

PointPredictive Job Description for Staff Data Scientist Position

July 16, 2018

Job Duties

- Data analysis and validation at AWS in UNIX (raw field distributions for tagged and untagged transactions)
- Support other team members with tasks as required
- Participate in brain-storming sessions with other team members and domain experts to create enhanced derived input features
- Build test models at AWS using Python scikit-learn and/or R
- Use model performance results to create ROC curves to compare model performances between versions and for reports to clients and prospects
- Help write performance reports for finished models
- Validate production models at AWS
- Support sales and marketing teams as requested with analysis and ideas

Job Skills

- **Necessary:**
 - Reasonably conversant with Linux/Unix commands and operations (e.g. gzip, sort, cc, cat, tail, pushd, popd, dirs)
 - Able to use Python and scikit-learn to build regression models (logistic and/or linear) and others
- **Useful**
 - Experience with AWS to analyze data and build models
 - Experience with Python and R to perform data analysis, validation and model building
 - Ability to do data visualization using tools such as Tableau and others
 - Familiarity with ROC curves and confusion matrix for performance evaluation of models
 - Familiarity with building supervised models and classification models (e.g., transaction fraud detection models)
 - Experience with large, tagged data sets (1 million+ records)
 - Experienced with predictive input feature creation and variable reduction
 - Ability to brainstorm with other team members and domain experts to create useful derived input features from raw data fields
 - Experience with creating explanations for model results for users
 - Familiarity with using categorical variables (100+ values – e.g., ZIP codes, Merchant Category Codes) as continuous inputs to models based on tagged training data (e.g., risk tables)
 - Some experience with
 - clustering
 - decision trees
 - random forests
 - neural nets
 - SVM
 - Knowledge of how to deal with low rates of bads in tagged data (e.g., 1:1000)