Density Based Repairing and Clustering

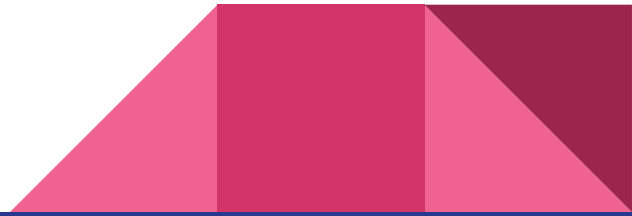
Guided By:
Dr. G.L. Prajapati

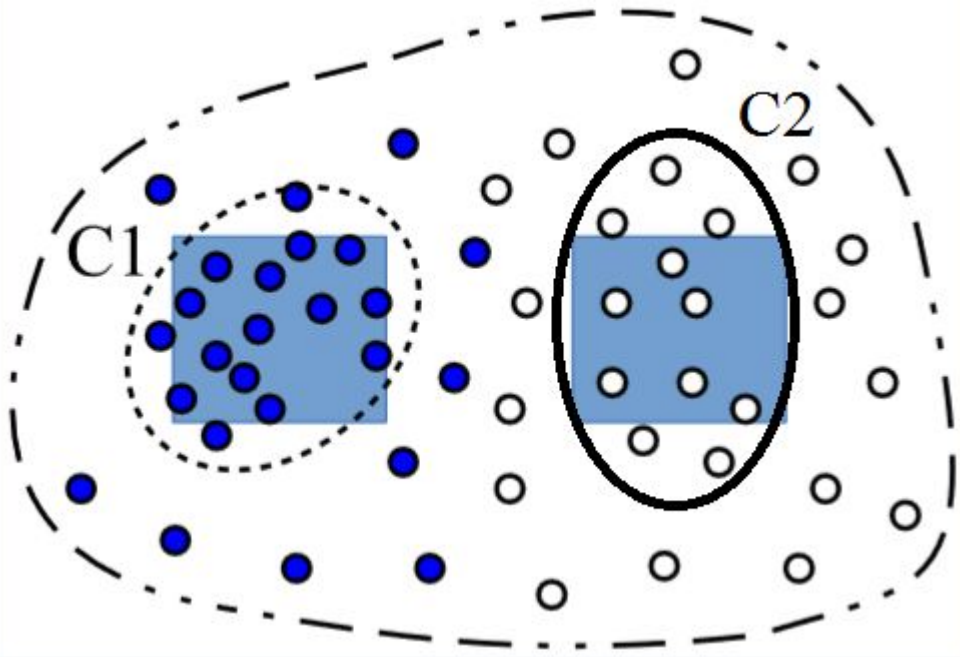
Submitted by:
Nitin Kumar Mittal (13C8032)
Pulkit Maloo (13C8037)
Tanay Kothari (13C8054)

Problem Domain

The need of optimal clustering is a problem faced in many domains in the world -

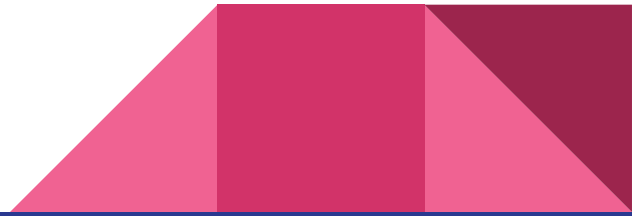
- ➔ GPS location accuracy
- ➔ Image compression
- ➔ Classification
- ➔ Market research
- ➔ Social network analysis




C1
C2

Solution Domain

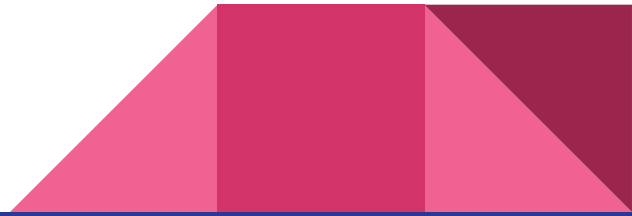
- ➔ Proposed solution
- ➔ Advantages over existing solutions
- ➔ Implementation



