

Join the Women in Data Science (WiDS) Datathon

- 280 teams
- a month to go (a month to go until merger deadline)

1. Description

In advance of the first 24-hour virtual Women in Data Science (WiDS) Worldwide Conference March 8, 2021, we invite you to build a team, hone your data science skills, and join us for the 4th Annual WiDS Datathon focused on social impact. [Register now!](#)

The WiDS Datathon 2021 focuses on patient health, with an emphasis on the chronic condition of diabetes, through data from [MIT's GOSSIS \(Global Open Source Severity of Illness Score\)](#) initiative. Brought to you by the WiDS Worldwide team at Stanford, the [West Big Data Innovation Hub](#), and the [WiDS Datathon Committee](#), this year's datathon is open until March 1, 2021. Winners will be announced at the WiDS Conference via livestream, reaching a community of 100,000+ data enthusiasts across more than 85 countries.

Background

Getting a rapid understanding of the context of a patient's overall health has been particularly important during the COVID-19 pandemic as healthcare workers around the world struggle with hospitals overloaded by patients in critical condition. Intensive Care Units (ICUs) often lack verified medical histories for incoming patients. A patient in distress or a patient who is brought in confused or unresponsive may not be able to provide information about chronic conditions such as heart disease, injuries, or diabetes. Medical records may take days to transfer, especially for a patient from another medical provider or system.

Knowledge about chronic conditions such as diabetes can inform clinical decisions about patient care and ultimately improve patient outcomes. Learn more about the global scale of diabetes on our [Datathon News page](#), and make sure you are subscribed to the [WiDS Datathon Mailing List](#) to receive the latest updates, tutorials, and articles.

Overview

This year's challenge will focus on models to determine whether a patient admitted to an ICU has been diagnosed with a particular type of diabetes, Diabetes Mellitus. Using data from the first 24 hours of intensive care, individuals and teams will explore labeled training data for model development. Participants will then upload predictions for unlabeled data to Kaggle and these predictions will be used to determine the public

leaderboard rankings, as well as the final leaderboard revealed at the close of the competition.

Data analysis can be completed using your preferred tools. Tutorials, sample code, and other resources will be posted throughout the competition at widsconference.org/datathon-resources and on the [Kaggle Discussion Forum](#).

The WiDS Datathon 2021 dataset is similar to the [WiDS Datathon 2020](#) dataset, but prior experience is not needed. This year's WiDS Datathon will also feature a Phase 2 hosted by the National Science Foundation [Big Data Innovation Hubs](#), to encourage research papers and further collaboration. Special thanks to Kaggle for supporting the suite of WiDS Datathon cash awards this year, totaling \$10,000 USD.

Who can participate

We invite anyone from those new to data science to veterans of the field to participate. For those who have never tried machine learning or worked with health data before, check out the [series of tutorials and webinars](#) from the WiDS Datathon Committee, workshop organizers, and community of participants to help you get started.

The WiDS Datathon aims to inspire women worldwide to learn more about data science, and to create a supportive environment for women to connect with others in their community who share their interests. Toward these ends, we open the datathon to individuals or teams of up to 4; at least half of each team must be women (individuals identifying as female participants). Participants can include students, faculty, and individuals with various roles in non-profit, academic, government, and industry organizations.

Acknowledgements

The WiDS Datathon 2021 is a collaboration led by the WiDS Worldwide team at Stanford University, the [West Big Data Innovation Hub](#), and the [WiDS Datathon Committee](#). WiDS Datathon 2021 cash prizes are provided by Kaggle and the Excellence in Research Award is supported by the National Science Foundation under Grants 1916573, 1916481, and 1915774, as part of a network of Big Data Innovation Hubs. Special thanks to the MIT GOSSIS Initiative and the Harvard Data Privacy Lab.



The WiDS Datathon is made possible by the following WiDS Global Visionary Sponsors:



2. Evaluation

Submissions for the leaderboard will be evaluated on the Area under the Receiver Operating Characteristic (ROC) curve between the predicted and the observed target (**diabetes_mellitus_diagnosis**).

Learn more about AUC in [this developer crash course](#), [this video](#), or [this Kaggle Learn Forum post](#).

Submission Format

For each **encounter_id** in the test dataset, you are asked to explore the columns of data (for example, patient laboratory results, demographics, and vital signs) and create a model that classifies whether the ICU patient has been diagnosed with Diabetes Mellitus.

