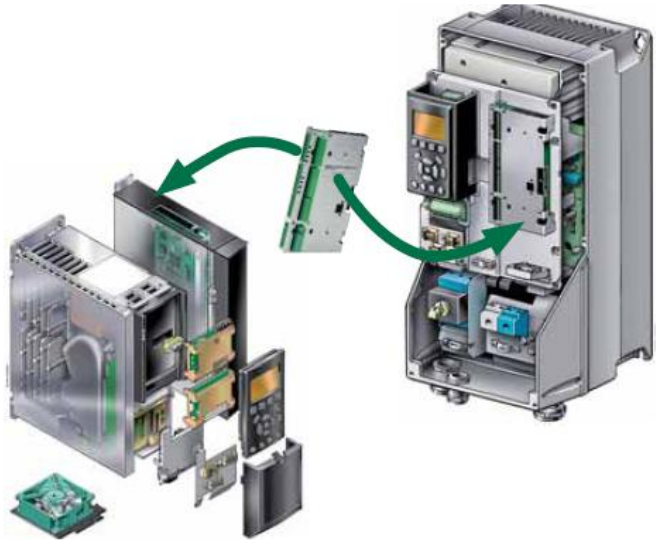
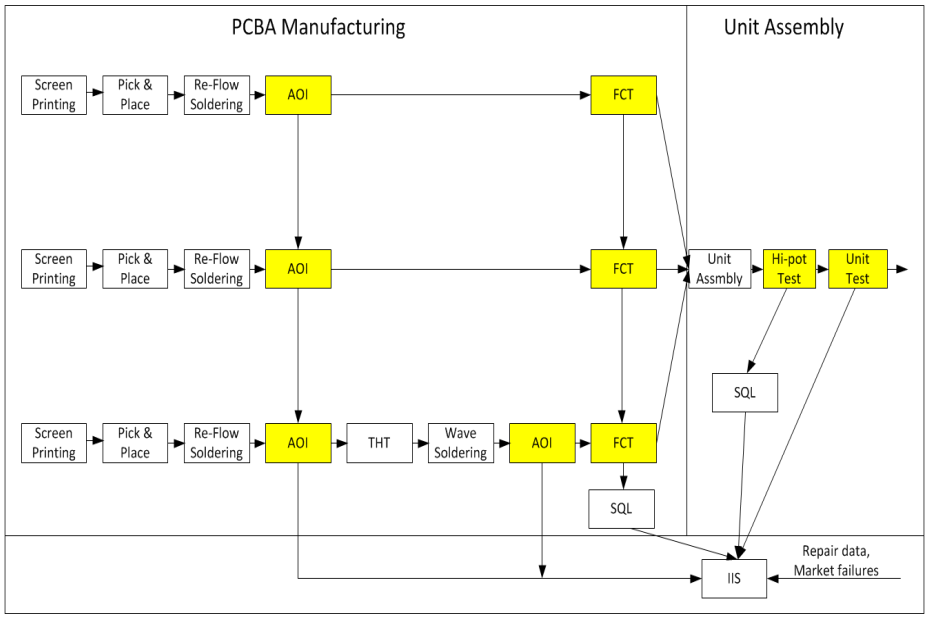
Analysis Of Quality Control Data: Faults Prediction In Case of Significant Noise and Imbalance

Abdul Rauf Khan, Henrik Schiøler, Torben Knudsen,
Murat Kulahci and Mohamed Zaki

UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Manufacturing Process (Of Industrial Partner)





UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Manufacturing Process (Of Industrial Partner)

Objective:

