DUKE UNIVERSITY ENERGY INITIATIVE

Position Title: Data Scientist

Occupational Summary

The Duke University Energy Initiative (energy.duke.edu) is a university-wide interdisciplinary collaboration focused on advancing an accessible, affordable, reliable, and clean energy system. The Initiative reaches across business, engineering, environment, law, policy, and the arts and sciences to educate tomorrow’s energy innovators, develop new solutions through research, and improve energy decisions by engaging business and government leaders.

As a part of the Energy Initiative, the Energy Data Analytics Lab is developing and applying advanced data analytics tools to transform diverse energy data into insights that lead to energy system performance improvements. Improvements include increased reliability, resiliency, environmental sustainability, and productivity, along with reduced costs. The Lab is a hub of research, education, and engagement activity with thought leaders in the business and the policy world around energy data analytics, and will engage with a broad spectrum of experts and stakeholders to provide energy data-focused solutions to challenging problems in the energy space. At the highest level, the types of research questions this position investigates are divided into three main categories: (1) describing and assessing the status quo of energy systems and markets; (2) making predictions and forecasts for energy resources, markets, and human behavior and (3) developing prescriptive pathways for the future evolution of modern energy systems.

This position will assist the Director, staff, and other faculty members on a range of data-intensive research projects and questions concerning energy and its interaction with the economy, environment and security.

Work Performed

Produce data analyses of complex energy data sets by applying computational tools and machine learning algorithms for analyzing data, producing written reports including journal publications, oral presentations including conference presentations, and web-based media such as interactive data visualizations. Additionally, the this position will assist with the analysis of unique data sets ranging from satellite imagery and remote sensing data, macro-economic energy market data and building-level smart meter data.

Work on signature projects including investigating how remotely sensed data (satellite imagery and lights at night data) can be used to provide information on energy systems infrastructure, characteristics, and resources for public policy and decision makers, as well as other researchers. This position will help develop and apply techniques for extracting information about the location and characteristics of energy infrastructure (solar photovoltaics, buildings and their associated energy consumption, transmission lines, power plants, regions of the world with access to electricity, etc.) automatically from remote sensing data using deep machine learning techniques.

Participate in additional research projects as needed such as investigating the integration of intermittent resources into the grid using statistical modeling and data analysis techniques; automated building energy
assessment through non-intrusive load monitoring, a machine learning based data analysis technique; as well as modeling data from project partners related to specific energy technologies, such as hot water heaters, to investigate how to increase their energy efficiency through data-driven analysis and modeling.

Assist with the broader activities of the Duke University Energy Initiative, as needed, including developing public-facing blog posts with data visualizations; providing expert advice regarding online data management and presentation; organizing energy data analytics research workshops; assisting with the development and implementation of energy data educational modules for classroom and laboratory use; student mentoring; and other duties as assigned.

Maintain and manage access to virtual machines and curate relevant data sets for sharing with researches both inside and outside of Duke University through managing data repositories.

Work with faculty and stakeholders to plan and conduct meetings, workshops, conferences, and events.

Perform other related duties incidental to the work described herein.

**Required Skills/Knowledge/Abilities**

- Proficiency in Python programming, preferably with experience using Numpy and Matplotlib.
- Conceptual understanding of machine learning (particularly predictive modeling) including linear and nonlinear regression; classification; supervised and unsupervised learning.
- Data processing skills including cleaning (handling missing or corrupt data), analysis (using statistical methods), and interpretation.
- Sound understanding of probability and statistical methods and a comfort with differential and integral calculus.
- An independent work ethic, requiring minimal daily supervision
- Ability to carry out multiple projects simultaneously.
- Exceptional organizational and project management skills, with attention to details.
- Knowledge of or willingness to learn about energy systems including electricity generation, transmission, distribution, and power market operation and technologies; transportation systems and fuels; and consumer energy efficiency decision-making and strategies.

**Desired Skills/Knowledge/Abilities**

- Deep learning (subfield of machine learning) expertise, or willingness to quickly learn Keras, Tensorflow, and running machine learning models on Graphics Processing Units (GPUs)
- Working knowledge of command line operations and the Linux operating system.
- Data visualization experience, including online and interactive data presentation such as D3.js, vega, or vega lite.
- Strong writing skills, as demonstrated by the ability to clearly describe statistical models and data-oriented research results.
- Polished verbal communication skills in presenting to both technical and non-technical audiences.

**Minimum Qualifications**
**Education/Training:** Work requires a master's degree in a relevant technical field including, but not limited to, engineering, computer science, statistics, economics, mathematics, or energy systems. Ph.D. preferred.

**Experience:** Academic or professional experience with advanced data analysis and computer programming required. One to two years of professional experience preferred.

**OR AN EQUIVALENT COMBINATION OF RELEVANT EDUCATION AND/OR EXPERIENCE**

The intent of this job description is to provide a representative of the types of duties and responsibilities that will be required of positions given this title. It shall not be construed as a declaration of all specific duties and responsibilities of any particular position, as employees may be directed to perform job-related tasks other than those specifically presented in this description.

**Application Process**

All applicants must apply through the Duke Human Resources system at http://www.hr.duke.edu/jobs/