A **diabetes_mellitus** value of 1 corresponds to a positive diagnosis (i.e., the patient has been diagnosed to have Diabetes Mellitus) and a value of 0 corresponds to a negative diagnosis.

Your submission file should contain a header and have the following format:

```
encounter_id, diabetes_mellitus  
1,0.793
```

3. Prizes

Prizes will be awarded for the top three leaders on the Kaggle leaderboard at the close of the competition, as well as through a [Phase 2](#) focused on research papers.

Leaderboard Prizes

1st Place on the Kaggle Leaderboard

- \$2,000 USD cash prize for the team
- 1 complimentary VIP registration per team member for the WiDS Worldwide Conference in 2022

2nd Place on the Kaggle Leaderboard

- \$1,500 cash prize for the team
- 1 complimentary VIP registration per team member for the WiDS Worldwide Conference in 2022

3rd Place on the Kaggle Leaderboard

- \$1,000 cash prize for the team
- 1 complimentary VIP registration per team member for the WiDS Worldwide Conference in 2022

Each leaderboard prize winner will be announced at the first 24-hour virtual WiDS Worldwide Conference, March 8, 2021.

Phase 2 Prizes

Excellence in Research Award

- \$3,000 USD cash prize for the team
- 1 complimentary VIP registration per team member for the WiDS Worldwide Conference in 2022
- A WiDS Datathon Excellence in Research Award Certificate for each team member
- An invitation to be featured in a #WiDStory article

In addition to the Excellence in Research Award, up to \$2,500 USD will be distributed for honorable mentions across submitted research papers.

All winners will be recognized on WiDS social media channels. Special thanks to Kaggle for providing and distributing the \$10,000 in cash prizes.

4. Datathon Phase 2: Excellence In Research Award

WiDS Datathon Phase 2: Excellence in Research Award

As you work with the WiDS Datathon 2021 data, are you...

...considering how your classification models perform across features such as gender, ethnicity, and age?

...diving into the implications of missing values and the "messiness" of real-world

Intensive Care Unit data?

...wondering how you can contribute to the broader health data community beyond the WiDS Datathon 2021 leaderboard?

To help encourage deeper exploration and the development of collaborative data science innovations, the [WiDS Datathon Committee](#) and the National Science Foundation [Big Data Innovation Hubs](#) are hosting a second phase of the WiDS Datathon — supporting a **WiDS Datathon Excellence in Research Award**.

We invite you to share more details about your work with WiDS Datathon 2021 data through a research paper of up to 3 pages, by **April 30, 2021, 11:59pm PT**. Check back here for the submission link in March!



Submissions will be reviewed by subject matter experts from the WiDS Datathon Committee, the National Science Foundation Big Data Innovation Hubs, and the MIT GOSSIS community. Papers will be evaluated on their **potential for real-world impact, rigor in scientific methodology, and clarity of communication**.

We request submission content in English to facilitate the review process, but welcome additional translations and encourage participants to share their papers broadly. Papers will also be eligible to be published on the [WiDS Datathon website](#).

Papers should address questions including but not limited to:

- What assumptions did you make in preparing your models and analyzing the data?
- What are the potential implications of your assumptions?
- What are your main insights and contributions?
- What are some of your 'lessons learned' and meaningful or surprising results?
- What unanswered questions do you have about the data, your analysis, and the analysis from the research community?
- How might you and others address these unanswered questions and/or advance related research?
- Are there open problems in mathematics, statistics, computer science, applied medical research or related fields that could benefit from deeper exploration of this data set?

Submissions should also include an **abstract** (4-5 sentence summary) and a **list of references** (e.g., cited research papers, books, tutorials, or other resources used). These research paper templates are provided for convenience: [Word format](#) and [PDF](#). Many other examples are available, e.g. <https://www.overleaf.com/latex/templates/tagged/academic-journal>.

Award Details

The individual or team will receive:

- **\$3,000 USD cash prize for the team**
- **1 complimentary VIP registration per team member for WiDS 2022**
- **A WiDS Datathon Excellence in Research Award Certificate for each team member**
- **An invitation to be featured in a #WiDStory article**

In addition to the Excellence in Research Award, up to \$2,500 USD will be distributed for honorable mentions across submitted research papers.

