

## 5 RESULTADOS

### 5.1 INTRODUÇÃO

Os resultados para a classificação de padrões via Rede de Kohonen e a análise fuzzy, utilizando o método descrito no capítulo 4, foram baseados em dados cedidos pela ELEKTRO para o município de Guarujá. Considerou-se, inicialmente, para construção dos *clusters* mais ou menos a metade dos clientes deste município (50.150 consumidores). Em seguida, fez-se uma comparação aumentando o tamanho da amostra e fazendo a “clusterização” com todos os clientes dessa localidade, cerca de 92.440 clientes.

As “clusterizações” foram feitas para um mapa auto-organizável de Kohonen com 16 neurônios, uma rede 4x4. Esta foi a dimensão da rede escolhida, pois quando tentou-se utilizar redes maiores, o número de neurônios sem PPH's tornou-se muito grande o que levava a ter-se neurônios que tinham um grande número de consumidores selecionados, mas que, ao mesmo tempo, não podiam ser utilizados para a análise fuzzy, já que possuíam um número insuficiente de pesquisas de posses e hábitos de consumo (menos de 20 pesquisas).

A seguir, são mostradas as análises para parte ou metade dos clientes de Guarujá e para todos os consumidores de energia elétrica desta localidade. Em seguida, é feita a comparação entre esses dois grupos de clientes. E finalmente, é feita a validação do modelo pela análise percentual de quantos clientes já identificados pela empresa ELEKTRO como fraudulentos e inadimplentes se encontravam nos *clusters* rotulados pela análise fuzzy como de consumidores fraudulentos ou inadimplentes.

## **5.2 ANÁLISE PARA METADE DOS CLIENTES DE GUARUJÁ**

### **5.2.1 Classificação por Rede de Kohonen**

Para a metade dos clientes de Guarujá (50.150 consumidores) obteve-se a “clusterização” que é mostrada na tabela 7. Além disso, é também apresentada, na seqüência, a tabela com cada um dos 16 *clusters* e o número de consumidores pesquisados em cada um destes. Destes *clusters* nem todos foram selecionados. Como dito anteriormente, apenas aqueles que possuíam no mínimo 20 pesquisas de posses e hábitos de consumo (PPH) foram tomados para análise fuzzy.

Tabela 7: Número de consumidores e clientes pesquisados por *cluster*  
(metade de Guarujá)

Clusters	Consumidores	Clusters	Pesquisas
1	10.480	1	111
2	220	2	2
3	3.128	3	164
4	3.593	4	144
5	2.151	5	52
6	0	6	0
7	2.138	7	74
8	1.795	8	15
9	1.209	9	23
10	2.769	10	138
11	4.063	11	216
12	0	12	0
13	5.860	13	0
14	873	14	9
15	5.398	15	129
16	6.473	16	161
<b>Total</b>	<b>50.150</b>	<b>Total</b>	<b>1.238</b>

Como se pôde ver na tabela anterior apenas dez *clusters* devem ser selecionados para a análise seguinte por possuírem o número mínimo de pesquisas de posses e hábitos de consumo, são eles: 1, 3, 4, 5, 7, 9, 10, 11, 15, e 16.

## 5.2.2 Análise Fuzzy

Escolhidos os *clusters* e as variáveis a serem utilizadas para a validação, que são aquelas escolhidas no capítulo 4 e que passaram pelos cortes definidos anteriormente, o passo seguinte foi construir as curvas de pertinência para os clientes normais, fraudulentos e inadimplentes. Estas curvas são apresentadas na seqüência, onde a figura 31 apresenta as curvas de pertinência dos consumidores adimplentes, a figura 32 mostra a curva de pertinência dos clientes fraudulentos e, finalmente, a figura 33 a curva dos consumidores inadimplentes.