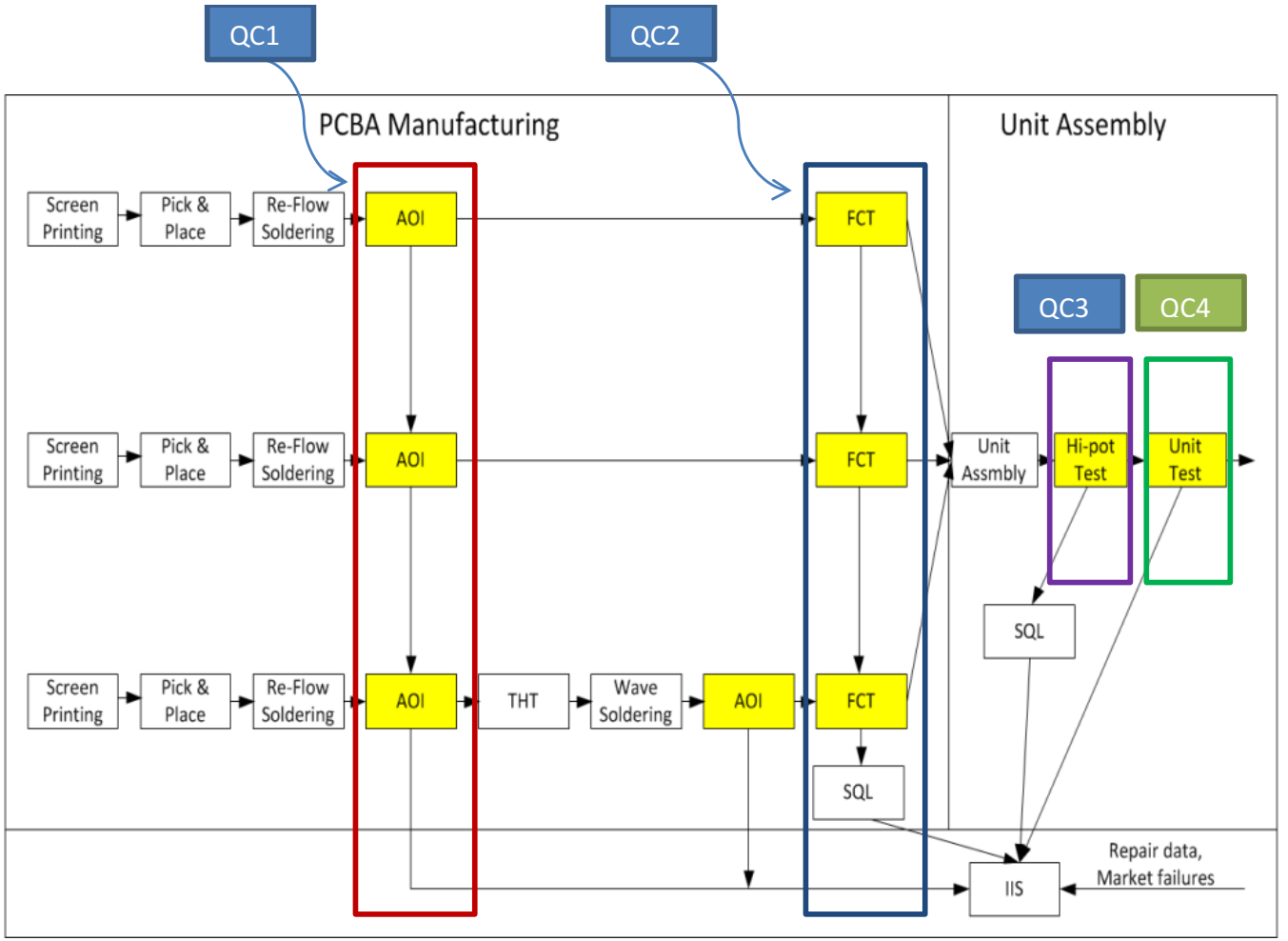
- Prediction of quality at last quality control stage

Challenges:

- Multi stage Manufacturing Process
- Heterogeneous Data streams

Gains:

- Optimization of resources
- Process understanding



UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Exploratory Analysis

A	B	C	D	E	F	G	H	I
Nr	Board 1 Vision	Board 1 Function	Board 2 Vision	Board 2 Function	Board 3 Function	VLT HSP	VLT Unit	Freq
1	F	F	F	F	F	F	F	0
2	P	F	F	F	F	F	F	0
3	R	F	F	F	F	F	F	0
4	F	P	F	F	F	F	F	0
5	P	P	F	F	F	F	F	0
6	R	P	F	F	F	F	F	0
7	F	R	F	F	F	F	F	0
8	P	R	F	F	F	F	F	0
9	R	R	F	F	F	F	F	0
10	F	F	P	F	F	F	F	0
11	P	F	P	F	F	F	F	0
12	R	F	P	F	F	F	F	0
13	F	P	P	F	F	F	F	0
14	P	P	P	F	F	F	F	0
15	R	P	P	F	F	F	F	0
16	F	R	P	F	F	F	F	0
17	P	R	P	F	F	F	F	0
18	R	R	P	F	F	F	F	0
19	F	F	R	F	F	F	F	0
20	P	F	R	F	F	F	F	0

DTU Technical University of Denmark

MADE Manufacturing Academy of Denmark

Danfoss

OF GE

Cambridge Service Alliance

Exploratory Analysis

Data Set 1

Test	Fail	Pass	Repair
AOI1	0	17505	1719
AOI2	52	18799	373
FT1	0	15527	3697
FT2	309	18892	23
FT3	299	18925	0
HV	2	19202	20
UT	4	19013	207

