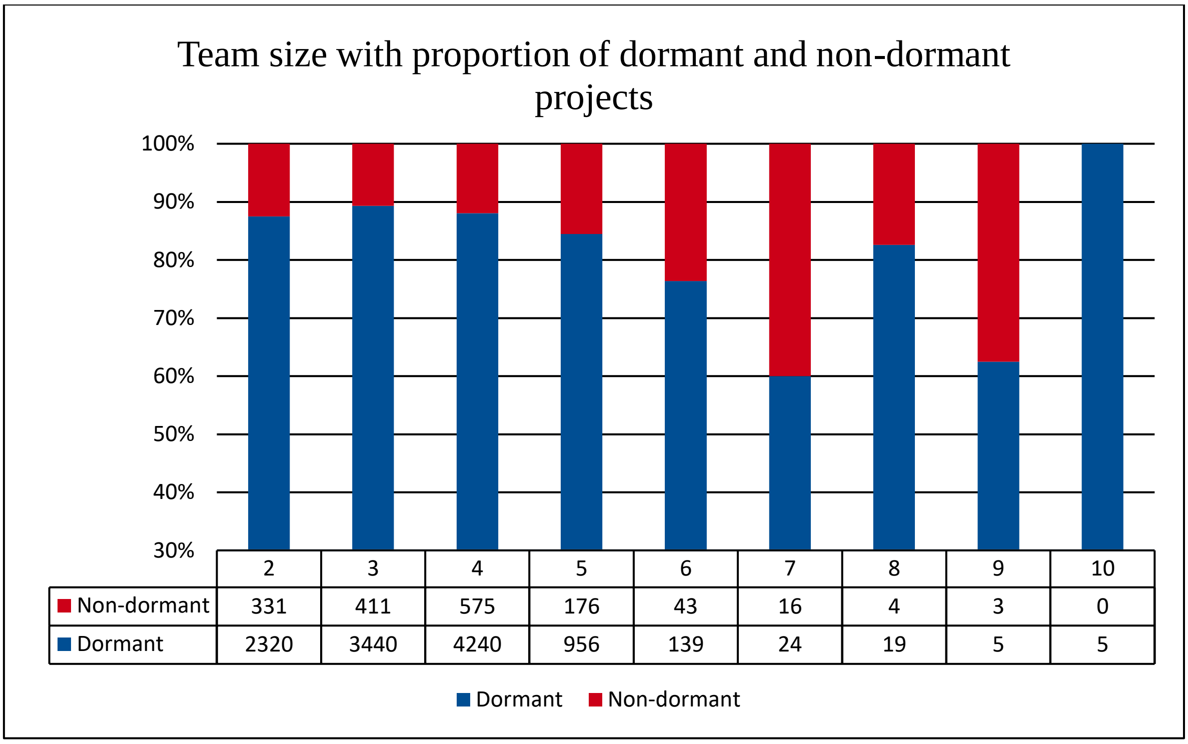


Figure 3. Proportion between dormant and non-dormant projects regarding team size.

Although having more than five members in the team increased the continuation rate significantly based on Figure 3 it was not a popular way to go, with only 258 projects out of 12 707 were larger than five in a team. On the other hand, a larger team can mean that the project has a higher chance of not being dormant after the hackathon by having more members that can continue with the idea.

The results of 12,707 projects regarding winning weight distributed between dormant and non-dormant projects are displayed in Figure 4.