Proposed Solution

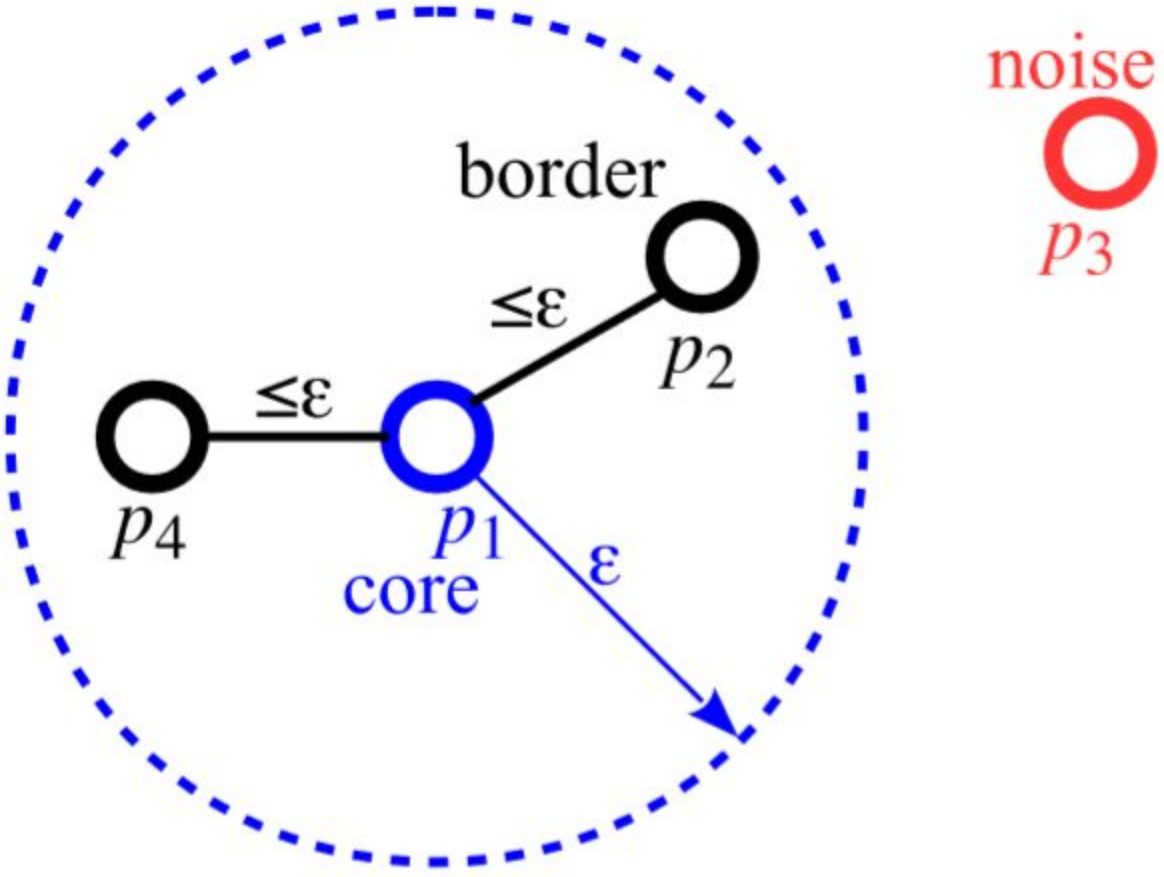
➔ Noise points are repaired

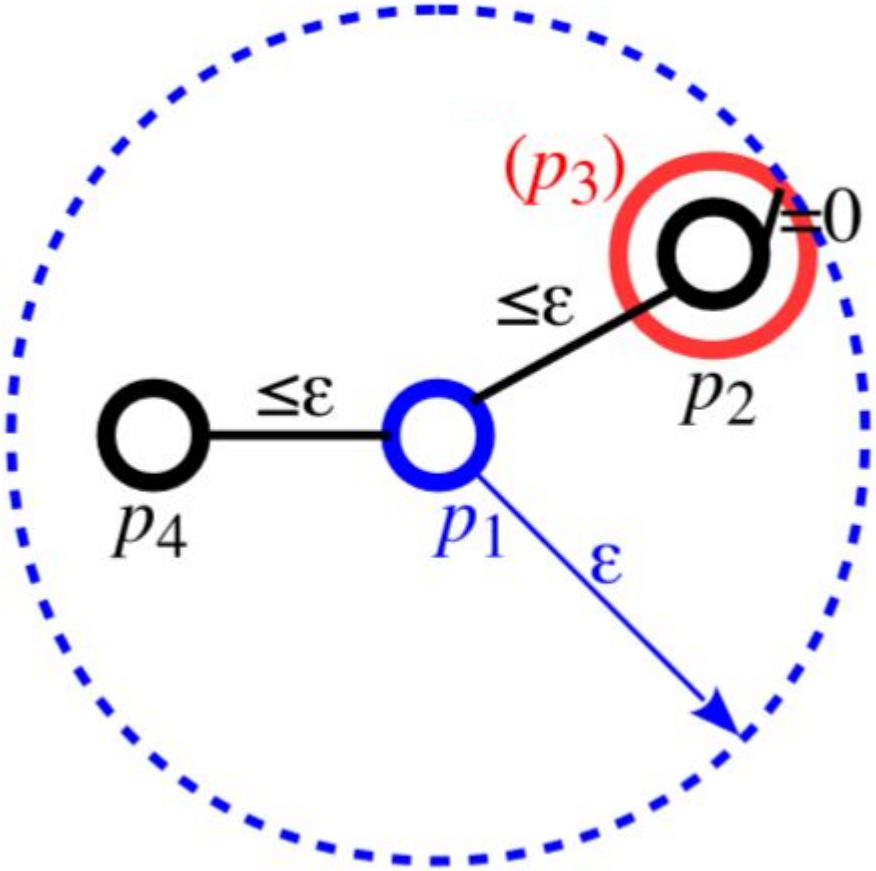
How?