Data Set 2

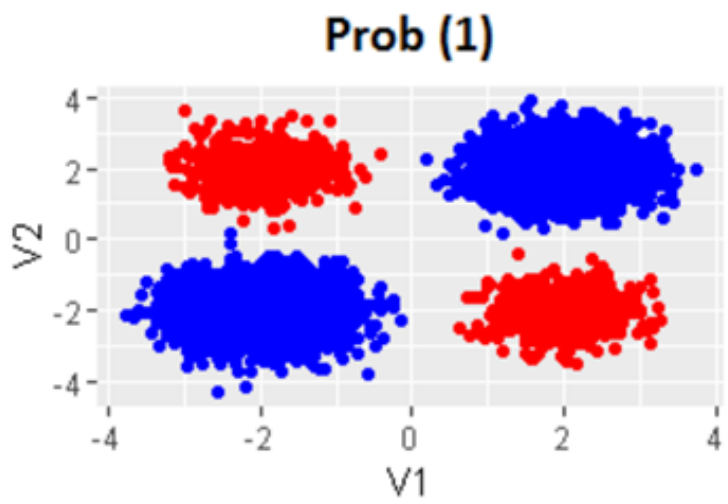
Test	Pass	Repair	Fail
AOI (Vision)	2213905	606315	0
FT (Function)	2059030	21813	40115
H.V (High Volt)	724244	2012	241
M.Load (Unit)	709287	21800	569

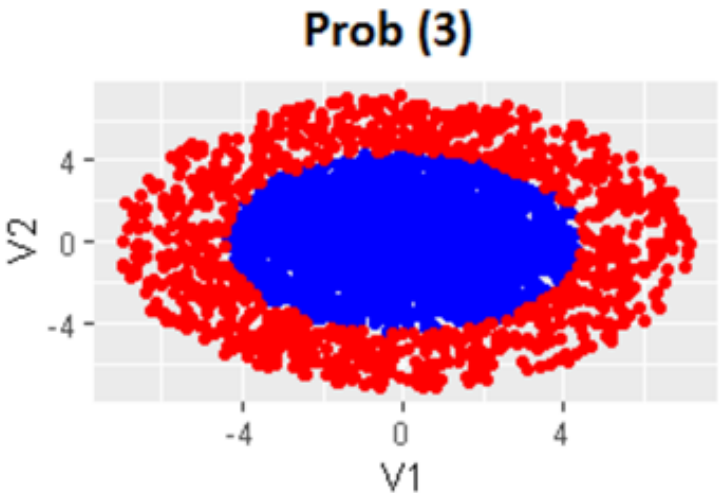
- **Primality of failure or repair at Unit test (UT)**

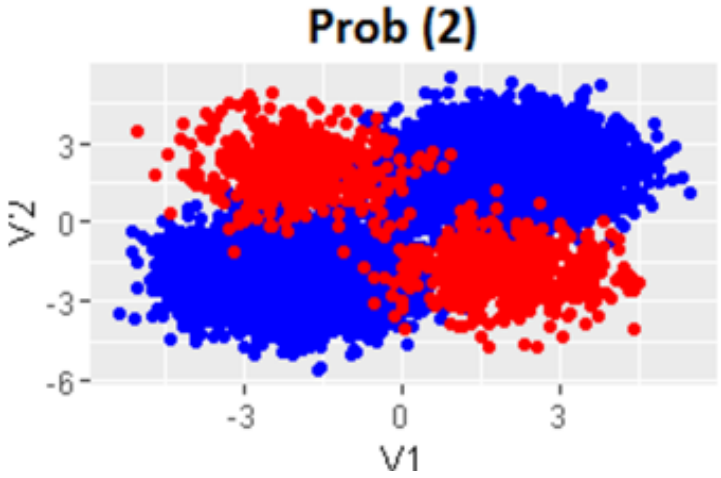
- Data Set 1: 0.0108
- Data Set 2: 0.0307

