**MATDAT18: Materials and Data Science Hackathon**

**Time and Place**

May 30-June 1, 2018

NSF Headquarters, Alexandria, VA

**Organizers**

Andrew Ferguson, Materials Science and Engineering, University of Illinois

Tim Mueller, Materials Science and Engineering, Johns Hopkins University

Sanguthevar Rajasekaran, Computer Science & Engineering, University of Connecticut

Brian Reich, Department of Statistics, North Carolina State University

*Primary Contact*: brian\_reich@ncsu.edu

*MATDAT18 Webpage*: <https://matdat18.wordpress.ncsu.edu/>

**Synopsis**

Increases in computing power and advances in high-throughput instrumentation has led to the generation of computational and experimental materials science data sets of unprecedented size. Researchers are increasingly turning to data science tools to analyze these data to extract understanding, and perform high-throughput screening and data-driven design. An impediment to success is that materials experts may not be experts in data science, and data scientists typically lack the domain-specific expertise in materials engineering. It is the goal of this 3-day “hackathon” to partner materials and data scientists within interdisciplinary teams to spark collaborative research partnerships. Materials researchers will develop fluency in statistical and machine learning techniques, and data scientists will be exposed to data-centric problems in materials engineering. **Full financial support is available to all participants.**

**Application Instructions**

**Step 1 – Solicitation of data-centric projects from materials researchers.**

***Deadline: January 15, 2018***

Materials scientists interested in proposing a project for the hackathon should complete the attached application form and submit via email to Brian Reich (brian\_reich@ncsu.edu). A (non-exhaustive) list of sample projects is provided below as examples of possible topics.

**Step 2 – Release and advertisement of materials projects to data scientists.**

***Deadline: February 15, 2018***

The organizers will sort the submitted projects, host them online, and advertise to the data science community through conferences, publications, and workshops. Data science applicants should complete the online application through the webpage at: <https://matdat18.wordpress.ncsu.edu/>. The organizers will pair teams, perform remote introductions, and support preliminary preparation and goal setting in advances of the hackathon.

**Step 3 – Solicitation of intention from data scientists to work on specific materials projects.**

***Deadline: March 16, 2018***

Data scientists interested in working on any specific materials project(s) identified by the organizers (in Step 2 above) should complete the attached application form and submit via email to Brian Reich (brian\_reich@ncsu.edu). Please save the file as MATDAT\_#name#\_#number#.docx where #name# is your name and #number# is the project number (1-19) given on the website project page.

**Step 4 – Announcement of identified teams**

***Deadline: April 1, 2018***

The organizers will identify teams based on the inputs collected from the materials and data scientists (in Steps 1 and 3 above). Each teach team will consist of ($\leq ) $2 materials experts and ($\leq ) $2 data scientists.

**Support and Sponsors**

Full financial support is available for participant travel, accommodation, and all meals.

Supported by the National Science Foundation under Grant No. DMR-1748198. Jointly supported by the Division of Materials Research (DMR), Division of Information and Intelligent Systems (IIS), and Division of Mathematical Sciences (DMS).

**Example Topics**

*Materials Science*

General:

• Data-driven design of experiment and simulation

• Inverse data-driven materials design

• Machine learning of quantitative structure activity relationship (QSAR) models

• Predicting the properties of materials

• Identifying descriptors of materials performance

• Identifying patterns in experimental data (e.g. micrographs).

• Dimensionality reduction, exploration, and exploitation of high-dimensional data sets

Specific:

• Discovery and design of sequence-defined cell-penetrating peptides and polymers

• Composition formulation of designer alloys

• Optimal design of substrate patterning for polymeric assembly

• Design of interaction potentials for self-assembling colloidal crystals

• Accelerated discovery of organic semiconductor materials

• Enhanced sampling in molecular simulation

• Materials discovery in large-scale databases

*Data Science*

• Bayesian data analysis

• Creation of databases

• Data integration

• Data reduction techniques

• Feature selection

• High Performance techniques

• Machine learning

• Out-of-core algorithms

• Spatial statistics

• Text mining

• Uncertainty quantification

**MATDAT18: Materials and Data Science Hackathon**

**DATA SCIENCE TEAM APPLICATION FORM**

**Complete and return via email to** brian\_reich@ncsu.edu **by 16 March 2018**

**Team Composition** (2 people max.)

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Department** | **Institution** | **Email** |
|  |  |  |  |
|  |  |  |  |

**Materials Project Title** (10 words max)

**Materials Project Synopsis** (2-3 sentences)

**A summary of the proposed initial data science solution** (2 to 3 sentences)

**A brief summary of the proposers’ qualifications**