**Data Science Working Group Call For Proposals (Spring 2020)**

Are you passionate about Data Science? Would you like to share your passion and knowledge with others? The Drexel Data Science Club is now accepting proposals for working groups in all fields related to data science. Students can propose topics, projects and workshops focused on field of their choice. Students can submit proposals as a group or individually and projects can be of any level of complexity.

The working groups would meet weekly and try to introduce students with new concepts and methods. Students should submit a plan for their working group including a weekly list of topics that they plan on covering, programming languages/ tools they are going to use, proficiency level required (beginner, intermediate, expert) and their resume/projects as credentials for authority on given topics. Please see past working group’s proposal on the next page as an example.

**Text Processing Working Group Plan (Winter 2020)**

Would you like to learn how to make a computer understand English? Are you starting out in Natural Language Processing and need teams to work with? The Drexel Data Science Club’s (DSC’s) Text Processing Working Group (TPWG) features interactive demos, projects, tutorials and discussions about text processing. We will start from the basics: TF-IDF, cosine similarity scores, regular expressions, topical modeling, Stemming, Tokenization, Lemmatization and likewise explore more advanced topics.

Throughout the term, the TPWG will take up a TP project as well, which, in addition to ad hoc topics brought by the WG’s members will be approached as an open-ended skill-building exercise.

Project Description:

We will scrape ratemyprofessors.com for all the reviews that students have posted for professors and attempt to summarize ratings for professors. We will, e.g., use TF-IDF scores to get most informative words, create word vectors and learn to generate insights from our text data.

Coding Language: Python

Meeting schedule:

|  |  |
| --- | --- |
| Tuesday, 1/07/2020  4:00 – 6:00 PM  3675 Market St., Room 928 | Introduction to Natural Language Processing.  Basic Regex. Extracting patterns from text. Introduction to TF-IDF scores. |
| Tuesday, 1/14/2020  4:00 – 6:00 PM  3675 Market St., Room 928 | Motivation for representing text as vectors. Introduction to TF-IDF vectors. |
| Tuesday, 1/21/2020  4:00 – 6:00 PM  3675 Market St., Room 928 | Stemming and Lemmatization. Word embeddings (Word2Vec and others). |
| Tuesday, 1/28/2020  4:00 – 6:00 PM  3675 Market St., Room 928 | Application of all previous concepts on ratemyprofessors.com data. |
| Tuesday, 2/04/2020  4:00 – 6:00 PM  3675 Market St., Room 928 | Free form application of concepts learned on new dataset. |