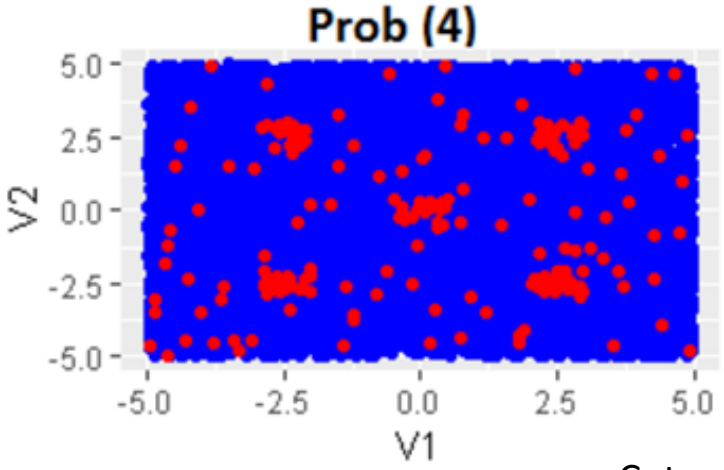
UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Imbalance and Noise vary among classification problems









- Category1: Blue
- Category2: Red

UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Existing Methods

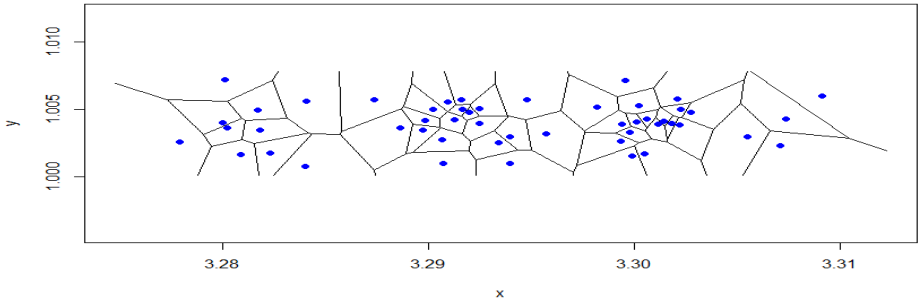
- State of the art

1. Pre processing methods
   - Resampling
   - Feature extraction (filter and wrapper techniques)

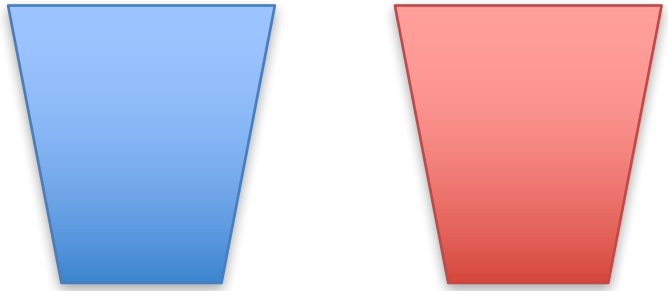
2. Cost Sensitive Methods
   - Boosting (adaboost, Gradient boost etc)

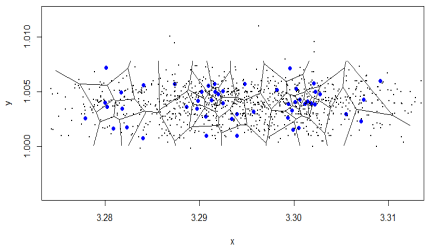
Proposed Methodology: Partition of Information Space and selection of best partition

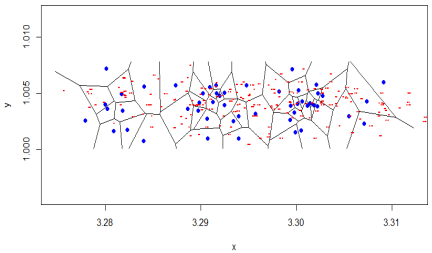
1. Partition of the feature space through Voronoi tessellation



2. Selection of the tiles according to their probability







UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Proposed Methodology to solve imbalance classification problem