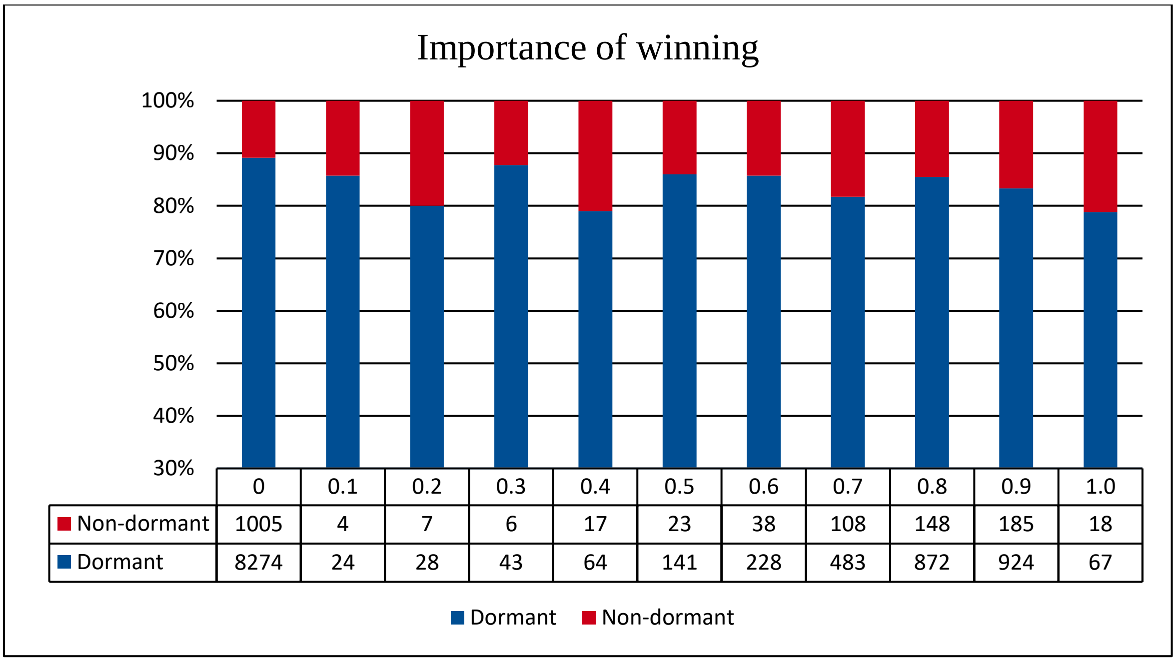
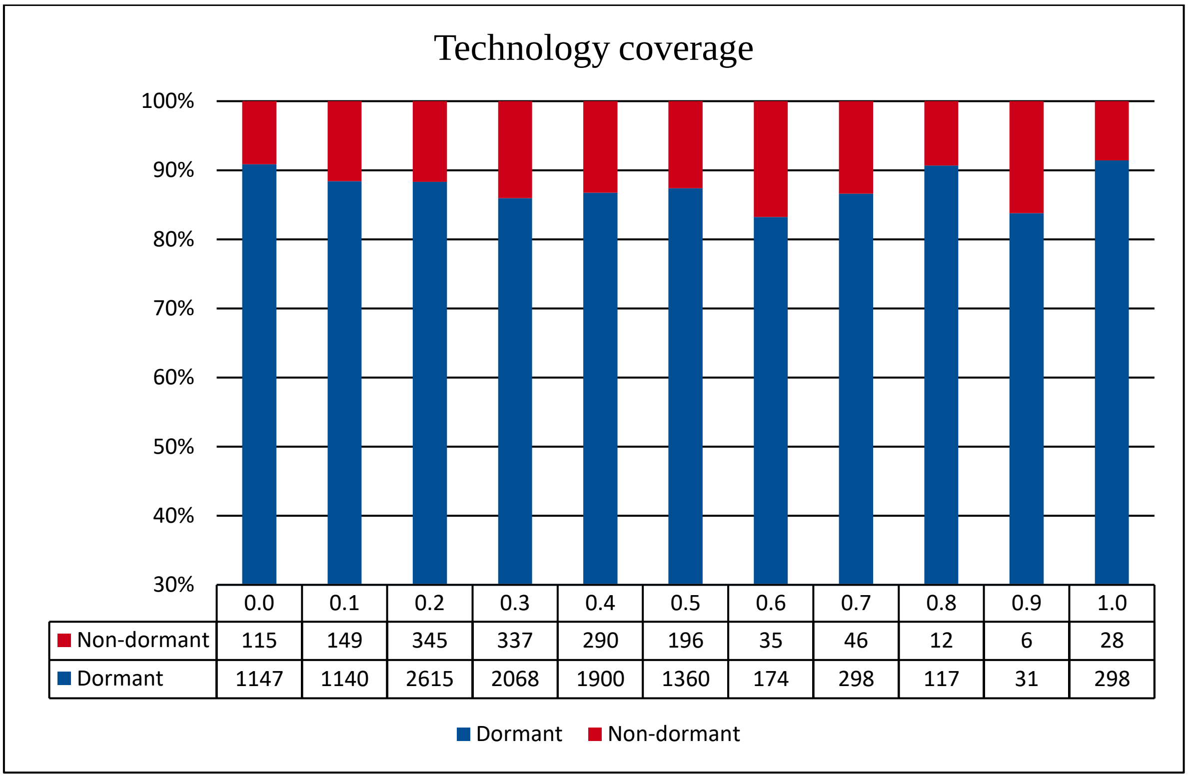


Figure 4. Proportion between dormant and non-dormant projects regarding win's weight.

Based on the data used for this research 27% of the teams won something, making it an achievement rather than a standard. Winning did have an effect on project's dormancy as projects that won something (win's weight higher than 0) had relatively higher percentage of being continued after the hackathon.

