Bing O'Dowd

Email : odowd.bing@gmail.com

github.com/bodowd & gitlab.com/bodowd

linkedin.com/in/bing-o-dowd

Summary

I am a results focused data scientist with a strong background in chemistry for biological applications and machine learning. Demonstrable experience delivering impact directly in projects, driving change, and developing collaborators to lift the organization to higher levels. Adept at extracting value out of large or small data, machine learning research, wet-lab chemistry & biochemistry. Motivated by delivering value and not by adding popular tools to my resume. An enthusiastic teammate, effective communicator, and creative thinker.

Experience

- **Bayer AG** — Monheim am Rhein, DE
  Data Scientist — *October 2020 - Present*
  - **Built data processing pipeline in order to perform first ever combined analysis**: Wrote Python programs to intake, extract, clean, and aggregate disparate excel sheets containing expensive experimental results from third party contractors. Compiled the data into a single data set for analysis and used PCA to discover insights to direct decisions of the following expensive, low-throughput experiments.
  - **Advanced projects with project-specific data analysis and models**: Built Random Forest model and backtested against our historical data to find a model that classifies at 70% accuracy and 93% recall compared to the baseline of manual selection of compounds which yielded 30% accuracy
  - **Implemented Conformal Predictors from literature and used for active learning in projects**: Used Conformal Predictors on a Random Forest model to quantify uncertainty, and used the uncertainty to implement selections of next experiments by active learning principles
  - **Implemented models from literature**: Developed internal tools for Conformal Predictors and Random Matrix Discriminant
  - **Continued developing and maintaining generative deep learning model in active chemical discovery projects**: In a team of two, continued maintaining and developing internal software for generative and point prediction deep learning on molecules. Initiated tutorials to make it easier to onboard new internal data science users to the software.
- **Developer**
  Online job board side project – chirole.com — *2019*
  - **Built online job board for computational jobs in life science companies**: Used Python/Flask for backend, Bootstrap, CSS, and HTML for front end, Postgres and SQLalchemy for the database, and deployed and hosted the app on Linode
  - **Integrated payments with Stripe**: Build feature so that customers could create job posting which would be automatically posted after payment
  - **Marketing on reddit, Twitter, and HackerNews**: Made posts to share on various subreddits, and HackerNews. Used Twitter to also broadcast jobs to reach a wider audience
- **Bayer AG** — Monheim am Rhein, DE
  PostDoc — *October 2018 - October 2020*
  - **Researched generative deep learning model**: Researched and published a variational autoencoder model for modeling molecular properties and generating chemical structures (Penalized Variational Autoencoder for Molecular Design https://chemrxiv.org/articles/preprint/7977131/2)
  - **Developed and applied generative deep learning model in active chemical discovery projects**: Jointly developed and maintained a variational autoencoder in Python and PyTorch in a 3-person team, and also applied the model in active chemical discovery projects
  - **Used generative model to generate novel molecule ideas yielding in a patent application**: Generated novel molecules and assisted in selecting ideas for synthesis and testing using the generative deep learning model, was found to have interesting properties, and lead to a patent application – "Novel heteroaryl-substituted pyrazine derivates as pesticides" (pending)
  - **Led the use of machine learning to uncover insights in a large business-critical project**: Identified new insights in long-standing project by using unsupervised machine learning. Implemented analysis and model in R, presented findings to stakeholders, and used the model to prioritize candidate molecules to synthesize and test

- **Dow AgroSciences** Champaign, IL, USA
  Data Scientist Intern, Cheminformatics *March - August 2017*
  - **Data pipeline development**: Refactored and further developed data cleaning pipeline to speed up data ingestion process, from annual and manual updates to weekly and automated updates
  - **Sped up 3D similarity searches**: Automated and sped up (from ~1 week to < 4 hours) 3D molecular similarity searches by building one interface to three separate programs, and by parallelizing the computation across multiple processors
- **University of Illinois at Urbana-Champaign** Champaign, IL, USA
  Research Assistant *August 2013 - May 2018*
  - **Introduced Machine Learning for chemical research to the lab and used ML to direct synthetic efforts**: Ruled out 31% of possible molecules to synthesize using PCA and a generalized linear model to prioritize chemistry efforts and reduce time (ca. 6mo-1yr) and financial costs
  - **Published work**: Discovered new a class of bacterial enzyme as an antimicrobial drug target, characterized novel inhibitor activity, and demonstrated previously unknown mechanism of action of bisphosphonates against cancer cell lines

Education

- **University of Illinois at Urbana-Champaign** Champaign, IL, USA
  PhD Chemistry *2013 – 2018*
- **University of California, San Diego** La Jolla, CA, USA
  B.S. Chemistry *2008 – 2012*

Skills

• Python, RDKit, Pandas, NumPy, Scikit-learn, PyTorch, SQL, Unix, git, statistics, machine learning, deep learning, organic chemistry, biochemistry, drug discovery, cheminformatics