**NATIONAL UNDERGRADUATE BIG DATA CHALLENGE 2021**

***Infodemiology for the Future of Digital and Public Health***

***Explore practical applications and insights of infodemiology to discover breakthrough connections in Digital and Public Health using open social, demographic, and health data***

Organized by [STEM Fellowship](http://stemfellowship.org/) in partnership with

[JMIR Publications](https://www.jmir.org/), [Roche](https://www.roche.com/), [SAS](https://www.sas.com/en_us/home.html), [Canadian Science Publishing](http://www.nrcresearchpress.com/), [Digital Science](https://www.digital-science.com/), [Overleaf](https://www.overleaf.com/)

Behind every innovation and discovery there are troves of foundational data. To drive and verify these innovations, representative data and the ability to handle it is a necessity. In 2020, it is projected that there will be 40 zettabytes (40 x 1021 bytes) of data. The organization and handling of this information has become a new science - the science of Big Data. Consequently, digital literacy and data analytics are increasingly becoming essential skills, allowing for more powerful results and applications in many different fields.

The Big Data Challenge (BDC) for undergraduate students is a unique inquiry-driven and experiential learning program that invites students from across the country to strengthen their problem-solving and critical thinking skills while gaining familiarity with the fundamentals of data science. By allowing students to undertake independent research projects that tackle real-world public health and bioinformatics problems, the BDC fosters scientific inquiry and prompts new and innovative ideas.

It is through educational systems that some of the brightest minds and innovators are found. However, education and access to it are not uniform. Gender, race, geographical region, socioeconomic status, and citizenship status are all significant, nuanced factors in this discussion. We challenge students to examine Open Data related to their learning experiences, and to bring their results to their professors/teachers, classrooms, and peers.

Teams of up to 4 students are each provided with data sets, workshops, learning resources and tools for data analysis. With the help of peer-mentors and the guidance of experts in academia and industry, teams undertake exploratory analysis of accessible and open health and socioeconomic data to develop public and individual health initiatives and solutions. Data analysis is combined with scientific writing, insofar that the teams present their research findings in the form of scientific manuscripts, which are then evaluated by academics and industry professionals. All aspects of the BDC, including the delivery of workshops, resources, and mentorship, will occur online and are equally accessible to all students regardless of their location or other circumstances.

At the end of the program, the research abstracts of all teams and the manuscripts of winning teams are published in the open access, peer-reviewed [NRC Research Press](https://www.nrcresearchpress.com/page/about) [STEM Fellowship Journal](https://journal.stemfellowship.org/journal/sfj) . The top teams are then invited to defend their findings in front of a panel of experts in competing for monetary and academic prizes, at the culminating finale event.

**PROBLEM STATEMENT**

|  |
| --- |
|  |

This year’s national and inter-university Big Data Challenge will encourage undergraduate students across the country to apply [infodemiological](https://www.jmir.org/themes/69) computational methods to open social and health-related data, with the ultimate goal of combating infodemics, to improve digital and public health. We look forward to your unique solutions and research into patterns of health misinformation and mechanisms, to validate publicly available data and use it to inform health outcomes. The challenge will be based on the analysis of [open data resources gathered](https://docs.google.com/document/d/1wSmMN2G2LotIh5oIvmmZv9SgS58xOCXxhv4nJUtycec/edit?usp=sharing) by STEM Fellowship and other open data sources identified by students.

**Through this challenge, students will:**

* **Investigate** how people search and navigate the internet for health-related information, as well as how they communicate and share this information, to provide valuable insights into the health-related behaviour of populations.
* **Identify** types of viral health information/misinformation patterns, and **suggest** solutions and decisions in various areas of health and science communication that impact both individuals and the general public.
* **Collect** and **analyze** data to investigate the validity of health claims within the media, by critically considering the implications of the data and attempting to establish objectivity.
* **Characterise**, **understand**, and **predict** the public health and socioeconomic effects of potentially disingenuous health claims within the media.
* **Hypothesize, Formulate** and **Present Findings** in the form of a scientific report.

**Why participate?**

* Develop **analytical** and **computational thinking** by using **computational techniques** in the context of current, real-world challenges in education.
* Learn **data visualization** to present student-found results from Big Data analytics.
* Engage in an **interdisciplinary**, problem space led by student-driven inquiry.
* Practice **scientific writing** and publish your ideas in the peer-reviewed [STEM Fellowship Journal](https://journal.stemfellowship.org/journal/sfj), through the largest national scientific publishing group.
* **Network** with individuals within industry and academia throughout the competition, in solving common goals.

**Research topics can include any fields of Infodemiology, such as, but not limited to:**

* The impact of spread of health misinformation on influencing public perception/actions, targeting certain cultural/ethenic communities, increasing polarization/stigmatization.
* Devising patterns, trends or models to differentiate between accurate and inaccurate information found on digital platforms (social media, search engines etc.).
* The influence of, or correlation between political, education, economic or social ideologies on promoting/interacting with digital and public health misinformation.
* Analyzing the effectiveness/validity of employed methods on digital platforms to remove/flag publicly shared data containing inaccurate information.
* Detecting and quantifying disparities in health information availability

The BDC offers top teams the highly sought-after opportunity to present their ideas to industry and academic experts, as well as publish their results in a scholarly journal.