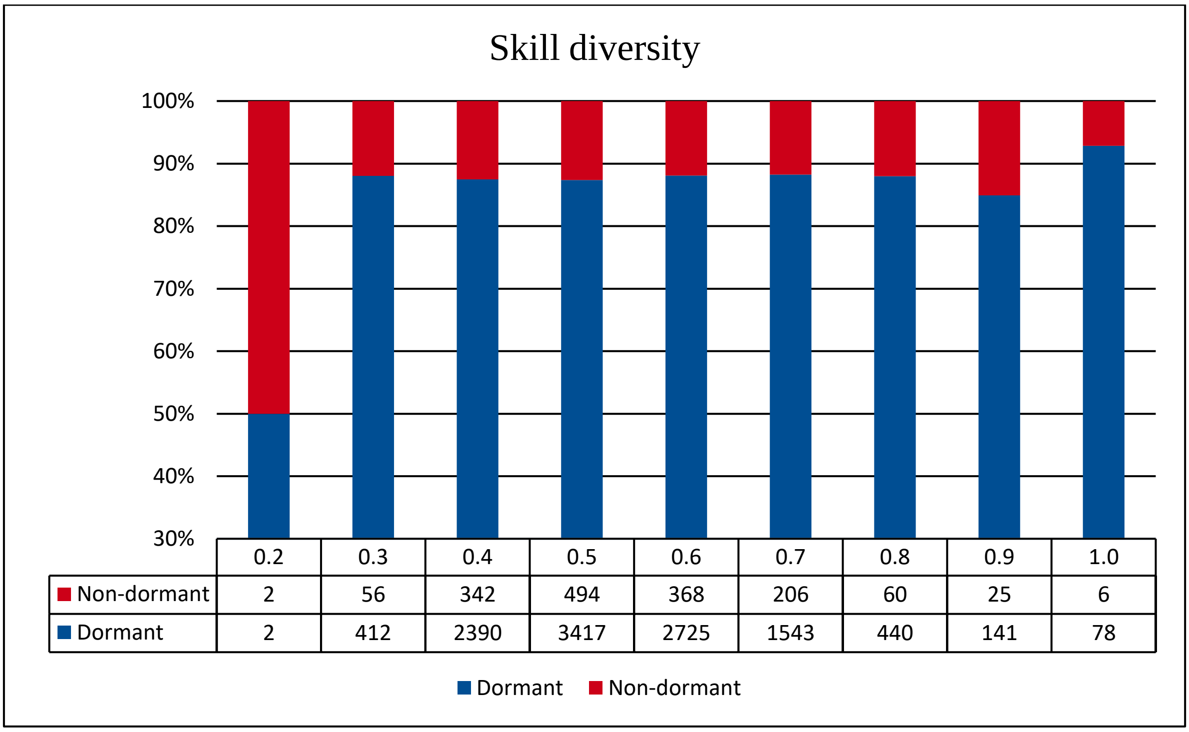
In the dataset, there were on average 15 winners per hackathon (SD = 9.0) with an average prize amount of $28 000 per hackathon, meaning that on average a winner received $1 850 in prizes, depending on the award setup, which is nowhere near to create something after the hackathon based on the prize. The amount of money in a prize pool ranged from having no cash rewards to $1 095 000 with a grand prize of one million dollars [30], which is a huge amount in terms of one hackathon. No other hackathons reached this level in the given data, as some of the hackathons had prize pools from $750 000 to $1 000 000 distributed between the winners. A large prize pool does not mean that the hackathon will be continued and the projects that come out of it are destined to succeed. In the given example of the hackathon with a grand prize of one million dollars, the outcome and decisions made by the board left many participants with questions and doubts about unbiased judging process [31]. To avoid negative publicity, the board decided to award two top teams both one million dollars, increasing the prize bool up to $2 095 000, which is unseen before on hackathons.

The amount of technologies used for a project ranged from 1-40 per team, the mean amount of technologies used was 5.5 with a standard deviation of 3.3, on the other hand the number of skills known by a participant ranged from 1 to 116, the mean amount of skills known was 4.7 with a standard deviation of 4.3.



	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Non-dormant	115	149	345	337	290	196	35	46	12	6	28
Dormant	1147	1140	2615	2068	1900	1360	174	298	117	31	298

Figure 5. Technology coverage distributed between dormant and non-dormant projects.

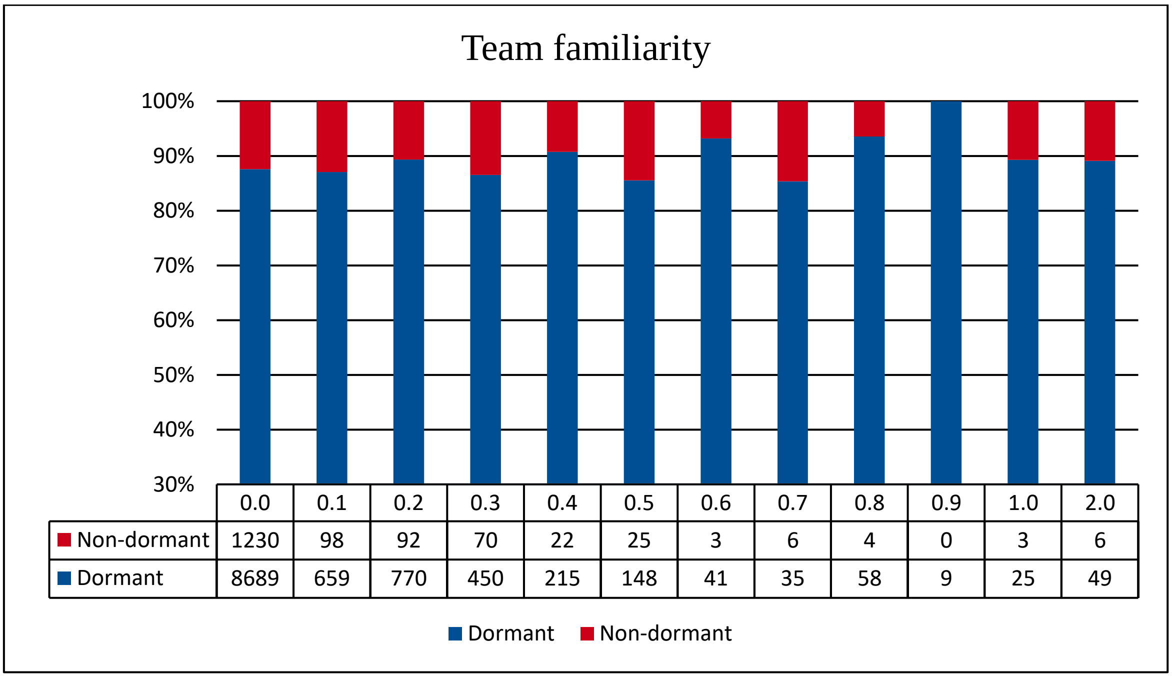


	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Non-dormant	2	56	342	494	368	206	60	25	6
Dormant	2	412	2390	3417	2725	1543	440	141	78

Figure 6. Skill diversity within the team distributed between dormant and non-dormant projects.