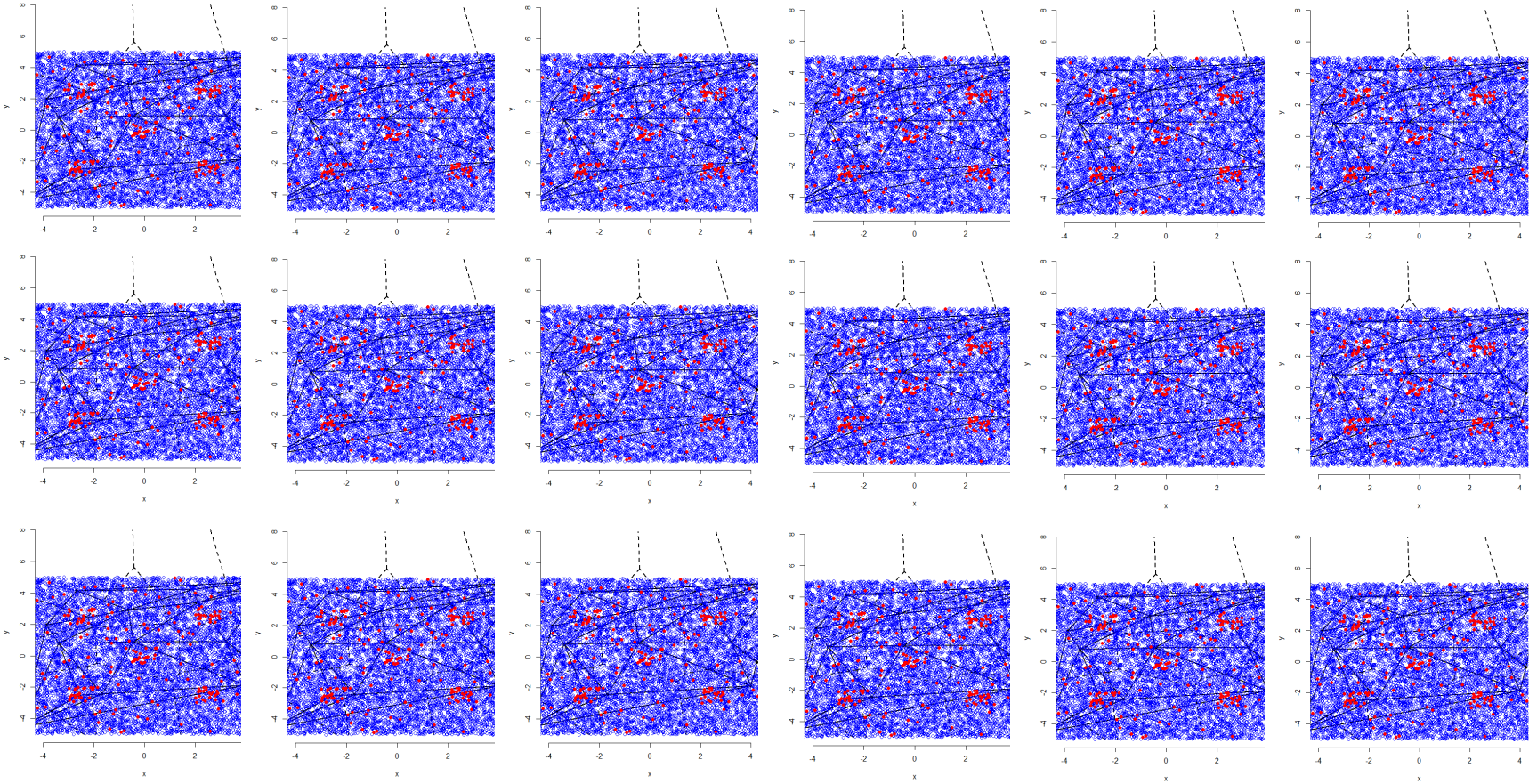
- Selection of Seed points

  Randomly
- LVQ to adjust the seed points

  SOM, the unsupervised way of finding the clusters
- LVQ output as final seed for voronoi tiles

  Voronoi tiles are based on the LVQ outputs
- Selecting voronoi tiles with the help of GA
  - Selection of the Number of tiles
  - Selection of Classes need objective function

Proposed Methodology: Partition of Information Space and selection of best partition

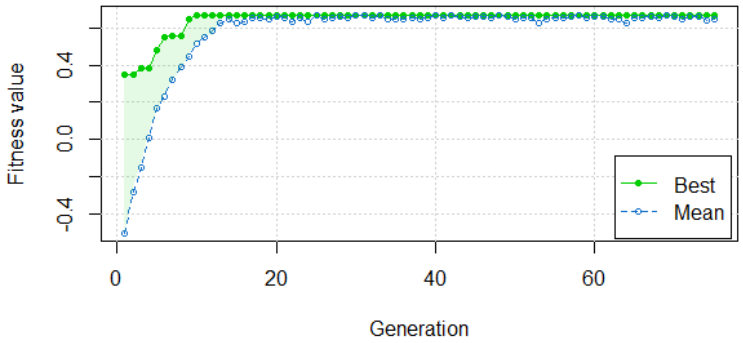


UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Assigning groups to the tiles (Results)

