1. Introduction

Semiconductor manufacturing processes are highly technologically complicated processes with over a thousand steps, inevitably being prone to error due to how delicate nature and repetitive stages they have to undergo. Normally wafers are built out of different layers tangled with circuits and, if the final product is defective, it results in waste for the companies in terms of not only money but also time.

The need for a practical and automated solution to detect failures in early stages is evident, not only to avoid material waste but mainly to avoid releasing faulty products to the market. Therefore, in this work statistical analysis and modelling algorithms were used to develop such a solution, testing many different approaches and methods to define the ideal option for maximizing the quality of the prediction model.

1. Business understanding

Business objective

Lower down cases of defective wafers so as to increase productivity and reliability by means of prediction using data provided by sensors installed on semiconductor process.

Business success criteria

Reducing faulty semiconductors during manufacturing process to get 100% Quality Semiconductor products.

Data goals

a. Come up with a feature selection technique that chooses from over the 500 features, which are provided by sensors scattered across different semiconductor fabrication, the ones with the most potential so as reduce considerably the model’s processing time;

b. Since Data is imbalanced due to the amount of pass and fail cases, it is necessary to look for a technique that could help to compensate and help the model to increase accuracy;

c. Develop a model to detect on-time any possible fault raising during the wafer production.

1. Data understanding

Data description

Based on the descriptions in Kaggle and Research Gate, the SECOM dataset, provided by Michael McCann and Adrian Johnston, contain real data on the semi-conductor manufacturing process. The data comes from 590 sensors and process measurement points spread through the whole production line, in this case consisting of 1567 entries.

1. Data Preparation
2. Feature Removal