Figures 5 and 6 show that having a higher skill diversity or more technologies covered in a project did not necessarily mean that the project was continued after the hackathon. Based on the figures above, project's technology coverage was rather low compared to the skill diversity within the team as majority of the teams had skill diversity of 0.4-0.7, while at the same time project's skill coverage was mainly 0.0-0.5. In the dormancy context, skill diversity did not influence project's dormancy as much as technology coverage did. A higher technology coverage did have some effect on the amount of project's dormant after the hackathon.

This can mean that teams indeed attempted to gather people with different skillsets around them when forming a team at a hackathon with almost none of the projects having a skill diversity of zero but having a higher skill diversity within the team did not seem to give an expected result of having higher project technology coverage. On the other hand, a low technology coverage indicates that participants did learn something new at a hackathon.



	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	2.0
Non-dormant	1230	98	92	70	22	25	3	6	4	0	3	6
Dormant	8689	659	770	450	215	148	41	35	58	9	25	49

Figure 7. Team familiarity within the team distributed between dormant and non-dormant projects.

The most amount of times team members had worked together was 11 times with one participant attending up to 56 hackathons. Although working with friends is important for participants, based on Figure 7 the majority (78%) of the projects had members in the team that had not worked together before and the rest of the projects having a rather low team familiarity. Knowing each other did not also influence project's dormancy in a visible way – a higher or lower team familiarity within the team did not influence project's continuation.

A rather low team familiarity can mean that the participants did not know each other before, which indicates that the participants did manage to gain new connections while networking with new people.

5.2.2 Univariate regression analysis

Regression analysis was applied to each variable alone against the dependent variable, with the results shown in Table 5.

Table 5. Summary of relationships between each variable alone against the dependent variable using Cox's regression model.

Variable	**Parameter estimate**	**P-Value**	**Hazard Ratio (95% CI for HR)**
Team size	-0.023	0.012	0.98 (0.96-0.99)
Technologies covered	-0.058	0.214	0.94 (0.86-1.00)
Skill diversity	0.104	0.150	1.11 (0.96-1.28)
Team familiarity	-0.009	0.790	0.99 (0.93-1.06)
Win's weight	-0.312	2e-16	0.73 (0.69-0.77)

The size of a team (M = 3.4, SD = 1.0) with a negative regression coefficient indicates that a larger team reduces the risk of a project being dormant after the hackathon. HR 0.98 (CI = 0.96-0.99) means that each additional member in the team reduces the risk of dormancy by 2%.

Technology's coverage (M = 5.5, SD = 3.3) in a project is not statistically significant (p-value = 0.214) but the negative regression coefficient and a HR of 0.94 (0.86-1.00) indicates that adding to technology coverage reduces the risk of dormancy by 6% that is higher than team size's impact. On the other hand, the confidence interval ends at 1.00, which indicates that it has a small contribution to the difference in hazard ratio.

Skill diversity (M = 4.7, SD = 4.3) and team familiarity were found to be influencing projects before but within this research they are not statistically significant (p-value = 0.150 and p-value = 0.790 respectively).

Winning has a positive regression coefficient suggesting that a higher winning weight reduces the risk of dormancy, based on Table 5. HR of 0.73 (0.69-0.77) means that a unit of growth in winning weight reduces the risk of a project being dormant after the hackathon by 27%.

5.2.3 Multivariate regression analysis

After applying regression analysis on each variable alone, different variable combinations are chosen at first by statistical significance and afterwards attempting to find the best combination by hand-picking. The best combination and the results including regression coefficient, hazard ratio and p-value are shown in Table 6 and Figure 8, which shows the probability of project survival over time across all three variables included in the multivariate analysis. Based on Table 6, only the importance of winning influences project's continuation significantly. Other predictors, team size and covered technologies, do not improve project's success significantly.

Table 6. Summary of multivariate Cox analysis.

Variable	**Parameter estimate**	**P-Value**	**Hazard Ratio (95% CI for HR)**
Importance of winning	-0.310	<2e-16	0.73 (0.69-0.77)
Team size	-0.015	0.100	0.98 (0.96-1.00)
Technologies covered	-0.082	0.070	0.92 (0.84-1.00)

In relationship with other co-variates team size and technology coverage are not statistically significant but on the other hand the cross-effect of variables implies that a higher value of technology coverage influences project's survival more than a larger team. The HR for team size has remained the same but coverage's influence has grown by 0.02 (2%) and when holding other covariates constant, having a higher technology coverage would reduce the risk of dormancy by 8% per unit of coverage compared to 6% in a univariate model.

