Job Description

We are looking for a Data Scientist that will help us discover the information hidden in vast amounts of data & help us make smarter decisions to deliver even better products. Your primary focus will be in applying data mining techniques, doing statistical analysis & building high quality prediction systems integrated with our products. **Depending on needs** develop "automate scoring using machine learning techniques", "build recommendation systems", "improve and extend the features used by our existing classifier", "develop internal testing procedures", "build system for automated manufacturing efficiency prediction", etc.

Responsibilities

- ❖ Selecting features, building & optimizing classifiers using machine learning techniques
- ❖ Data mining using state-of-the-art methods
- ❖ Extending company's data with third party sources of information when needed
- ❖ Enhancing data collection procedures to include information that is relevant for building analytic systems
- ❖ Processing, cleansing, & verifying the integrity of data used for analysis
- ❖ Doing ad-hoc analysis & presenting results in a clear manner
- ❖ Creating automated anomaly detection systems & constant tracking of its performance

Desired Skills & Experience

3 to 6 years of proficient experience working & developing SDKs for any platform

Skilled/Expert Level development in:

- ❖ Excellent understanding of machine learning techniques and algorithms, such as k-NN, Naive Bayes, SVM, Decision Forests, etc.
- ❖ Experience with common data science toolkits, such as R, Weka, NumPy, MatLab, etc. D**epending on our specific project requirements** . Excellence in at least one of these NumPy or R is highly desirable.
- ❖ Experience with data visualisation tools, such as D3.js, GGplot, etc.
- ❖ Proficiency in using query languages such as SQL, Hive, Pig, NiFi **{{actual list depends on what you are currently using in your company}}**
- ❖ Experience with NoSQL DB's such as InFleuxDB, MongoDB, Cassandra, HBase
- ❖ Good applied statistics skills, such as distributions, statistical testing, regression, etc.
- ❖ Good scripting & programming skills PHP, Slim, SQL, Laravel
- ❖ Data-oriented personality
- ❖ **Hadoop, HDFS, NiFi,**
- ❖ **Any certification related to BigData is require**

Proficient in understanding the user requirement based on the data availability

Proficient in using a sample data to conduct a POC for the given problem statement

How to organize and "wrangle" large datasets so that you can get actionable insights from them. This may include finding innovative ways to combine fields of data that don't naturally mesh together.

How to collect data from all kinds of different sources, from web APIs, to internal databases encoded, SQL, Influx, etc,...

Education

MTech./ MS/Mechatronics / Computer or equivalent