Phase 2 winners will be recognized via WiDS and the Big Data Innovation Hubs' social media channels. Special thanks to Kaggle for providing the cash prizes for the winners.

Eligibility

To be eligible for the award, all entrants must have participated in the first phase of the WiDS Datathon 2021 on Kaggle. Individuals and teams should submit their papers with the same names and contact emails used for WiDS Datathon 2021 registration. The Submittable website will open by March 9, 2021, and papers are due April 30, 2021.

5. FAQ

Feel free to ask questions in the [Kaggle Discussion Forum](#) throughout the competition. Below are some frequently asked questions and details. We look forward to your submissions!

I plan to participate in the WiDS Datathon with a team of up to 4 people. Do we all need to register separately?

Yes, each individual who participates in the WiDS Datathon must register with us by filling out [this form](#).

What local Datathon-focused workshops are being hosted?

[WiDS 2021 ambassadors](#) are hosting WiDS Datathon in-person mentorship workshops, webinars, and tutorials worldwide to get you acquainted with the WiDS Datathon. Check out the [WiDS Datathon Workshops Map](#) and browse some of the [tutorials and resources](#).

How many authors are allowed for the Excellence in Research Award submission?

We encourage individuals and author teams of up to 4 people to work together, since we can support up to 4 VIP tickets and prizes for the [Excellence in Research Award](#) winner. You are welcome to acknowledge additional collaborators and mentors directly in the paper, through an extended blogpost or [Kaggle Discussion Forum](#) post, or other venues.

I am part of a team for the first phase of the Datathon on Kaggle, but only a portion of the team is interested in writing a paper. Can I still participate in the Datathon Phase 2: Excellence in Research Award?

Yes, you are welcome to collaborate with individuals in a “new” team for the second phase of the Datathon, as long as (1) each person participated in the first phase of the datathon on Kaggle and (2) your new team still includes 50% or more individuals identifying as female. The WiDS Datathon aims to inspire women worldwide to learn more about data science, and to create a supportive environment for women to connect with others in their community who share their interests.

When/how will the papers be reviewed for the Excellence in Research Award? Can I submit my work for publication elsewhere?

After the April 30, 2021 research paper deadline, submissions will be reviewed for their potential for real-world impact, rigor in scientific methodology, and clarity of communication, by subject matter experts from the [WiDS Datathon Committee](#), the National Science Foundation [Big Data Innovation Hubs](#), and [MIT GOSSIS](#) community. The paper content will be treated as confidential in order to enable participants to submit their work to scientific journals and conferences if they wish (please refer to specific guidance from publishing venues).

Does my leaderboard standing impact my eligibility and consideration for the Excellence in Research Award?

No, leaderboard placement does not impact the WiDS Datathon Excellence in Research Award. As long as you have submitted an entry in the first phase of the Datathon (by March 1, 2021), it does not matter if you are ranked first or last on the leaderboard. All are encouraged to write a paper that shares their perspective and contributions.

Can I include results and work that I complete after the Kaggle leaderboard closes March 1, 2021, in the Excellence in Research Award paper submission?

Yes, the second phase of the datathon is intended to encourage deeper exploration. If you are interested in working with similar data even beyond the April 30, 2021 research paper

deadline, you can visit the [eICU Collaborative Research Database](#), [MIMIC](#), and [PhysioNet](#). Please consider sharing your stories and future work with us using #WiDSDatathon and via email to widsdatathon@stanford.edu!

6. Timeline

Datathon Timeline

February 26, 2021: Team Merger deadline. This is the last day competitors may join or merge teams.

March 1, 2021: Final Kaggle submission deadline to be eligible for leaderboard prizes.

March 8, 2021: Datathon Leaderboard Winners will be announced at the WiDS Worldwide Conference.

March 9, 2021: Datathon Phase 2: Excellence in Research Award submissions open, on Submittable.

April 30, 2021 11:59pm PT: Datathon Phase 2: Excellence in Research Award submissions close.

August 2021: Datathon Phase 2: Excellence in Research Award winners announced.

All deadlines are at 11:59 PM UTC on the corresponding day unless otherwise noted. The competition organizers reserve the right to update the contest timeline if necessary. Please join the [community mailing list](#) to receive all updates and announcements.