Ding CHENG

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

Experience

Senior Statistical Data Analyst Apr 2021 – Present

Global Statistics and Data Science *BeiGene*

- Utilize analysis expertise in clinical trial and other data science with business knowledge, and collaborate with biostatisticians, to identify, define and solve the problems and to provide strategic input
- Use SQL and Python to extract data analysis from databases such as Symphony and Flatiron, build statistical models to support patient recruitment, estimate market share, and clinical trials study set-up
- Use R/Shiny to build an automated statistical analysis platform for biomarker identification and recommendation

Clinical Data Engineer Feb 2018 – Apr 2021

PD Data Sciences *Roche*

- Implemented data curation and integration from Clinical Trial data of EDC, Omics Data and RWD within harmonized standards and control terminology for secondary use
- Developed SQL/SAS programs for extracting and processing data with CDISC standards and lead the data collection of clinical database development with clinical data EDC programming
- Assisted stakeholders to work with data-related technical issues and deliver automation solution and insights generation

Research Assistant May 2017 – Sep 2017

Groningen Research Institute of Biomedical *Groningen, The Netherland*

- Built a random forest model using genomics and proteomics data to train a classification model for CODD patients
- Conducted data visualization, including volcano plots, violin plots and interactive user interface and realized the data upload and analysis

Related Projects

Safety and Real World Data Anlysis | *R, SQL, Python, Power BI*

- Built a safety signal identification dashboard to monitor adverse events in patients from post-hoc and marketing using statistical models
- Worked with Clinical Operations to estimate the number of patients meeting the admission criteria and their geographic distribution in United States
- Worked with the Market and Strategy Department, insurance data is used to speculate on the projected market size and patient distribution of the relevant treatment options

IR prediciton model with machine learning | *Python, Scikit Learn, Streamlit, Matlab*

- Evaluated the effects of SVM, PLS, ANN, GP and other machine learning models in infrared concentration regression
- Verified the Gaussian process approach through spectroscopic datasets and worked with the manufacture unit to deploy predictive models into the cell culture monitoring process

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*

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, Mandarin
Programming Languages: Python, R, SQL, C#, SAS, JavaScript
Tools: Tableau(Certificated), Power BI, TIBCO Spotfire, Microsoft Office
Libraries and Frameworks: Pandas, Scikit Learn, NumPy, Shiny, Tidyverse, Roxygen