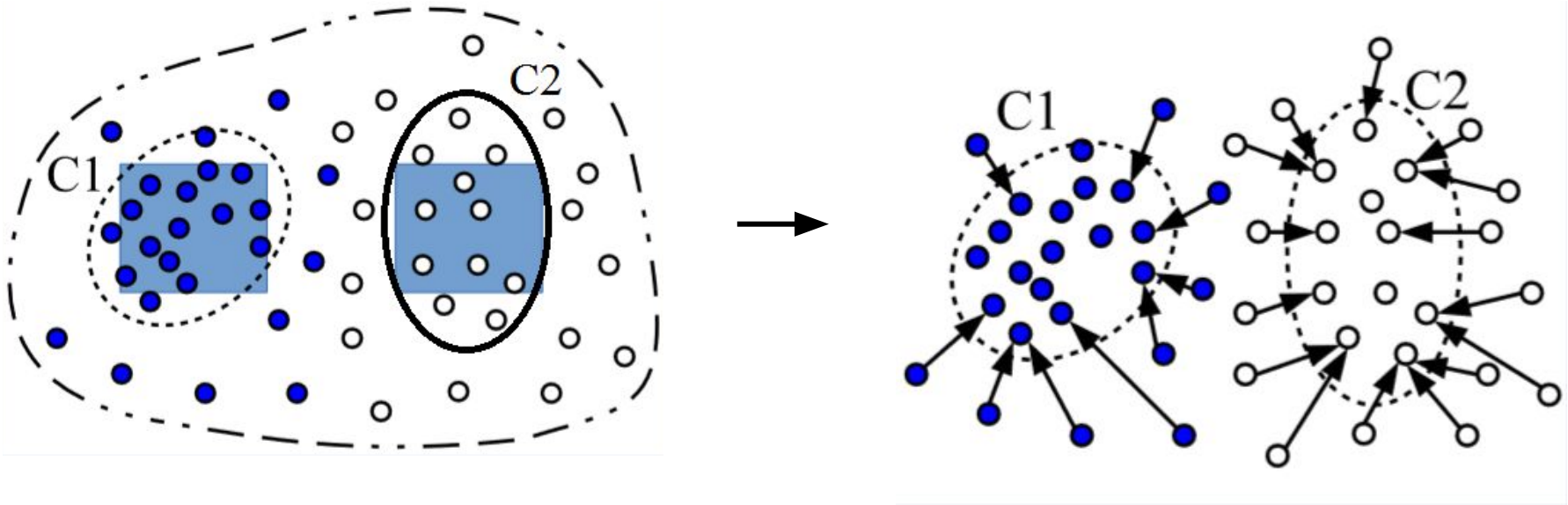


Proposed solution - Parameters

- ➔ Core Points
- ➔ Border Points
- ➔ Noise Points
- ➔ η (Density)
- ➔ ε (Distance)

