Data Engineer

This position will be located within (Dev/IT) and work closely with computer scientists, IT and data scientists to deploy and optimize machine learning models in the production system environment.

Responsibilities:

- Work closely with IT/Computer Scientists on technical aspects of deploying machine learning models in production.
- Work closely with Data Scientists to understand, implement, refine, design, and test deployment of machine learning models in production.
- Optimize the environment for production machine learning models to access and handle data more efficiently and ensure scalability.
- Designs new processes and builds large, complex data sets needed for machine learning processes.
- Serve as SME on machine learning technology and recommend acquisition of appropriate technology for production purposes.
- Advise and assist IT/infrastructure on install and configuration of machine learning systems.
- Explore, design, and implement a robust production-grade data processing pipeline that can ingest, aggregate, and transform large datasets.
- Independently conduct literature search to keep informed of best practices and new methods.
- Serve as on-call for production issues related to machine learning processes.

Requirements:

- BS degree in Computer Science or related field with 5 years machine learning engineering experience or MS/PhD degree in Computer Science or related field with 3+ years of machine learning engineering experience.

- 3+ years hands-on experience deploying production-level machine learning algorithms and productionizing them at scale in a distributed computational environment.
- 1+ year experience with R. Working knowledge of R required.
- Experience working with large, messy real-world data.
- Experience with SQL, Ruby, Python, C#, Pig and other query and programming languages.

- Experience with machine learning database tools and platforms such as HBase, Mongo, Hive, Cassandra, MySQL, SQL Server, PostgreSQL, Hadoop, Spark.
- Experience with machine learning optimization tools and related technologies such as H2O, Theano, TensorFlow. Experience with H2O required.
- Experience with machine learning platforms for production models such as Apache, Pattern, Shogun.

- Strong expertise in computer science fundamentals: data structures, performance complexities, algorithms, and implications of computer architecture on software performance such as I/O and memory tuning.
- Working knowledge software engineering fundamentals: version control systems such as Git and Github, workflows, ability to write production-ready code.
- Strong knowledge of data architecture and system/pipeline and data processing engines such as Spark and Hadoop.

- Working knowledge of R and Rstudio.
- Working knowledge of SQL, Pig, Python, and other query languages.
- Knowledge of C++, PHP, Java and other languages.

- Knowledgeable with machine learning tools and frameworks like Python, Spark, H2O, Theano, mlpack, TensorFlow.
- Knowledge of machine learning platforms such as Amazon, IBM Watson, Azure, Google Predict, BigML

- Strong trouble-shooting skills.
- Knowledge of technical infrastructure.
- Knowledge of installation and configuration of machine learning systems/technology.
- Strong technical aptitude.
- Basic knowledge of statistics, calculus and probability, experimental design, and machine learning techniques to enable conceptual understanding of Data Scientist's models.
- Has strong critical thinking skills and the ability to relate them to the products of Paycom.
- Possesses a combination of creative abilities and business knowledge.

- Demonstrates excellent verbal and written communication skills as well as the ability to bridge the gap between data science and business management.
- Displays exceptional organizational skills and is detailed oriented