**Complex Data Analysis Interface Designer & Developer (CODEX)**

Imagine the world before Excel. Working with large tables of numbers was painful, slow, and often done by writing custom computer scripts. This is the state of machine learning today: expert knowledge is required in both machine learning and software programming in order to just explore data from a data-driven perspective.

CODEX (the Complex Data Explorer) will permit users from novice to expert understanding in a fast, beautiful interface to load in their dataset and immediately begin building intuition about their data. What looks normal or odd? Are there troublesome values? Do these numbers usefully predict these other numbers, and if so how? CODEX, when complete, will be a transformative product that lets everyone "play" with their data using machine learning, reducing analysis that previously took weeks to a matter of hours. Complex data science concepts will suddenly become similar to making a slide in PowerPoint or a graph in Excel.

We are seeking a highly skilled, professional, experienced JavaScript front-end developer to implement this ambitious vision, help define and implement the user interface, and do so with the highest software engineering integrity and Dev-Ops technology. We’ve got the machine learning and Python back-end covered.

Some of CODEX’s key features to make it clear what we’re aiming for:

 - Each algorithm has an image of what your actual data would look like (subsample) if you used this approach on it, along with time and memory predictions if used on your full dataset.  
 - Can visually select the one that looks best W.R.T desired outcome, memory available and runtime.

 - Can select hyper parameters in the same way: visually. Shows you the ramifications of selecting each setting before you commit to long processing jobs.  
 - All sliders and choices made have “effect” graphs beside them. For example, if you are filtering on a threshold, have a histogram of how much data would be filtered above the slider to help guide.  
 - <?> icons abound, permitting ML-expert-based guidance on what each algorithm does, why you would use it, and its shortfalls including web links for additional reading.  
 - Linked graphs: if you highlight data on one graph, instantly see it on all other graphs to permit very fast data exploration (you’ve got to try this… it’s amazing!)  
 - Numerical time-series data is current focus, not images or text.  
 - Multiple sub-selections of the data can be saved, processed separately, or visually overlaid to discover relationships.  
 - Clustering, Dimensionality Reduction, Classification, Regression, Curve Fitting, etc.  
 - Browser-based so you can use it anywhere… can also load server on your local machine so you don’t even need the net.  
 - Saves out what you’ve done so far in Python or Matlab code, including graphs, so you can pick up in your favorite language after your CODEX session  
 - Initial “bad value” scanning to discover NaN’s, Inf’s, outliers, repeating (likely bad) values in data before you start work… discover how they relate to each other (bad feature? bad rows/samples? Neither?)  
 - Correlation Matrices between features  
 - Estimation of “modality” of input features. Unimodal? Bimodal? Trimodal? This is vital when it differs from what you expect.

Requirements:

* Demonstrated expertise in visualizing complex data (e.g. science, engineering, statistical, medical, or map-based data)
* Demonstrated expertise with complex, large-scale JavaScript applications (Tornado, D3, React, Jest, NodeJS, Charts, SockJS)
* Experience with optimizing, performance tuning, and profiling asynchronous web data transmission with remote server computational engine
* Experience with graphing technical data (scatter plots, heat maps, histograms)
* Production-quality software engineering skills (unit testing, end-to-end validation, continuous integration via Jenkins, advanced Git usage)

Desired:

* Experience with research environment (moving targets, evaluating experimental results, building new solutions, deep problem solving)
* Knowledge of Machine Learning (classification, regression, feature selection)
* Statistics & Linear Algebra
* Python