Sample	F	K
1	1	1
2	3	2
3	7	1
4	8	1
5	9	1
6	11	2
7	23	2
8	37	1
9	39	2
10	45	1
11	46	1
12	59	1

ChiMerge Discretization

•Statistical approach to Data Discretization

•Applies the Chi Square method to determine the probability of similarity of data between two intervals.

Sample	F	K	Intervals
1	1	1	{0,2}
2	3	2	{2,5}
3	7	1	{5,7.5}
4	8	1	{7.5,8.5}
5	9	1	{8.5,10}
6	11	2	{10,17}
7	23	2	{17,30}
8	37	1	{30,38}
9	39	2	{38,42}
10	45	1	{42,45.5}
11	46	1	{45.5,52}
12	59	1	{52,60}

ChiMerge Discretization Example

•Sort and order the attributes that you want to group (in this example attribute F).

•Start with having every unique value in the attribute be in its own interval.

ChiMerge Discretization Example

Sample	F	K
1	1	1
2	3	2
3	7	1
4	8	1
5	9	1
6	11	2
7	23	2
8	37	1
9	39	2
10	45	1
11	46	1
12	59	1

•Begin calculating the Chi Square test on every interval

Sample	K=1	K=2	
2	0	1	1
3	1	0	1
total	1	1	2

Sample	K=1	K=2	
3	1	0	1
4	1	0	1
total	2	0	2

ChiMerge Discretization Example

Sample	K=1	K=2	
2	0	1	1
3	1	0	1
total	1	1	2

$\mathbf{E_{11}} = (1/2)*1 = .05$
$\mathbf{E_{12}} = (1/2)*1 = .05$
$\mathbf{E_{21}} = (1/2)*1 = .05$
$\mathbf{E_{22}} = (1/2)*1 = .05$

$\mathbf{X^2} = (0-.5)^2/.5 + (1-.5)^2/.5 + (1-.5)^2/.5 + (0-.5)^2/.5 = \mathbf{2}$

Sample	K=1	K=2	
3	1	0	1
4	1	0	1
total	2	0	2

$\mathbf{E_{11}} = (1/2)*2 = 1$
$\mathbf{E_{12}} = (0/2)*2 = 0$
$\mathbf{E_{21}} = (1/2)*2 = 1$
$\mathbf{E_{22}} = (0/2)*2 = 0$

$\mathbf{X^2} = (1-1)^2/1+(0-0)^2/0+ (1-1)^2/1+(0-0)^2/0 = \mathbf{0}$

Threshold .1 with df=1 from Chi square distribution chart merge if $\mathbf{X^2} < 2.7024$

ChiMerge Discretization Example

Sample	F	K	Intervals	Chi^2
1	1	1	{0,2}	2
2	3	2	{2,5}	2
3	7	1	{5,7.5}	0
4	8	1	{7.5,8.5}	0
5	9	1	{8.5,10}	2
6	11	2	{10,17}	0
7	23	2	{17,30}	2
8	37	1	{30,38}	2
9	39	2	{38,42}	2
10	45	1	{42,45.5}	0
11	46	1	{45.5,52}	0
12	59	1	{52,60}	

- Calculate all the Chi Square value for all intervals
- Merge the intervals with the smallest Chi values

ChiMerge Discretization Example

Sample	F	K	Intervals	Chi^2
1	1	1	{0,2}	2
2	3	2	{2,5}	4
3	7	1	{5,10}	5
4	8	1		
5	9	1		
6	11	2	{10,30}	3
7	23	2		
8	37	1	{30,38}	2
9	39	2	{38,42}	4
10	45	1	{42,60}	
11	46	1		
12	59	1		

- Repeat

ChiMerge Discretization Example

Sample	F	K	Intervals	Chi²
1	1	1	{0,5}	
2	3	2		1.875
3	7	1		
4	8	1	{5,10}	
5	9	1		5
6	11	2		
7	23	2	{10,30}	
8	37	1		1.33
9	39	2	{30,42}	
10	45	1		1.875
11	46	1		
12	59	1	{42,60}	

•Again

ChiMerge Discretization Example

Sample	F	K	Intervals	Chi²
1	1	1	{0,5}	
2	3	2		1.875
3	7	1		
4	8	1	{5,10}	
5	9	1		3.93
6	11	2		
7	23	2		
8	37	1	{10,30}	
9	39	2		3.93
10	45	1		
11	46	1		
12	59	1	{42,60}	

•Until

ChiMerge Discretization Example

Sample	F	K
1	1	1
2	3	2
3	7	1
4	8	1
5	9	1
6	11	2
7	23	2
8	37	1
9	39	2
10	45	1
11	46	1
12	59	1

Intervals Chi^2

{0,10}

2.72

{10,30}

3.93

{42,60}

•There are no more intervals that can satisfy the Chi Square test.