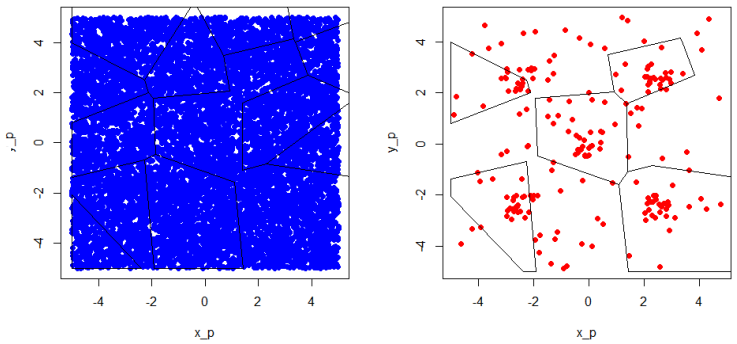
Status	Class B_{pass}	Class B_{fail}	
pass	$X_{pass,B_{pass}}$	$X_{pass,B_{fail}}$	$X_{pass.}$
fail	$X_{fail,B_{pass}}$	$X_{fail,B_{fail}}$	$X_{fail.}$
Total	$X_{.,B_{pass}}$	$X_{.,B_{fail}}$	$X_{..}$

Table 1: Contingency table for partion $\{B_{pass}, B_{fail}\}$



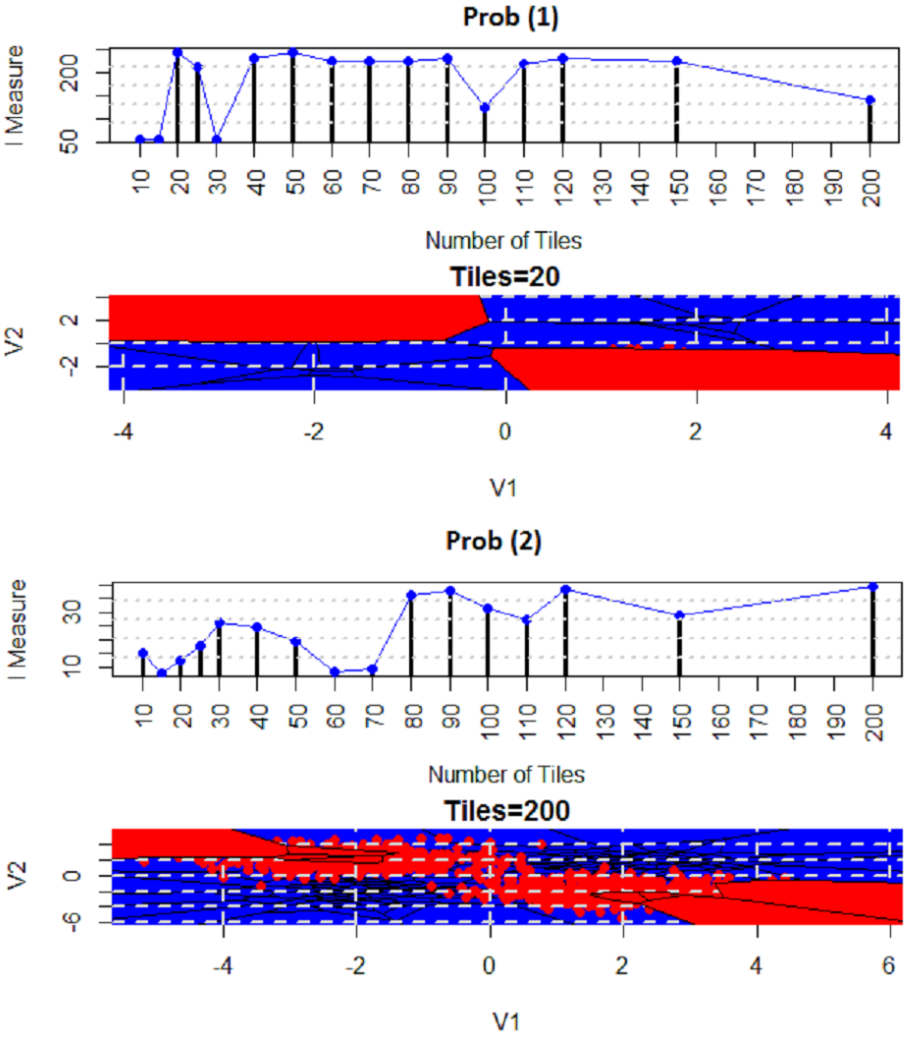
Objective Function

$$I = \frac{L_{fail} - U_{pass}}{(U_{pass} - L_{pass}) + (U_{fail} - L_{fail})}$$