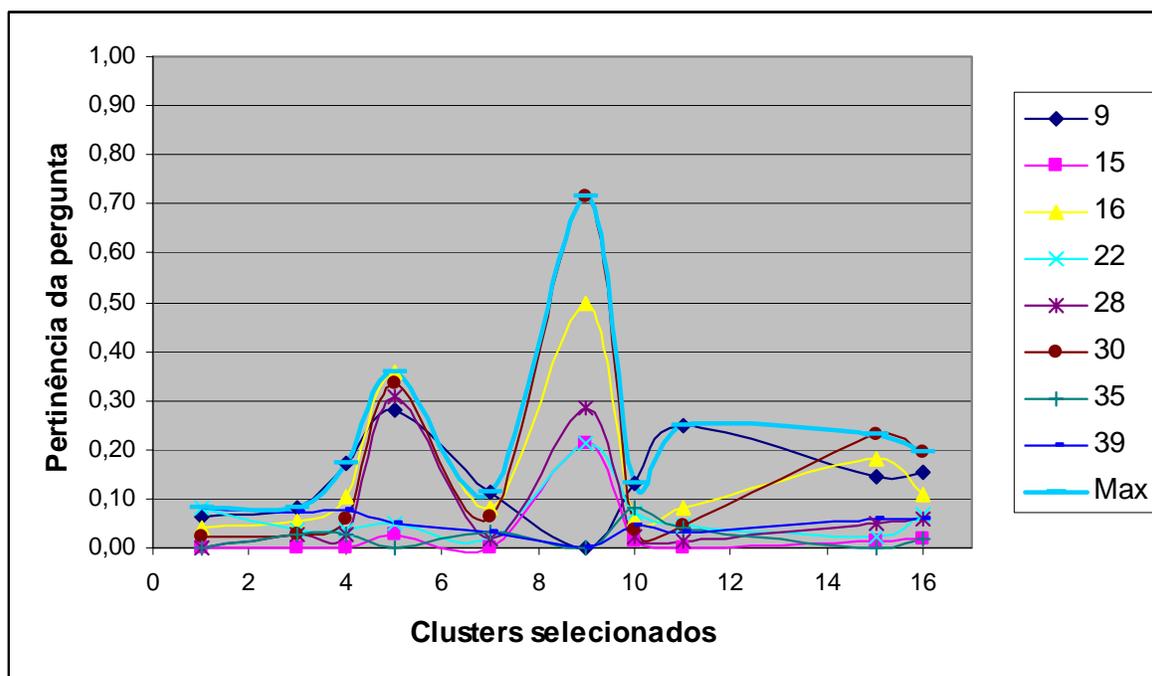


Figura 31: Curvas de pertinência de clientes adimplentes para metade de Guarujá

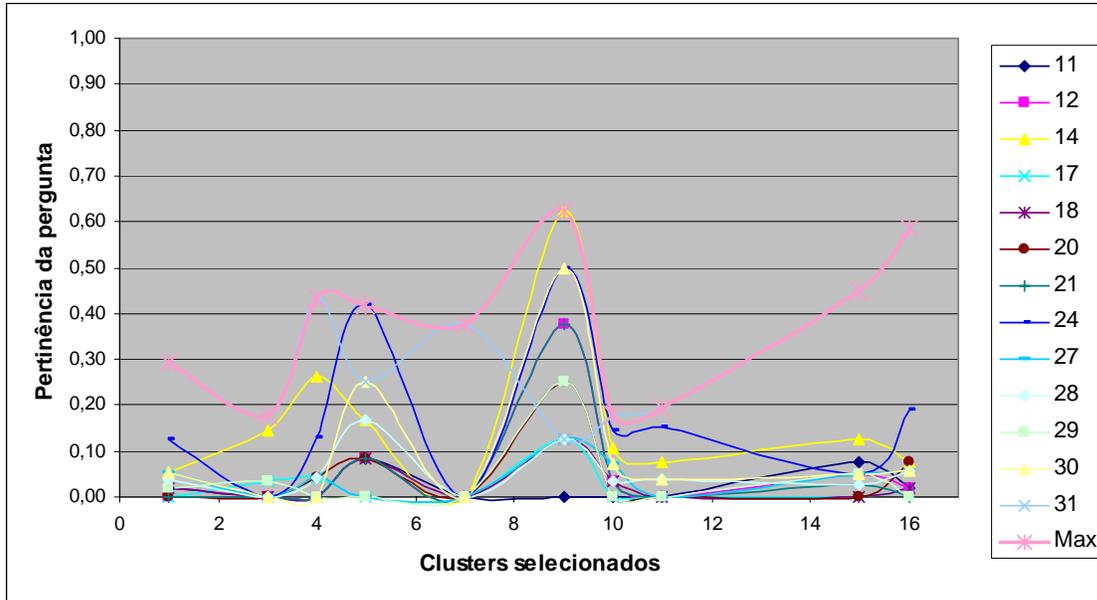


Figura 32: Curvas de pertinência de clientes fraudulentos para metade de Guarujá

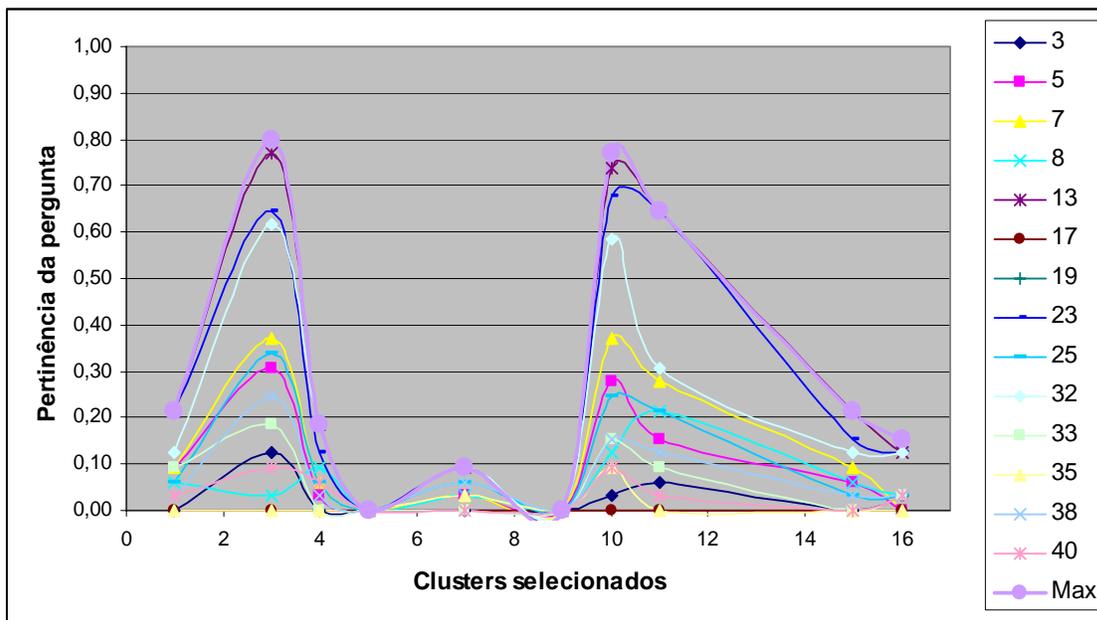
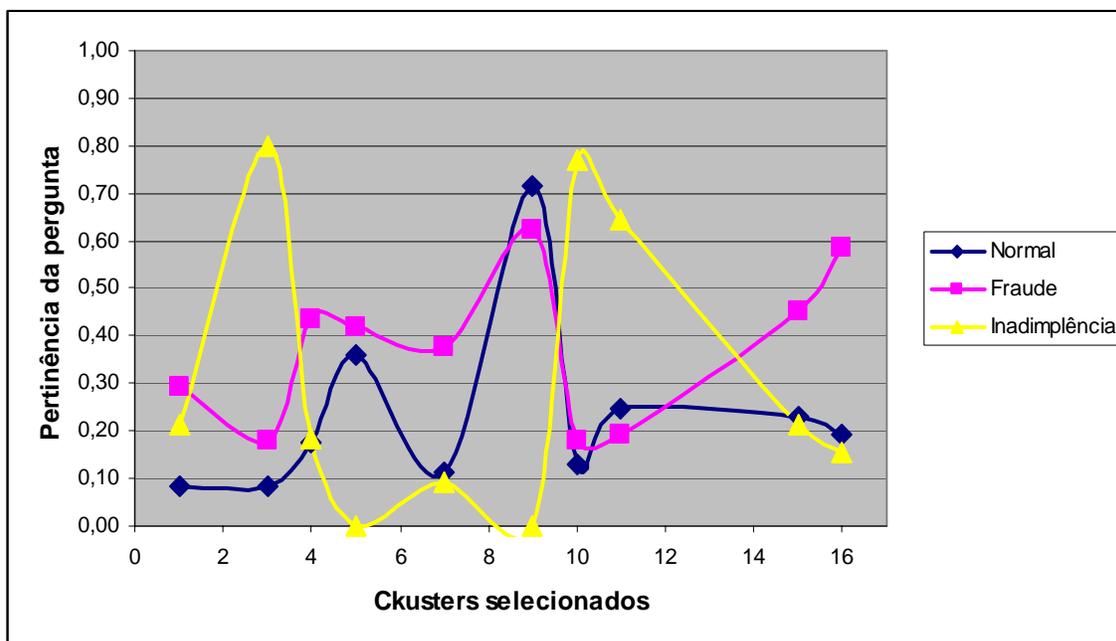


Figura 33: Curvas de pertinência de clientes inadimplentes para metade de Guarujá

Plotadas as curvas de pertinência e definidas as curvas de máximo pôde-se então partir para análise final destas curvas que é mostrada na figura 34. Nesta, constam as curvas de máximo obtidas a partir das curvas

de pertinências dos clientes normais fraudadores e inadimplentes mostradas nas figuras 31, 32 e 33.



**Figura 34: Resultado final da classificação fuzzy dos clusters para metade de Guarujá**

Analisando-se a figura 34, pode-se notar que os clusters 3, 10 e 11 apresentam o maior número de clientes inadimplentes e que a existência de fraude nestes três grupos é pouco expressiva. Já quando a análise é feita para os consumidores fraudulentos é visível que os clusters 9, 15 e 16 são os que têm maior representatividade deste “tipo” de cliente. Deve-se ressaltar que o grupo 9 possui um grande número de clientes normais, é possível especular que muitos destes podem estar “camuflados” e serem possíveis fraudadores de energia elétrica da empresa. Vê-se também que os clusters 4, 5 e 7 também possuem um número considerável de clientes fraudulentos e devem ser considerados na busca por fraudes.