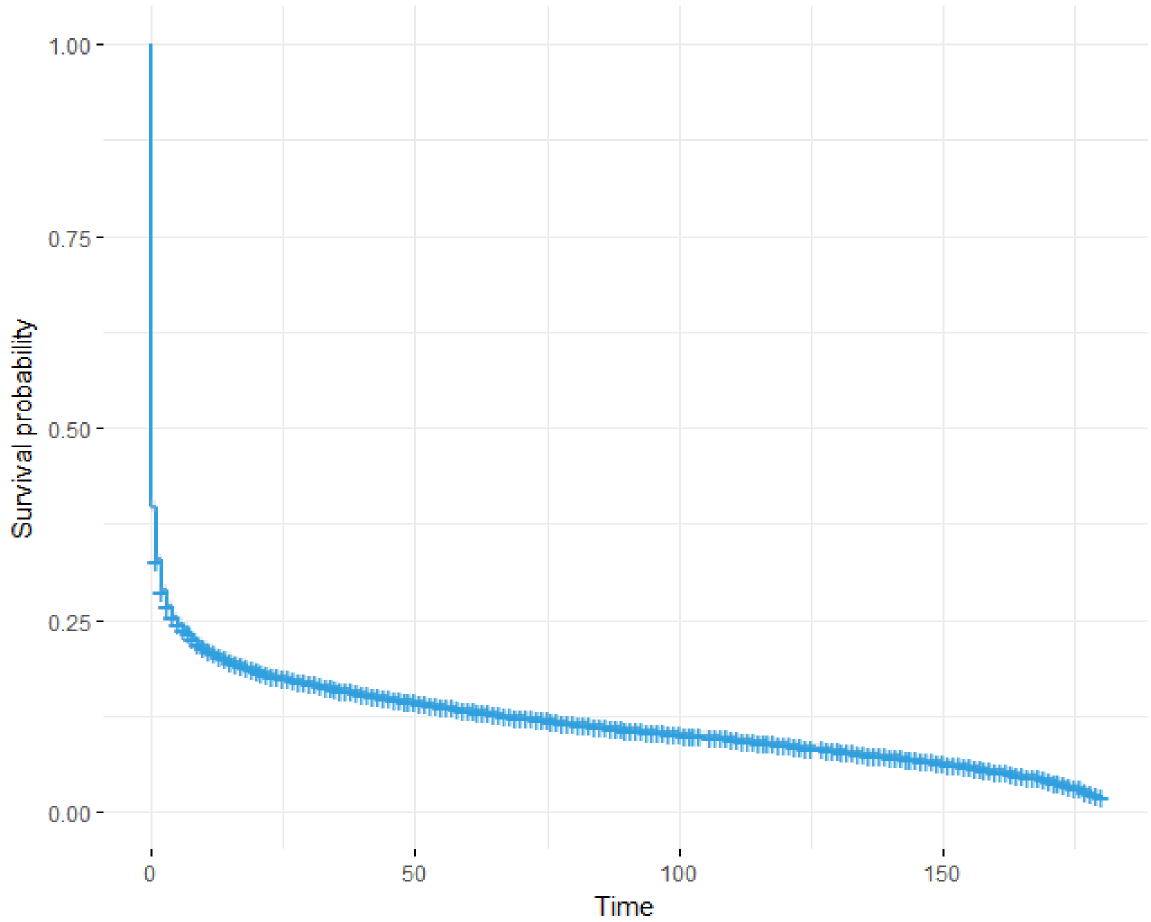


Figure 8. Overall probability of survival across the dataset.

Similarly to univariate model, the importance of winning continues to be statistically significant, with a HR of 0.73 implying its strong relationship between a higher winning weight and reduced risk of a project being dormant. When holding other covariates constant, a higher win's weight reduces the risk of dormancy by 27% per unit.

6 Discussion

The aim of the study was to explore collegiate hackathon projects on a large basis. Research on hackathons before had mainly focused on the hackathon event itself [6] and the ones focusing on hackathon projects were with a much smaller scope, usually referring to 4-5 projects at maximum [1], [3], [6], [11]. This thesis provides an understanding of what happened to the hackathon projects after the hackathon on a large basis (RQ1) and what could be the possible factors that could promote or hinder a project's continuation after the hackathon (RQ2).

This analysis led to identify that although there are many hackathons organized with each hackathon producing several projects, not all of the projects produce something tangible. It has been found before [1], [7], [14] that there is a 'clean-up' period after the hackathon to finish and produce presentable projects, this research adds to this knowledge by showing on a large basis that during the first five days the amount of projects turning inactive was the largest with 17% of growth.

On the other hand, the percentage of active projects turning inactive during the period from five days to the observed time of 6 months after the hackathon did not slow down with it capping at 97% of the active projects turning inactive during four to six months after the hackathon. This can mean that the clean-up or finishing the product was done much later than five days after the hackathon but at the same time the contribution per project was three to four times smaller during the first five days after the hackathon. This indicates that the real work to finish the product was not done during the first five days but just to clean up the code.

To research impact of factors on project's continuation, a choice of factors had to be made based on if there is any measurable data available for each of the factors. One of the chosen factor's was importance of winning, which has been found before to be important predictor in promoting project's continuation [2]. This research adds confidence to the belief by showing that the importance of winning is a very strong factor in promoting project's survivability after the hackathon with each unit of winning weight's growth the probability of continuing increases significantly. The percentage of dormant projects is relatively lower for teams that won something at a hackathon with only 27% of the teams winning a prize at a hackathon making it an achievement; on the other hand, there is no information of what the prizes were other than just funding. Then again, it is found that funding at some hackathons reached to an unreal level of having a prize pool from $750,000 up to over one million dollars, where a grand prize was for one team one million dollars in cash [30]. This can help a team to skyrocket their idea and it will definitely motivate the team but then again it is found to have negative publication after the hackathon due to suspects of unbiased judging resulting in the board giving two top teams one million dollar, raising the prize bool to over two million dollars [31].

Skill diversity within a team is something that has not been so deeply researched before but the researches have indicated that skill-matching is important to give the team a multi-dimensional view of the problem and to distribute the tasks based on each individual's skills [1], [5]. This research adds to the knowledge that teams attempt to gather participants with different skills to their team with the majority of the teams having a rather high skill diversity within the team but at the same time, skill diversity does not contribute significantly alone or in a relationship with other factors in promoting project's survivability. On the other hand, a high skill diversity within the team does not reflect in the technology coverage, which is rather low for majority of the teams. This indicates that new technologies are in use and the participants within the team do learn something new while participating at a hackathon which has been found before as one of the motivating factors for attending a hackathon [1], [2], [5], [6]. Technology coverage is found to not contribute significantly alone but in a relationship with win's weight, the significance of technology coverage improves considerably but is still considered as not significantly affecting the survivability. Participants have stated before the need to add confidence to their skills [5], [6], this research adds to this knowledge by indicating that although the technology coverage is rather low it still insists that participants did add experience to the skills they had previously known.

