Ding CHENG

+86 17521619529 | mkbo@me.com | github.com/dipcarbon

EDUCATION

Fudan University — Shanghai, Part-time
Master Candidate of Software Engineering, Data Science and Application — *2021 – Present*

Fudan University — Shanghai, Full-time
Bachelor degree of Chemistry, Minor in Data Science — *2014 – 2018*

EXPERIENCE

Senior Statistical Programmer — Apr 2021 – Present
Global Statistics and Data Science — *BeiGene*
- Use SQL and Python to extract data analysis from databases such as Symphony and Flatrion, build models to support patient care, estimate market share, and clinical trials study set-up
- Utilize programming expertise in clinical trial and other data science with business knowledge, and collaborate with MA biostatisticians, to identify, define and solve the problems and to provide strategic input and safety signal identification
- Use Shiny to build an automated statistical analysis platform for biomarker identification and recommendation

Data Engineer/Clinical Programmer — Feb 2018 – Apr 2021
PD Data Sciences — *Roche*
- Work with stakeholders to assist with data-related technical issues and explore ways of automation and insights generation
- Develop SAS programs for extracting and processing data with SDTM implementation and CDISC standards and lead the data collection of clinical database development with clinical data EDC programming

Research Assistant — Jan 2021 – Present
Biomedical Informatics Laboratory(Part-time) — *Fudan University & Zhongshan Hospital*
- Construct deep learning models to train classification models of liver cancer patients using convolutional neural networks and NLP
- Implement OHDSI OMOP common data model(SNOMED CT, MedDRA) and build up a federated analysis platform using R packages
- Improve EHR Data usability and quality based on natural language processing and machine learning(Bi-LSTM)

RELATED PROJECTS

Data Curation and Integrated Analysis | *R Shiny, SQL, Python, REST API*
- Implemented data curation and integration from Clinical Trial data of EDC, Omics Data and RWD within harmonized standards and control terminology for secondary use
- Supported continuous delivery using Jenkin and Docker and developed R packages(*CurateR, Scrubj*) and a full-stack R Shiny app(*Curation Workshop*)
- Built exploratory and analytic tools that utilize the data pipeline to provide key metrics and clinical outputs(TLGs) with R frameworks

Crovalimab Phase I/II Trials Study | *SAS, R, Pinnacle 21, Spotfire*
- Ensure timely and quality development and validation of datasets and reports for study-documents
- Designed and reviewed the eCRF and data structures and comply with company and industry standards (CDISC SDTM) and processes, created the review of programming specifications

PUBLICATION

First Author: Ding Cheng, Liang Qiao*, Peter Horvatovich*, *Towards Spectral Library-free MALDI-TOF MS Bacterial Identification*, J. Proteome Res., 17 (2018), 2124–2130

TECHNICAL SKILLS

Languages: English(TOFEL 108), Mandarin
Programming Languages: Python, R, SQL, C#, SAS, JavaScript
Tools: Tableau(Certificated), Power BI, TIBCO Spotfire, Microsoft Office, Final Cut Pro
Libraries and Frameworks: pandas, Flask, NumPy, Matplotlib, Shiny, Tidyverse, Roxygen