



RESEARCH PARTICIPANT CONSENT FORM

Creating Impact in COVID-19 Data Science Contests

Principle Investigator Sabine Brunswicker

Research Center for Open Digital Innovation (RCODI)

Purdue University

Key Information

Please take time to review this information carefully. This is a research study. Your participation in this study is voluntary which means that you may choose not to participate at any time without penalty or loss of benefits to which you are otherwise entitled. You may ask questions to the researchers about the study whenever you would like. If you decide to take part in the study, you will be asked to sign this form, be sure you understand what you will do and any possible risks or benefits. The purpose of this study is to learn more about platform-enabled data-science contests. Such contests have become an important mechanism for solving COVID-19 data science challenges. In such contests, competitors have 2 to 4 weeks to design and improve their solutions in order to win.

What is the purpose of this study?

The purpose of this study is to evaluate how to design data science contests to increase the contests participant's productivity in performing data analysis related COVID-19. We specifically focus on virtual and platform-enabled data science contests (also referred as challenges). Such contests enable a large number of participants to use large volumes of machine-readable COVID-19 related data along with other platform features outside of physical and temporal constraints. In these contests you can win monetary prizes depending on your performance throughout the contest. In this study, you will be using an online platform called IronHacks (www.ironhacks.com) that has been developed by the RCODI over the last 4 years. It shares similar features like Kaggle.com, Topcoder.com. On this platform contests are called "IronHacks". More than 600 participants have finished a multiphase IronHacks contest. In this upcoming COVID-19 data science contests we aim to enroll up to 1000 participants to find a novel statistical model that help us fight the COVID-19 pandemic. Participants need to be older than 18 years to participate in this study.

What will I do if I choose to be in this study?

If you choose to take part in the study, you will be asked to register on our platform, IronHacks (www.ironhacks.com), select the "IronHacks" that is open for registration, and to start a multiphase hacking process. Each challenge has a specific number of large volumes of datasets (big data) associated with it that you can use for free for the purpose of this research study. The platform offers you access to cloud-based programming environment that allows you to manipulate this data without having to worry about local storage and computing. There are additional features that will facilitate your work (e.g. tutorials, result page, forum).

Throughout the multiphase challenge process, you are asked to submit several versions of your solutions. You are free to decide when you make such submissions. As long as your submission is submitted for evaluation, we will evaluate your progress. We will publish your evaluation scores very fast (e.g. within 24 h). This gives you the opportunity to improve. Winners will be awarded at the end of the contest. We will use an objective evaluation method that is detailed in the task description (e.g. the accuracy of your prediction compared to observed data). Winners will be awarded at the end of the contest based on a rank-order logic (see below section incentives). Only those participants that complete all phases in the contests are eligible for awards. Throughout the contest, we will send you short surveys that ask about the tools and techniques that you used, and your experiences during the contest.

If you start a contest, and then decide that you no longer wish to participate, you are able to leave the contest at any time. Participation in the challenges is voluntary, as is the completion of the survey.

How long will I be in the study?

This study will cover the length of contest. It may vary between 2 to 4 weeks. There will be between 2 to 5 phases and submission deadlines. You are not required to work on the task at specific time of the day or week. It is up to you how much time you invest in order to improve your standing in the contest. In the past, participants spend between 60 minutes to several hours to prepare a new version of their solution for each submission phase. As long as you meet the deadlines for submissions and short surveys we will evaluate your progress and your current standing in the contest. There will be a short survey at the beginning of the contest, and additional short surveys at each time of the submission, followed by a final survey. The surveys at the beginning and the end take about 20 minutes. The short surveys throughout take about 5 to 10 minutes.

What are the possible risks or discomforts?

There are minimal risks associated with participation in this study. Activities undertaken by participants are similar to those in most virtual contests and challenges run on well-known platforms like Kaggle.com, TopCoder.com, and Challenge.com used by millions of people. You might face some discomfort due to low internet connectivity at your home/working location. While the interconnectivity at your place is out of our control, we try to accommodate for such discomfort by ensuring that your work is saved in the cloud, even if you get disconnected. Breach of confidentiality is always a risk with data, but we will take precautions to minimize this risk as described in the confidentiality section.

Are there any potential benefits?

There are potential direct benefits from participating in this study. You can apply your skills in data science by working on real-world data that you typically would not have access to. Further, we offer you cloud-based infrastructure that allows you to work with big datasets without having to spend money on hardware or software. You will also have the opportunity interact with our partners like the Chief Data Scientists of the State of Indiana or other representatives from governmental agencies that are supporting our work.

Will I receive payment or other incentive?

Following other data science contests offered on similar platforms like IronHacks (e.g. Kaggle.com) we will be offering place prizes and complementary monetary incentives. In total, you can win between \$40 to \$500 dollars. These payments are slightly lower but still comparable to those on other contest platforms. We will award the place prizes to the top-3 solutions based on their ranks achieved in the contest: (1) Best Solution, (2) Runner-up, and (3) Third Place. The Best Solution may receive a prize money up to \$500. The prize award for the second or third place will be slightly lower. If you do not make it to the top 3 you can win up to \$40 in categories like improvement spirit and community mindset. As specified in the task statement of the contest, such awards are either based on your individual or our group's improvement throughout the contest. Prizes will be paid in gift cards (e.g. via Amazon) to facilitate digital distribution.

Are there costs to me for participation?

There are no anticipated costs to participate in this research.

Will information about me and my participation be kept confidential?

The project's research records may be reviewed by the study sponsors, US DHHS Office for Human Research Protections, and by departments at Purdue University responsible for regulatory and research oversight. This study is financially supported by the National Science Foundation as well as the Research Center for Open Digital Innovation (RCODI).

During the contest, your confidentiality will be maintained using a pseudonym system, where you will remain anonymous on pages of the website that are accessible by other participants (e.g. forum). Personal identifiers (e.g. registration emails and location) will only be available only to the research team, and will be maintained on our secured technical infrastructure. Only de-identified data will be used for future research and/or publication.

At the end of the contest, we will request permission from the creators of the top 3 solution to reveal the identity so that they can attend a final virtual award ceremony with sponsors and partners (e.g. the Chief Data Scientists of Indiana). This award ceremony will be live streamed. After that, we will feature winners and other selected solutions on our website and via social media channels. We will not display the name of the participants unless you give us permission to do so.

What are my rights if I take part in this study?

You do not have to participate in this research project. If you agree to participate, you may withdraw your participation at any time without penalty.

If you have questions, comments or concerns about this research project, you can talk to one of the researchers via email to opendigital@purdue.edu. If you have questions about your rights while taking part in the study or have concerns about the treatment of research participants, please call the Human Research Protection Program at (765) 494- 5942, email (irb@purdue.edu) or report anonymously via Purdue's Hotline: see www.purdue.edu/hotline.

Documentation of Informed Consent

I have had the opportunity to read this consent form and have the research study explained. I am prepared to participate in the research study described above. I can screen print this informed consent document for my own records. By checking the appropriate checkbox in the survey AND clicking the link "NEXT", I agree to participate in the study described above.