

Background

A complex modern semi-conductor manufacturing process is normally under consistent surveillance via the monitoring of signals/variables collected from sensors and/or process measurement points. However, not all of these signals are equally valuable in a specific monitoring system. The measured signals contain a combination of **useful information**, **irrelevant information** as well as **noise**. It is often the case that useful information is buried in the latter two. Engineers typically have a much larger number of signals than are actually required. If we consider each type of signal as a feature, then **feature selection** may be applied to identify the most relevant signals. The Process Engineers may then use these signals to predict yield excursions downstream in the process. This will enable an increase in process throughput and decreased time to learning, and reduce the per unit production costs.

The dataset represents a set of such features where each data record represents a single production entity with associated **measured features** and the labels represent a simple **pass/fail yield** for in house line testing and associated date time stamp, where -1 corresponds to a pass and 1 corresponds to a fail and the data time stamp is for that specific test point.

Datasets

➤ **Two datasets**

secom.data: 1567 observations, 591 variables (features)

secom_labels.data: classification (pass/fail) and time stamp

Requirements

- You are asked to implement the idea given in Slide 1 using, but not limited to, materials covered in the class.
- You may read additional literature and learn new methods/R codes to solve the problem.
- **Teaming:** Each team should include two students.
- **Submission:** Each team should submit a set of Power Point slides (**pdf** format) as if you were to give a presentation (no actual presentation).
 - First, title page: title of project and team members (UINs)
 - Then: strategy, methods, justification, results, conclusions
 - Finally: insights gained from doing the project
 - Page limit = **12** slides (including title page); extra slides will be ignored.
 - NOT include any codes.

Grading

- Totally 10 points
- Based on quality of data analysis and quality of slides
- Mandatory point deductions:
 - **Slides format:** should be in pdf format. If not, 1 point will be deducted.
 - **Late submission:** penalized by 0.5 point per hour it is late

Deadlines

- **Team formation: November 9 (Thu) @11:59pm**
 - Download the spreadsheet on eCampus, fill in and submit.
 - Each team member must submit a duplicate copy of your team formation sheet.

- **Power point slides: December 7 (Thu) @11:59pm**
 - Each team member must submit a duplicate copy of your slides.