Hao Lin Ph.D.

haolin.work@gmail.com (226) 972-9951 linkedin.com/in/hao-lin-cs/ github.com/atuSpirit

TECHNICAL SKILLS

- Languages: Python, Java, SQL (SQLite, MySQL), R, Perl, knowledge of HTML and XML
- Machine Learning techniques: Linear regression, Logistic Regression, Decision Tree, Random Forest, SVM, Neural Networks, Clustering, NLP, feature engineering, knowledge of deep learning (CNN, RNN)
- Tools: Pandas, Numpy, Matplotlib, scikit-learn, BeautifulSoup, Git, Agile and Scrum (Certified Scrum Master)
- Skills: Data analytics, data structure, algorithm design and analysis, software quality assurance

EXPERIENCE

Insight Data Science Toronto, Ontario

Data Science Fellow May, 2019-Present

- Developed a web app that leverages machine learning and NLP techniques to predict escalating customer complaints and suggest actionable responses. The app, Complaint Assistant, is developed with Flask and deployed on Amazon Web Services
- Multiclass-classification of complaints into product categories based on complaint narratives embedded by TF-IDF
- Combined TF-IDF embedding features and sentiment analysis metrics to classify an imbalanced data set with 16K narratives

University of Waterloo, Cheriton School of Computer Science Waterloo, Ontario

Research Associate, Computational Biology June 2018 - Present

- Designed pipelines and built software using Java for the challenging polyclonal antibody sequence assembly problem, which could deliver significant advances in the drug design industry

Bioinformatics Solutions Inc. Waterloo, Ontario

Software Quality Assurance Manager Nov 2014 - June 2016

- Led the software quality control team of three staff through three full development cycles of commercial software PEAKS in fast paced Agile development process; Gained positive feedback on software quality from customers with new releases
- Communicated effectively with the owner and developers to improve designs, features, and algorithms to meet customer's requirement; Hired, trained and mentored QA team members
- Initiated and implemented an automatic data-driven regression test framework using Java, which played important role in identifying software errors in each development cycle

Bioinformatician Feb 2013 - June 2016

- Created and implemented data analysis pipelines to validate software results and hunt improvements for quality and performance using Python and Perl scripts
- Carried out statistical analysis using R to collaboratively improve PEAKS 7.5 label free quantification module; Ranked #1 among world-wide research groups and companies in iPRG 2015 study
- Used linear regression to model the feature quality score and the feature pair significance score in proteomics data, which was a key factor to the success of a newly developed labeled protein quantification method

University of Waterloo, Cheriton School of Computer Science Waterloo, Ontario

Postdoctoral Fellow, *Computational Biology* July 2008 - Jan 2013

- Designed and developed efficient algorithms and pipelines to provide practical solutions to analyze large scale genomics data and proteomics data in three cutting-edge research projects
- Collaboratively developed and commercialized software ZOOM and ZOOM Lite(version with GUI) for mapping huge scale DNA sequencing data, one of the earliest efficient software for the emerging Next Generation Sequencing Technology
- Proposed a novel combinatorial model for LC-MS retention time alignment and implemented it in PEAKS 7, which served as the base of the label free quantification module in PEAKS 7

Sectional Instructor, Introduction to Computer Science 2 May 2013 - Aug 2013

- Taught data structures and algorithms using Scheme and Python to 110 undergraduate students
- Designed assignments, exams and marking schemes; Managed three academic support staff for tutorials and marking

EDUCATION

PhD in Computer Science and Technology March 2008

Chinese Academy of Sciences, Institute of Computing Technology Beijing, China

BS in Computer Science and Technology July 2002

Sichuan University Chengdu, China