Similarly to skill diversity but unlike found before, this analysis led to identify that team familiarity does not have an impact on promoting project's survivability after the hackathon. The familiarity within the team was rather low with the majority (78%) of the projects having participants who had not worked together at all, although many researches before have found that knowing each other is important for the team members themselves [2], [6], [7], [14] and that it impacts project's continuation [1], [6]. On the other hand, a low familiarity in the team forces to connect with new people, which has been found important for personal growth before [2], [6], [7], [13], as it requires knowledge of fellow team members to work efficiently in a team.

Research done before has found that too large teams, in a particular case containing 11 members, felt overwhelming and was not effective as people started to step on each other's toes with taking each other's tasks [5]. This research identifies that teams with more than 11 members do not exist and majority of the teams ranged from 2-4 members per team within the used dataset, although it is identified that the number of participants in a team is significant in a univariate model even though having a very small contribution to promote project's continuation. In relationship with other factors, the factor is not statistically significant.

Out of five variables, only one ended up being statistically significant to promote project's continuation after the hackathon, which may mean that the operationalization of factors was done incorrectly, wrong factors were chosen or there are factors more important that this research did not find.

6.1 Implications

The findings presented in this paper have a number of implications for future research, some of them had been found before but the knowledge was expanded with this research, and what promotes or hinders project's continuation after the hackathon.

The results of this research indicate that the majority of the projects became inactive within the first five days after the hackathon, although the percentage of active projects turning inactive remained the same in the observed period of two month intervals. Not all projects turned out to have tangible outcomes but in some cases the project was finished and presented as a finished product [28]. In some occasions, a project became open-source after the hackathon for the community to make contributions on [25] and in some other occasions a new perspective was found and the original idea was halted [26], [27].

Participants that are looking to form a team at a hackathon with an idea of continuing their project after the hackathon ends should keep in mind factors that could promote their project's continuation. Winning a prize at a hackathon was found to have significant impact on promoting project's continuation but it is something that cannot be influenced directly by the hackathon participants, though the participants should prepare beforehand in order to enhance their winning possibilities. Together with the importance of winning, coverage of technologies in a project was found to have some positive impact on project's continuation. Concluding from this, technologies used in a project should be covered by as many team members as possible as it has some positive effect on project's continuation. Team size alone had an impact on project's continuation but the effect was rather small, each additional member decreased the chance of being dormant by 2%, to make any conclusions based on it.

On the other hand, familiarity within the team, skill diversity and team size in relationship with other variables did not have a significant impact on promoting or hindering project's continuation after the hackathon.

This research did raise new questions for future research. How to measure motivation within a team? What happens to teams before the hackathon on a large basis? Would a skill requirement before a hackathon have an impact on project's continuation? What happened to the projects that had their repositories turned private? Having mentors for each team in a large hackathon is impossible but is there be any other solution for experienced developers to help teams at big events like these? Are hackathons profitable for companies that organize them?

6.2 Limitations

The aim of the study was to see what happened to the hackathon projects and what could possibly promote or hinder project's continuation after the events had concluded on a large basis of 12 707 projects with 34 067 unique participants.

However, there were many limitations with the setup of this research. The original dataset included 60 478 projects with 103 026 unique participants but it had to be reduced due to massive invalid data and data not meeting the requirements, which means that a lot of possibly valuable information was lost. An additional limitation was the possibility to make a repository private, which reduced the amount of projects by 471.

Although before this research there had been found a handful of predictors that could possibly influence project's continuation, in this research many of them could not be measured due to limits of the dataset in hand. It was not possible to measure the motivation or how did the team work during the hackathon, what exactly a participant learned or did the participant connect with someone at a hackathon nor was it possible to measure popularity of a project, which could all be factors that could promote or hinder project's continuation at a hackathon and after it. A possible way to measure the missing factors would have been to make a survey within the teams but this did not fit in the research workload.

On the other hand, it was possible to measure if a team won something but it was limited to a true or false value, which meant that there was no information about what the prize that the team won was, limiting the research. In this research, hackathon location and the size of a hackathon was possible to measure with the available data but these values were not used in regression model because of being on a different level compared to the other variables. For example, projects in a hackathon all have the same location and amount of projects in the event but they are different in terms of other factors (skill diversity, team size …) making them incomparable.

7 Summary

This study focused on what happens to the projects after the hackathon and what could be the possible predictors that could promote or hinder a project's continuation after the hackathon where the participants are students. This research is different from others by the scale of the data used within the research, including 60 478 projects at first which was then reduced to 12 707 projects compared to similar studies consisting of up to 10 projects per study.

To answer the research questions, a data gathering was conducted to collect additional information of each project in the research that was followed by a data analysis and regression modelling to identify factors that could promote or hinder project's continuation after the hackathon.

From the analysis, it was found that not all teams produce something tangible and most of the projects die during the first weeks after the hackathon but it was also found that the percentage of active projects turning inactive during the period from five days to the observed time of 6 months after the hackathon did not slow down with it capping at 97% of the active projects turning inactive during four to six months after the hackathon, meaning that the clean-up is not always done during the first weeks.

Analysing the factors – team size, importance of winning, technology coverage, skill diversity and team familiarity resulted in importance of winning being the only factor in relationship with other predictors that influences project's continuation with a higher value of win's weight resulting in having a better chance of not turning dormant after the hackathon. When factors are kept separated, a larger team has a positive impact on project's survival.