Note: Undergraduate students from all majors and programs at any Canadian institution are welcome to participate and there is no minimum GPA required.

**PRIZES**

|  |
| --- |
|  |

**ACADEMIC PRIZES:**

* **JMIR Internships:**For select competition winners, [JMIR Publications](https://www.jmir.org/) will be providing internship opportunities allowing students to work on projects related to Infodemiology. JMIR Publications is one of the first open–access publishers in the world, helping scientists disseminate innovations, ideas, protocols, and research results to the widest possible audience. They innovate in the scholarly communication space itself, experimenting with novel metrics, new business models, new models of peer review and dissemination, and new technologies.
* **Opportunities with Roche**:

[Roche Canada](https://www.rochecanada.com/) will provide a winning team the opportunity to interact with Roche to learn how big data can be used in the pharma setting to address pertinent questions related to health outcomes for patients. Students will also have the opportunity to network with various Roche employees and learn more about careers in Pharma and big data.

* **Scholarly publication** of manuscripts for 3 winning projects in the peer-reviewed [STEM Fellowship Journal](http://journal.stemfellowship.org/),published by the [NRC Research Press](http://www.nrcresearchpress.com/)**.**

**MONETARY PRIZES:**

* Hoffmann-La Roche Infodemic Research Solution Awards
  + 1st Prize: $1000 + Publication
  + 2nd Prize: $600
  + 3rd Prize: $400
* Overleaf Outstanding Science Communication Awards
  + 1st Prize: $1000 + Publication
  + 2nd Prize: $600
  + 3rd Prize: $400
* JMIR Infodemic Innovation Award
  + $1000 + Publication

**SCHEDULE & DEADLINES**

|  |
| --- |
|  |

**May 21, 2021 - Registration Deadline**

* Form your team of up to 4 students and register [here](https://stemfellowship.org/unbdc-2021-registration-page/).

**May 22, 2021 - Competition Begins**

* All teams will have been matched with peer and expert mentors and provided with datasets and learning resources by this date.
* Work together with your team, making use of your mentors, teachers, and the provided resources to analyze your data and propose solutions.
* Tell the story of your data discovery through a scientific report. Use [Overleaf](https://www.overleaf.com/) to prepare your project report, and submit it to us at [bigdata@stemfellowship.org](mailto:bigdata@stemfellowship.org)
* In addition to formal mentorship, the STEM Fellowship Data Science Team members will support participating students throughout the challenge period. [Access Workshop Schedule, Meeting/Registration links, Resources etc.](https://docs.google.com/spreadsheets/d/1TjT2Qi2_Xa6USgbca-sQoZVDN8KA7WZmzOigWCRoVbA/edit?usp=sharing)
  + **May 22nd 2021 5:00 PM - 6:00 PM ET:** Statistics and Data Handling in Jupyter
  + **Individual Workshop:** Data Visualization
  + **Individual Workshop:** Linear Regression
  + **May 25th 2021 5:00 PM - 6:00 PM ET:** Classification using Logistic Regression + Clustering (K-means algorithm)
  + **May 27, 2021 3:00 PM - 4:00 PM ET:** Overleaf Webinar - Intro To Overleaf

**May 31, 2021 - Project Report Submission Deadline**

* All participating students will submit a project report (developed in Overleaf) before the deadline **11:59PM PT** to [bigdata@stemfellowship.org](mailto:bigdata@stemfellowship.org), for evaluation by a team of academic and industry experts.

**June 21, 2021 - Finalists Announcement**

* The 15 finalists will be announced!
* Finalist teams will submit a video presentation by July 9th, improve their project, and be judged by an expert panel at Big Data Day - the culminating event for the BDC.

**July 17, 2021 - Big Data Day and Announcement of Winners**

* Competing finalist teams will each be questioned on their 10 minute pre-recorded presentation to a panel of academia and industry experts, who then determine the award winners.
* There is an open roundtable discussion (featuring academia and industry experts) relating to the BDC theme.
* Students network with dignitaries, academic and corporate leaders at Big Data Day.

**COVID-19 CONCERNS**

|  |
| --- |
|  |

At STEM Fellowship, we care about the health and safety of our participants and the community at large. The Big Data Challenge will be conducted entirely remotely and online this year. We will have the Big Data Challenge finalists submit a video presentation and have judges evaluate them remotely.

Further, we will host video-conference interviews with academia and industry experts, as well as video presentations of speakers instead of the live round table and presentations at Big Data Day.

**WHO TO CONTACT WITH QUESTIONS**

|  |
| --- |
|  |

Anish Verma - Chief Data Officer [anish.verma@stemfellowship.org](mailto:anish.verma@stemfellowship.org)