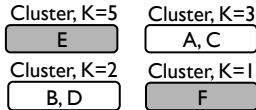


# Unlabeled Dataset

Instances	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	Label
Inst. A	3	1	3	0	5	1	9	?
Inst. B	1	1	2	0	7	3	8	?
Inst. C	2	3	2	5	5	2	1	?
Inst. D	1	0	8	1	0	1	9	?
Inst. E	5	0	2	5	5	10	8	?
Inst. F	0	1	1	4	2	1	1	?

Mean	2	1	3	2	4	3	6
------	---	---	---	---	---	---	---

(a) Clustering



K = the number of attributes whose feature values are greater than Mean.



Generated training set								
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	<i>Label</i>
Inst. E	5	0	2	5	5	10	8	Buggy
Inst. F	0	1	1	4	2	1	1	Clean

(b) Labeling



Final training set				
$X_1$	$X_5$	$X_6$	$X_7$	Label
5	5	10	8	Buggy
0	2	1	1	Clean

(c) Feature Selection