|  |  |
| --- | --- |
| **UW DIRECT Capstone Project Proposal**  The project should allow trainees to cement the acquisition of data science skills and develop proficiency in the conduct of team-based interdisciplinary research | |
| **PROJECT NAME:** Application of molecular data science to study phthalate plasticizers | |
| **SPONSOR NAME:** Cargill, Inc.  **May we list you on our website as a partner DIRECT project partner?** Yes  **Will graduate students be asked to sign a non-disclosure agreement?** No | |
| **PROJECT DESCRIPTION:** This team will work with a researcher from Cargill and Professor Jim Pfaendtner to study molecular scale descriptors of plasticizers, in particular those derived from phthalate chemistries. We will perform a descriptor-based analysis of known plasticizers searching for key features common across a database of plasticizers. Following that we will compare against public repositories such as the GDB-17 database.  **DESCRIPTION OF DATA TO BE USED:** UW has a subscription to multiple electronic plasticizer database/handbooks. The GDB-17 database and related databases are publicly available. | |
| **PROJECT START DATE:** 3/30/20 | **PROJECT END DATE:** no later than 6/19/20 |
| **PROBLEM TO SOLVE/OBJECTIVE:**  There is a great need to develop alternatives to phthalate chemistries for uses such as plasticizers in commercial polymers. Cargill, inc seeks to understand how ML techniques can be used across classes of common molecules to assist in the development of common descriptors or discovery of common features across sets of molecules. This information can be in turn used to assist in identifying potential new alternate molecules. | **TIMELINES AND DELIVERABLES:** A proposed work plan is:   1. Data scavenging and wrangling from one or more plasticizer databases present in the UW libraries (2-3 weeks) 2. Re-purposing existing code from prior Pfaendtner group projects to calculate descriptors and perform clustering or other unsupervised learning techniques (3-4 weeks) 3. Calculation of parameter sets across for large public molecule databases and rank-ordering similar molecules (2-3 weeks) 4. Other tasks as decided on with project mentors in pursuit of the problem solution |
| **PROJECT MENTOR(S):**  Tim Abraham [tim\_abraham@cargill.com](mailto:tim_abraham@cargill.com) – plasticizer chemistry  Todd Kurth [todd\_kurth@cargill.com](mailto:todd_kurth@cargill.com) – plasticizer chemistry  Giota Kyriakou [Panagiota\_kyriakou@cargill.com](mailto:Panagiota_kyriakou@cargill.com) – data scientist (first line of communication if not obvious)  Spencer Schaber [spencer\_schaber@cargill.com](mailto:spencer_schaber@cargill.com) – data scientist  Alper Yarasik [alper\_yarasik@cargill.com](mailto:alper_yarasik@cargill.com) – business / applications specialist  **UW FACULTY CO-ADVISOR:** Jim Pfaendtner  **PROJECT TEAM MEMBERS:** | |