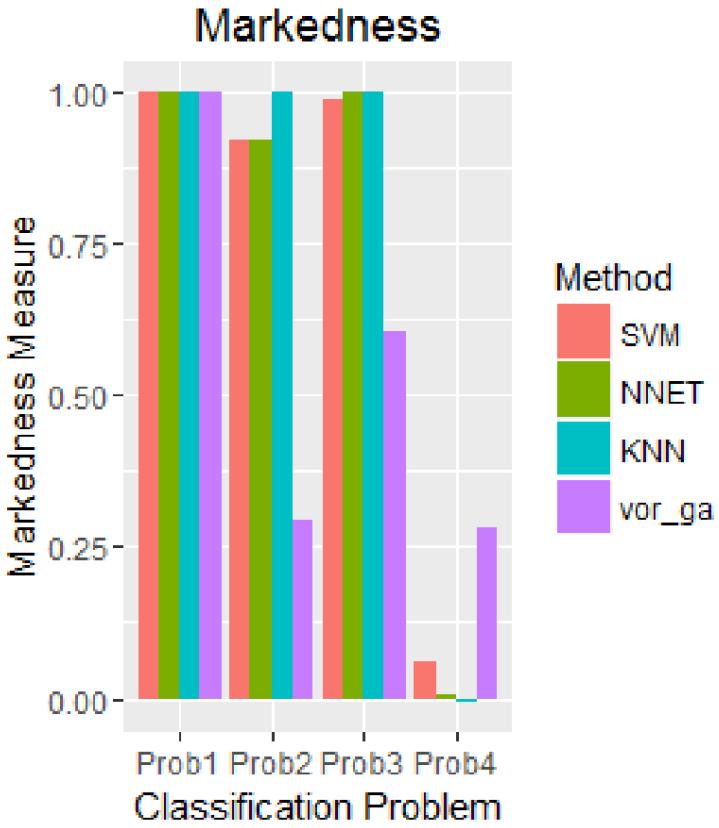
UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

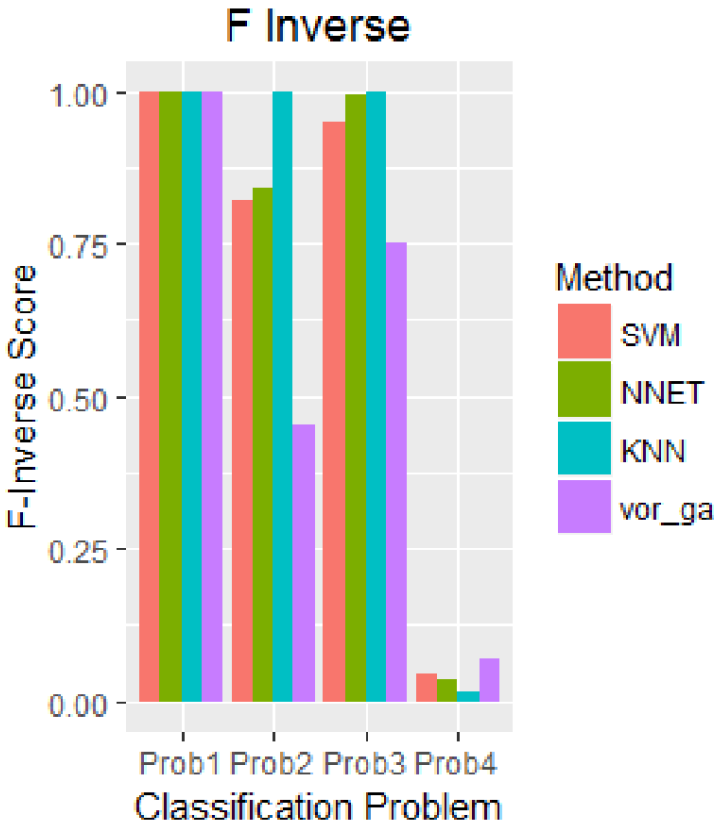
Classification through proposed Methodology (Results)



UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Performance of the proposed methodology (Results)





$$Markedness = Precision + Inverse\ Precision - 1 \quad (7)$$

Where,

$$Precision = \frac{Actual\ Positives \cap Predictive\ Positives}{Predictive\ Positives}$$

