Title:Semi-Conductor Manufacturing Process

**Team:**

**Shivam Raj Solanki**

**Mayank Mishra**

**Shirish Pandagare**

**Divyank Garg**

**Randhir Nalage**

1. What is the problem being addressed here? Who would care for this solution? How important and valuable is this solution?

The problem being addressed is of a semi-conductor manufacturing process. We plan to build a predictive classification model for fault detection(yield failure) with the response being simple Pass or Fail during the manufacturing process using classification.

Semi-Conductors are very widely used and almost all OEMs and electronics companies would care for a problem like this.

The Solution will help predict a failure using the sensor parameters and various other inputs which will avoid failures and reduce downtime by ensuring proper maintenance on time if required.

1. What is the analytics problem solved here? What is the response (if any), what are the inputs?

The analytics problem being solved is Prediction of a binary response i.e. Classification. The inputs are in the form of sensor variables and parameters.

1. What data set will you use?

We will be using the SECOM dataset from the UCI website. It has 591 attributes and 1567 observations(rows) (<http://archive.ics.uci.edu/ml/datasets/secom>)

1. How do you plan to solve? List the chain of techniques you will employ?

We will be using all classification tools learnt in class. The steps are:

Imputation->Classification using all known techniques->Cross Validation/ANN/Clustering->AUC/ROC->Selecting best technique->Final Resulting Model.

1. What is the new technique you will learn and apply? What is the R library you will employ for this? Show that it can work for your kind of data.

We are proceeding with the data from a new tool with which we are planning to build our model is the SMOTE function from the DMwR (Data Mining with R) Library. We are also looking at C5.0 found under package C50 for the task. However, we may change our tool/model in the later stage depending on the outcome of the model. We are not very confident in the classification model as compared to the regression model but we are keen interested in learning Classification/ANN/Clustering equally well as we did in Linear regression so we are choosing classification model over the regression.