The analysis did reveal new outcomes, identifying that the clean-up is not bound to happen right after the hackathon and although not all the projects produce something tangible there are projects that produce valuable outcomes. On the other hand, only one, the importance of winning, out of five variables ended up significantly influencing project's continuation after the hackathon.

8 References

[1] A. Nolte, E. P. P. Pe-Than, A. Filippova, C. Bird, S. Scallen, and J. D. Herbsleb, "You Hacked and Now What?: - Exploring Outcomes of a Corporate Hackathon," Proceedings of the ACM on Human-Computer Interaction, vol. 2, no. CSCW, pp. 1–23, Nov. 2018.

[2] G. Briscoe and C. Mulligan, "Digital Innovation: The Hackathon Phenomenon".

[3] M. Komssi, D. Pichlis, M. Raatikainen, K. Kindstrom, and J. Jarvinen, "What are Hackathons for?," IEEE Software, vol. 32, no. 5, pp. 60–67, Sep. 2015.

[4] E. H. Trainer, A. Kalyanasundaram, C. Chaihirunkarn, and J. D. Herbsleb, "How to Hackathon: Socio-technical Tradeoffs in Brief, Intensive Collocation," in Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16, San Francisco, California, USA, 2016, pp. 1116–1128.

[5] D. Cobham, C. Gowen, B. Hargrave, K. Jacques, J. Laurel, and S. Ringham, "From hackathon to student enterprise: An evaluation of creating successful and sustainable student entrepreneurial activity initiated by a university hackathon," presented at the International Conference on Education and New Learning Technologies, Barcelona, Spain, 2017, pp. 789–796.

[6] A. Nandi and M. Mandernach, "Hackathons as an Informal Learning Platform," in Proceedings of the 47th ACM Technical Symposium on Computing Science Education - SIGCSE '16, Memphis, Tennessee, USA, 2016, pp. 346–351.

[7] D. Cobham, C. Gowen, K. Jacques, J. Laurel, and S. Ringham, "From appfest to entrepreneurs: Using a hackathon event to seed a university student-led enterprise," presented at the International Technology, Education and Development Conference, Valencia, Spain, 2017, pp. 522–529.

[8] "Devpost - The home for hackathons," Devpost - The home for hackathons. [Online]. Available: https://devpost.com/. [Accessed: 05-Dec-2018].

[9] B. Kitchenham, "Procedures for Performing Systematic Reviews," p. 33.

[10] M. Drouhard, A. Tanweer, and B. Fiore-Gartland, "A Typology of Hackathon Events," p. 4.

[11] N. Taylor and L. Clarke, "Everybody's Hacking: Participation and the Mainstreaming of Hackathons," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18, Montreal QC, Canada, 2018, pp. 1–12.

[12] H. Lapp *et al.*, “The 2006 NESCent Phyloinformatics Hackathon: A Field Report,” Evolutionary Bioinformatics, vol. 3, p. 117693430700300, Jan. 2007.

[13] A. Carruthers, “Open Data Day Hackathon 2014 at Edmonton Public Library,” Partnership: The Canadian Journal of Library and Information Practice and Research, vol. 9, no. 2, Dec. 2014.

[14] B. Busby, M. Lesko, and L. Federer, “Closing gaps between open software and public data in a hackathon setting: User-centered software prototyping,” F1000Res, vol. 5, May 2016.

[15] G. Ray, J. B. Barney, and W. A. Muhanna, “Capabilities, business processes, and competitive advantage: choosing the dependent variable in empirical tests of the resource-based view,” Strategic Management Journal, vol. 25, no. 1, pp. 23–37, Jan. 2004.

[16] L. Hynes, V. Jordan, F. O’Dea, and M. Byrne, “Strength In Numbers Hackathon: Using a novel technology-focused brainstorming activity to engage stakeholders in intervention development.,” p. 8.

[17] M. Dima, Q. Mary, and Tarek E. Virani, “Hackathons: Why Co-location?,” p. 6.

[18] “About Major League Hacking,” Major League Hacking. [Online]. Available: https://mlh.io/about. [Accessed: 02-Dec-2018].

[19] K.-M. Uiga, “University of Tartu Software Engineering master thesis github: ‘What happens to all these hackathon projects?,’” 01-Mar-2019. [Online]. Available: https://github.com/martinuiga/ut_masterthesis. [Accessed: 20-Mar-2019].

[20] M. Valiev, B. Vasilescu, and J. Herbsleb, “Ecosystem-level determinants of sustained activity in open-source projects: a case study of the PyPI ecosystem,” in Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2018, Lake Buena Vista, FL, USA, 2018, pp. 644–655.

[21] S. Chatterjee and A. S. Hadi, Regression Analysis by Example. John Wiley & Sons, 2015.

[22] P. K. Andersen and R. D. Gill, “Cox’s Regression model for counting processes: A large sample study.” .

[23] R. Singh and K. Mukhopadhyay, “Survival analysis in clinical trials: Basics and must know areas,” Perspect Clin Res, vol. 2, no. 4, pp. 145–148, 2011.

[24] “Cox Regression (Proportional Hazards, Hazard Ratio) - StatsDirect.” [Online]. Available: https://www.statsdirect.com/help/survival_analysis/cox_regression.htm. [Accessed: 04-Mar-2019].

[25] “Bottender | Make Bots Your Way, Fast and Flexible,” Bottender. [Online]. Available: https://bottender.js.org/. [Accessed: 10-Mar-2019].