By above definition,

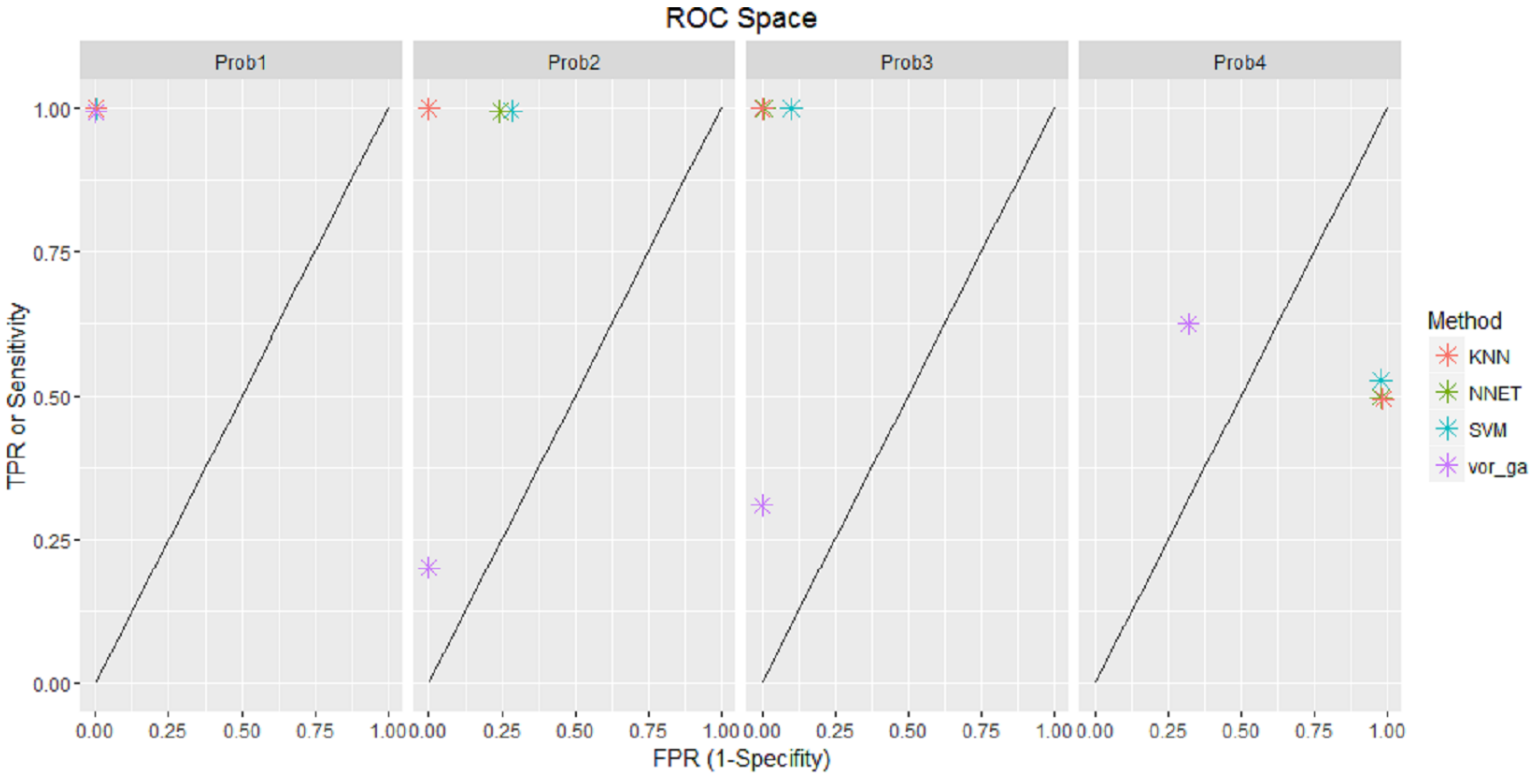
$$Precision = \frac{TP}{(TP+FP)} \text{ and } Inverse\ Precision = \frac{TN}{(TN+FN)}$$

		Prediction outcome: p	Prediction outcome: n	Total
Actual value	p′	True Positive (TP)	False Negative (FN)	P′
	n′	False Positive (FP)	True Negative (TN)	N′
Total		P	N	

$$F_1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Performance of the proposed methodology (Results)



$$\mathrm{FPR} = \frac{\mathrm{FP}}{N} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} = 1 - \mathrm{TNR}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

UNIVERSITY OF CAMBRIDGE
Cambridge Service Alliance

Conclusion

- A Novel methodology to classify imbalanced and noisy data

- Application of methodology in Industrial manufacturing

✓ Optimized quality control
✓ Optimization of Resources
✓ More Knowledge/understanding about process dynamics

Vision of next generation factory