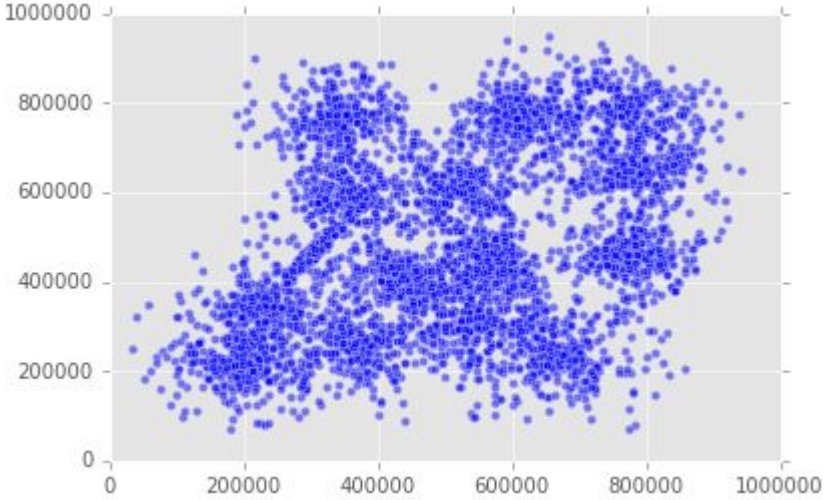

noise
p_3
border
$\leq\varepsilon$
p_2
$\leq\varepsilon$
p_4
p_1
core
ε


(p_3)
$=0$
$\leq\varepsilon$
$\leq\varepsilon$
p_2
p_4
p_1
ε

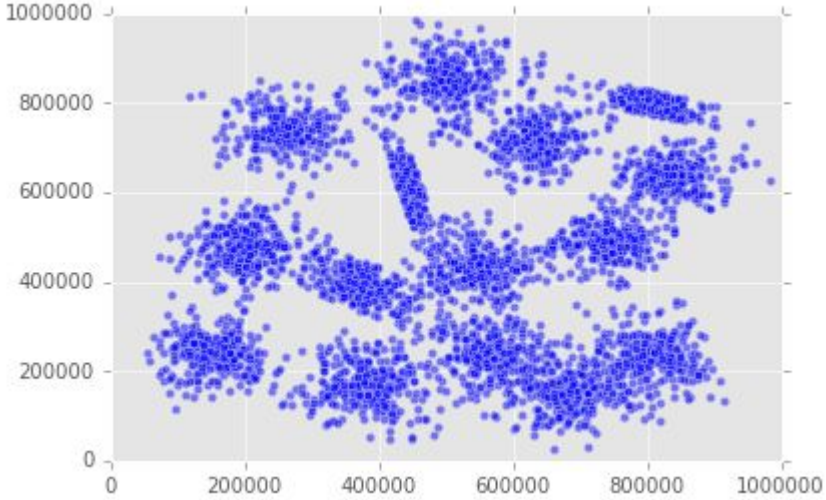


Noise is repaired instead of being discarded.