[26] ClientComm facilitates better communication between clients and case managers to increase the number of people who succeed on community supervision.: codeforamerica/clientcomm-rails. Code for America, 2019.

[27] “A communications platform for case managers in the Criminal Justice System.: slco-2016/clientcomm,” 30-Jul-2018. [Online]. Available: https://github.com/slco-2016/clientcomm. [Accessed: 12-Mar-2019].

[28] “FOLD is a platform allowing storytellers to structure and contextualize stories,” 31-Aug-2018. [Online]. Available: https://github.com/readFOLD/FOLD. [Accessed: 12-Mar-2019].

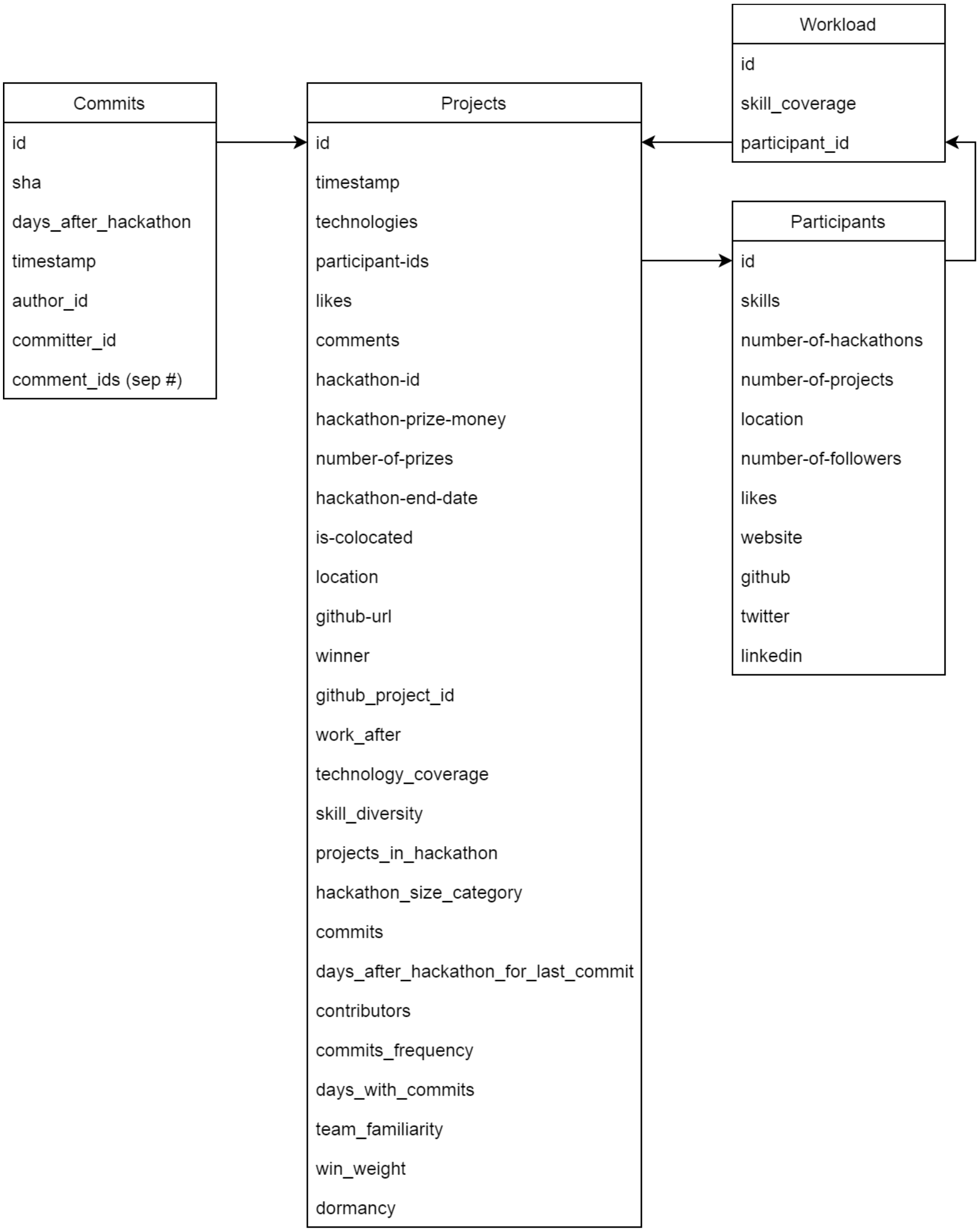
[29] S. Devalapurkar, “A web app that tutors can use to manage their students and schedule lessons: sdevalapurkar/schedulearn,” 21-Feb-2019. [Online]. Available: https://github.com/sdevalapurkar/schedulearn. [Accessed: 12-Mar-2019].

[30] “Salesforce $1 Million Hackathon,” Salesforce $1 Million Hackathon. [Online]. Available: https://salesforce1million.devpost.com/. [Accessed: 20-Mar-2019].

[31] “Update on Salesforce1 Hackathon Feedback & Review | Developer Force Blog.” [Online]. Available: https://developer.salesforce.com/blogs/developer-relations/2013/12/update-on-salesforce1-hackathon-feedback-review.html. [Accessed: 21-Mar-2019].

Appendix

I. Structure of the dataset



II. License

Non-exclusive licence to reproduce thesis and make thesis public

I, **Karl-Martin Uiga**,

(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

What happens to all these hackathon projects?,

(*title of thesis*)

supervised by Alexander Nolte.

(*supervisor's name*)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.
4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Karl-Martin Uiga

15/05/2019