Define Data Performance Metrics for RCASE algorithm

Background

Warwick Analytics provides ground breaking software that identifies root causes of faults and inefficiencies in manufacturing industries. The main algorithm (RCASE) was developed from over 10 years academic research and spun out of Warwick University.

RCASE requires a failure marker (usually from a warranty claim or test case) and some data from the life cycle of the product – this could include tolerance data, production data, testing data and user data. It does not require hypothesis and provides results even with dirty/incomplete data sets.

In 2013, Warwick Analytics won Demo God at Demo Fall and SAP’s worldwide most innovative start up as well as raising initial investment from Jensons Solutions.

Project

Data provided to RCASE is never clean by definition as not all failure markers will be identified (e.g. products yet to fail, customers not reported fault). The result of an RCASE calculation is a region defined by variables with their upper and lower limits. Within this region, there will be a high concentration of failure markers but almost certainly “normals” (those not identified as failed) that will probably fail.

WA needs a set of rules that, based on the characteristics of the data, define how much data is required to get a result. Currently, WA advises customers to give everything!

By creating sample datasets, it should be possible to understand the characteristics of datasets that will lead to strong regions. From this, a way of analysing a dataset to give a confidence metric of the likelihood of finding a results. RCASE deals with a multi-dimensional search space and so this is a non-trivial problem.

Deliverables

WA will provide access to the RCASE implementations, example datasets, dataset generators and the algorithm developers.

The deliverables are expected to be

1. Data analysis tool that scans data to give a confidence metric for RCASE