RESULTS

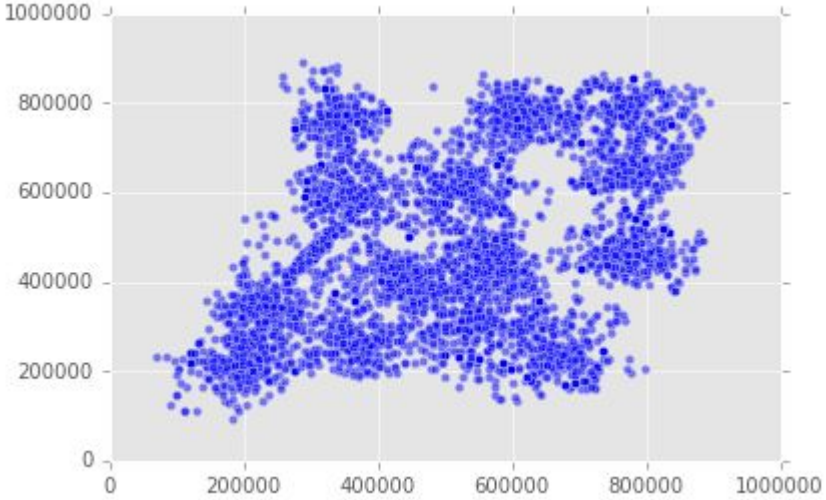


S2

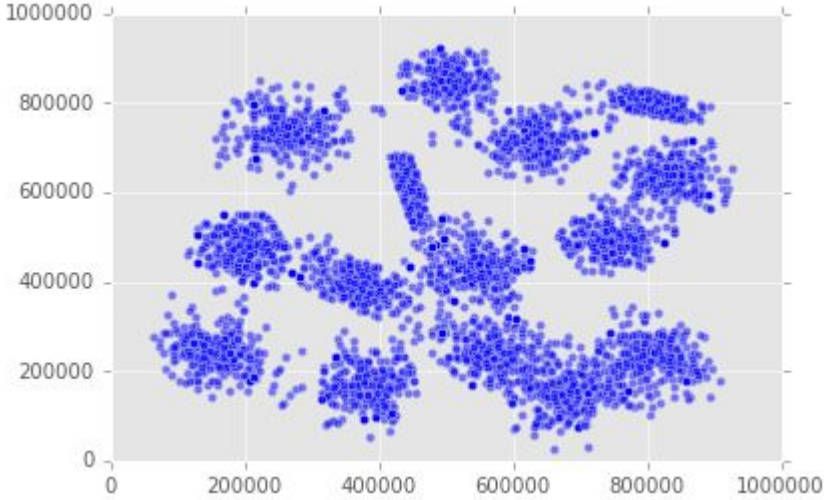


S3

Results - S2 & S3

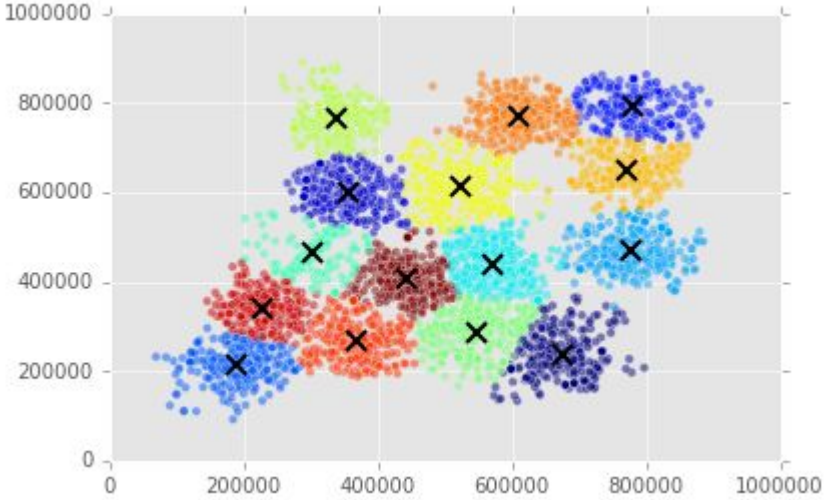


S2

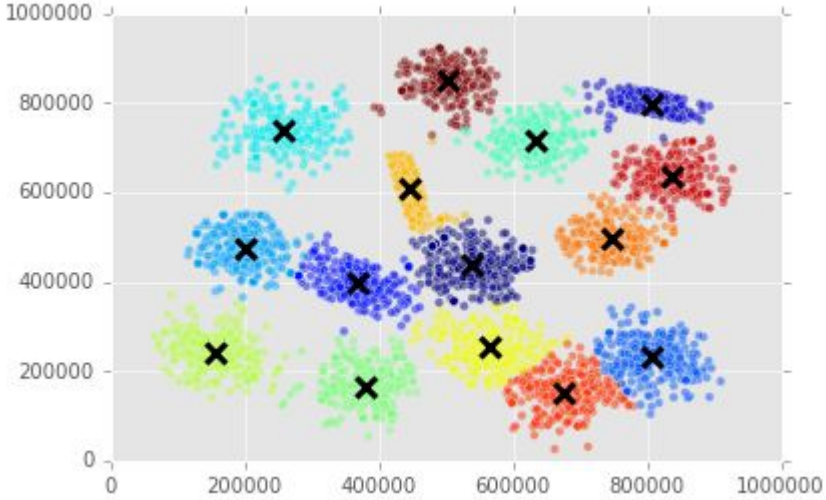


S3

Results - S2 & S3

