Job Title: Data Scientist – Research Informatics

Requisition #: 9555

POSITION SUMMARY & KEY RESPONSIBILITIES:

The Data Scientist will be inquisitive in his/her approach to figuring out the right biomedical data set that needs to be mined from multiple clinical and clinical research systems at City of Hope (COH). Resource should be comfortable in asking questions about the data source, figure out the business rules around how it was put together, and look forward to aid scientists in their biomedical data identification and retrieval needs:

Will use data mining, statistics, statistical modeling and machine learning to understand relationships and patterns in data. The insights gathered from this analysis are translated into actionable steps and best practices

Responsible for developing tools to enable visualization to translate analytics into information that can be used by clinical and operational leadership

Will take a data driven approach to decision making in projects.

The Data Scientist with Clinical Research Focus position requires exceptional communication skills, a collaborative approach and attitude of service with a sense of urgency. The candidate must be self-driven and eager to learn new methods of data munging, data visualization and data analysis.

Main Responsibilities Includes:

Integrate multiple systems and perform analytics on large data sets

Perform explanatory data analyses, generate and test working hypotheses, prepare and analyze historical data, identify patterns and correlations among various data points and discover new insights

Perform machine learning, natural language, and statistical analysis methods

Meet with researchers to explain data, data analysis, insights and relationships within data

Analyze and interpret research data, extract scientific relevance and formulate

Produce reports from research data to meet end-user needs

Collaborate and develop working relationships with clinical informatics, bioinformatics and other computational groups at City of Hope

Collaborate with other database and software developers.

Minimum Education and Experience:

Masters degree in statistics, mathematics, informatics, machine learning, computer science, epidemiology or a related field

5-8 years of experience with data analytics

3+ years of applicable experience developing and successfully implementing statistical models for data mining, predictive risk modeling, clustering, and classification

2+ years of experience with decision trees

Experience in using analytics to generate and validate insights from large data sets

Experience with relational databases and SQL queries

Other Experience or Certification (Preferred):

Ph.D. degree preferred

Experience in a health care environment

Experience in health plans/payer environment

EHR experience

3+ years of applicable experience developing and successfully implementing statistical models for data mining, predictive (risk) modeling, clustering, and classification

2+ years of experience with decision trees

Skills and Abilities:

A doctorate degree in relevant discipline (e.g., medicine, basic sciences, clinical sciences, Health Sciences, applied statistics, data mining, machine learning, etc.)

Deep familiarity with clinical research (clinical trials), clinical or healthcare domain is required

Five years of experience in working with biomedical datasets in cancer research environment

A good understanding of statistical and predictive modeling concepts, machine-learning approaches, clustering and classification techniques and stratification analysis.

Proficiency in SQL, MySQL and/or PostgreSQL.

Expertise in one of the programing languages

Demonstrated understanding of machine learning and predictive analytics techniques

Ability to mine large datasets and analyze large, complex, multi-dimensional datasets

Must be knowledgeable in at least one statistical and analytic packages such as R, MATLAB, SAS, SPSS or Weka

Proficiency in at least one scripting language (e.g. Python, Perl)

Ability to exercise independence and use creative approaches to problem-solving.

Experience in data visualization of biomedical datasets.

Comfortable communicating complex technical subjects to non-technical audiences using practical examples, and prototypes