



Data Scientist

Location: Frederick, Maryland

Term: Full time, Permanent

Industry: Consulting – Drugs/Biologics/Devices

Reports to: Director, Biostatistics

BioStat Solutions, Inc. (BSSI) is a Health IT corporation in service to the pharmaceutical, biotechnology, medical device and biodefense industries. BSSI provides niche advanced analytics services to the healthcare sector that include the development and implementation of strategies for biomarker and –omic studies, biodefense countermeasure development, and medical device studies.

Responsibilities include, but are not limited to:

- Develop, automate and deploy algorithms based on novel advances in machine learning, predictive analytics and optimization. Must to be able to interpret ML models on biology data.
- Familiar with multiple Therapeutic Areas (TA), and able to quickly learn the necessary biology knowledge of new TA. Having research experience on gastrointestinal diseases is a big plus.
- Develop high quality programs in R or other languages (e.g. Python) for use in data management, statistical and bioinformatics analysis, and visualization of biomarker and –omic studies.
- Perform statistical analysis of molecular association across various therapeutic areas but with a focus on gastrointestinal disease and assess statistical properties of molecular variation (e.g. RNA-seq counts, single nucleotide polymorphisms, protein expression).
- Develop and execute data processing, quality control and analysis pipelines for biomarker data.
- Assist in design, analysis and interpretation (including visualization) of whole transcriptome sequencing studies and/or whole genome sequencing studies and high throughput proteomic studies.
- Develop algorithms and tools to process, analyze and integrate large data sets with millions or billions of rows.
- Develop analysis plans and reports, data and programming specifications, and documentation of reusable functions and packages.
- Perform analytical programming validation and application testing.
- Develop new web applications used for client scientists to analyze –omic and clinical datasets.
- Interact and collaborate with colleagues and clients to provide programming support in a fast-paced team environment and building a data-driven product pipeline.
- Keep abreast of new state-of-the-art software technologies and best-practices.

Qualifications:

- Ph.D. (Preferred degree in Computational Biology, Bioinformatics, Data Science, Biostatistics, or related field) or M.S. degree with 4-6 years' relevant work experience.
- Proficiency in R (Preferred experience in building R/Python packages with foreign language interfaces).
- Experience with gastrointestinal, neurological, immunological, or inflammatory diseases is preferred.
- 2+ years of experience with data visualization tools.
- Experience with machine learning and deep learning packages (i.e. caret, mlr, TensorFlow, Keras, etc.).
- Ability to work efficiently in a Unix/Linux environment.
- Team player with excellent communication and interpersonal skills.
- Experience with analytically processing biomarker data such as high-throughput or targeted gene expression assays (e.g. NanoString, RNA-Seq), whole genome sequencing, , or other specialty lab data (e.g., bacTRAP, immunoassays, IHC, qPCR, TCR sequencing, etc.).
- Experience with -omic data, big data platforms, and / or public databases (e.g. 1000 Genomes, TCGA, dbGaP).
- Experience in complex matrix algebra and high dimensional predictive models (thousands+ of features) is preferred.
- Experience with linear/nonlinear mixed model tools (nlme, lme4, NONMEM, etc.) is preferred.
- Experience with relational databases such as MySQL is preferred.

BioStat Solutions, Inc. is a voluntary equal opportunity employer. In compliance with federal law, all persons hired will be required to verify identity and eligibility to work in the United States and to complete the required employment eligibility verification form upon hire

Please apply to hr@biostatsolutions.com

www.biostatsolutions.com