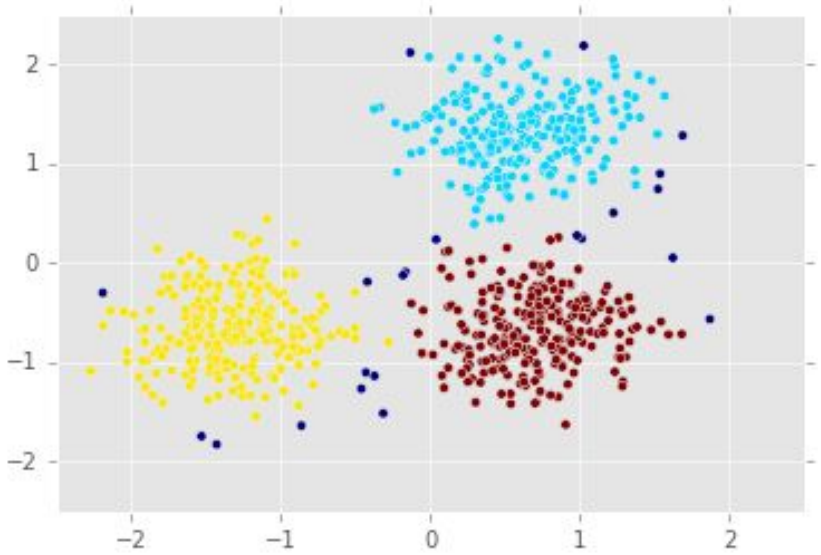


S2

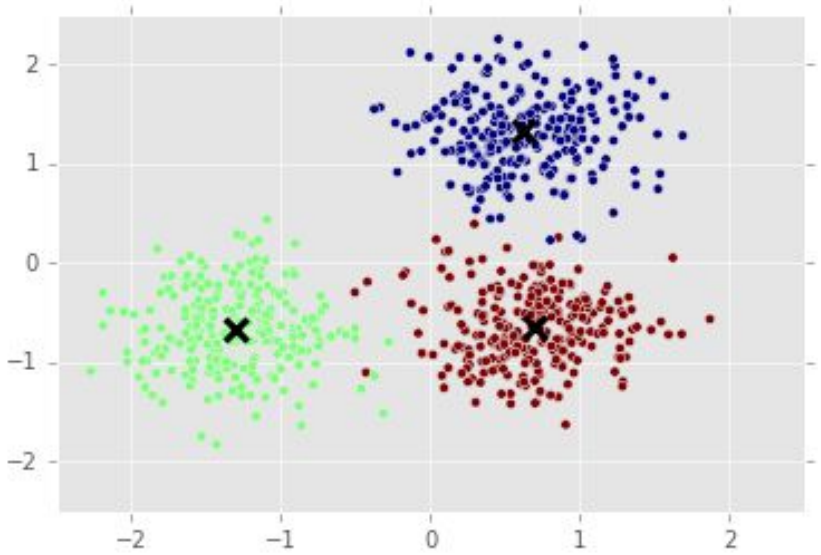


S3

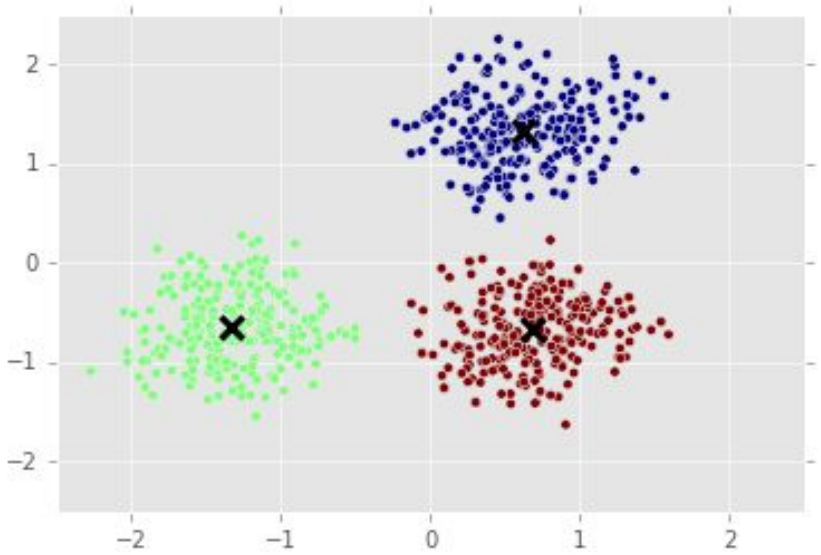
Results - S2 & S3



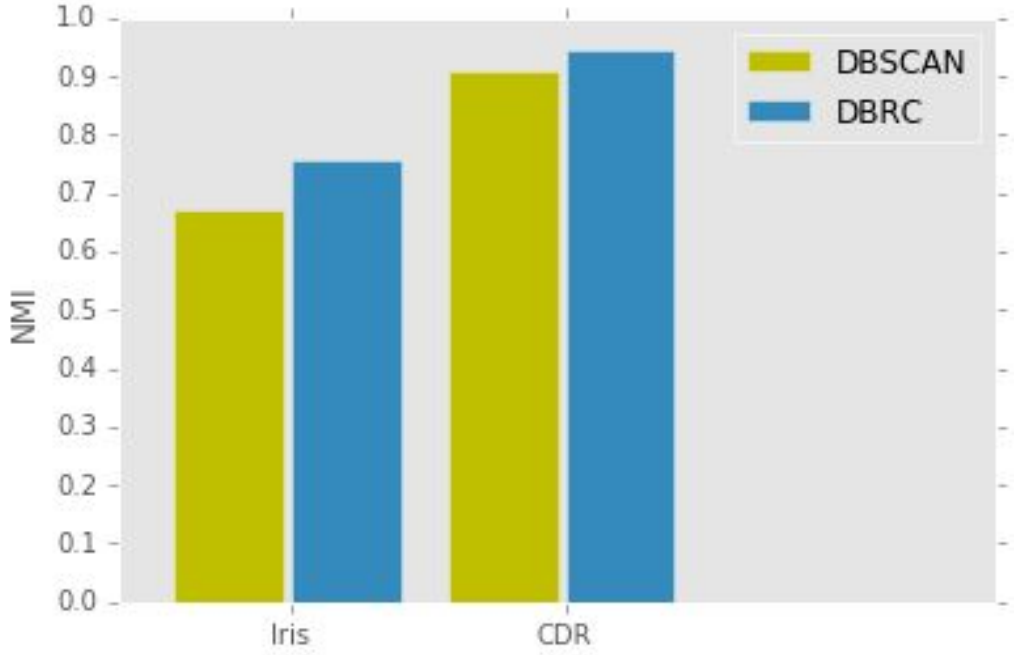
Application over Call Data Records



Application over Call Data Records



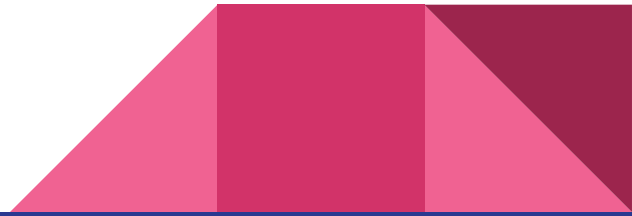
Application over Call Data Records



NMI Accuracy - Iris & CDR

Advantages Over Existing Solutions

- ➔ By utilizing noise in unclean data, more accurate clusters are formed
- ➔ No information lost



Thank You!