

Electrics Company: Applying Machine Learning Tools to Highlight Anomalous Data



The Challenge

Over the last two years, we have been working with a cutting edge electronics company to build advanced algorithms using R and Python. Since they are at the research & design stage of the development process, their data structure is unique and challenging.

Each week they generate thousands of data sets that have a spatial, temporal and experimental component. Their data structure is unique - around 10,000 voltage curves distributed on a disk. This necessitates the use of novel machine learning algorithms, coupled with classical experimental design.

In theory, each disk should be identical. In practice, this isn't the case. The goal of the project was to develop a set of algorithms and R packages that would identify outliers and deviations from the canonical data. The ultimate goal was to optimise product yield. A novel challenge to this was that any issues detected were then fixed. In fact, our training data set consisted of issues that were resolved as they improved their processes.

The Solution

Data from their systems was obtained via APIs. We developed R software to automatically assess the data quality and remove any rogue readings. We then created machine learning tools to automatically highlight any potential failures and outliers. Finally, we proposed new experimental formulations.

Our Results

This was a fast-moving project, where the data type was changing and the data sources were being updated. We developed adaptable software that the client is still using. As the client only has a small data science team, we provided on-site training and regular assistance after the project has finished.