## 5.3 ANÁLISE PARA TODOS OS CLIENTES DE GUARUJÁ

### 5.3.1 Classificação por Rede de Kohonen

Já para total dos clientes de Guarujá (92.440 consumidores) obteve-se a “clusterização” que é mostrada na tabela que se segue. Além disso, é também apresentada a seguir a tabela com cada um dos 16 *clusters* e o número de consumidores pesquisados em cada um destes.

Tabela 8: Número de consumidores e clientes pesquisados por *cluster* (Guarujá)

Clusters	Consumidores
1	6106
2	4203
3	5150
4	5573
5	4313
6	858
7	4
8	8090
9	12540
10	808
11	2015
12	5365
13	0
14	13156
15	13791
16	10468
<b>Total</b>	<b>92440</b>

Clusters	Pesquisas
1	291
2	3
3	24
4	0
5	7
6	6
7	0
8	24
9	52
10	6
11	88
12	365
13	0
14	83
15	158
16	131
<b>Total</b>	<b>1238</b>

Pode-se notar na tabela 8 que apenas nove *clusters* podem ser selecionados para a análise fuzzy por possuírem o número mínimo de 20 de pesquisas de posses e hábitos de consumo, são eles: 1, 3, 8, 9, 11, 12, 14, 15, e 16.

### **5.3.2 Análise Fuzzy**

Tomadas as variáveis e selecionados os *clusters* a serem utilizados para a validação do modelo, o passo seguinte foi construir as curvas de pertinência para os três “tipos” de clientes. Estas curvas são apresentadas a seguir, onde a figura 35 apresenta as curvas de pertinência dos clientes normais, a figura 36 mostra a curva de pertinência dos consumidores fraudadores e na figura 37 a curva dos consumidores inadimplentes é apresentada.

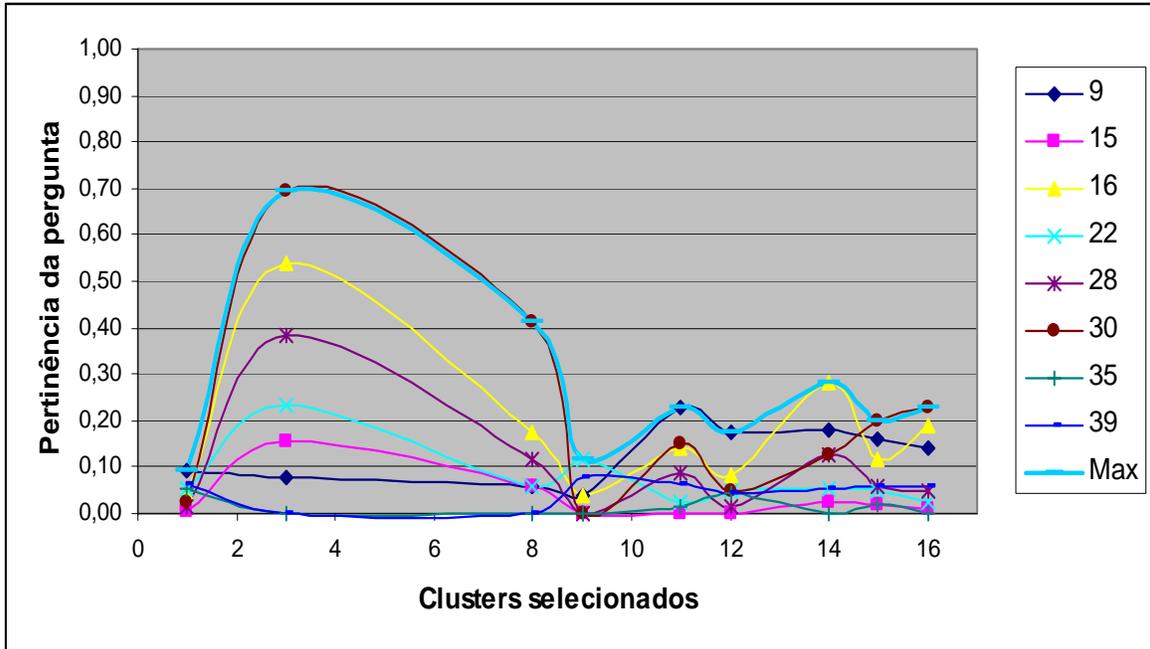


Figura 35: Curvas de pertinência de clientes adimplentes de Guarujá

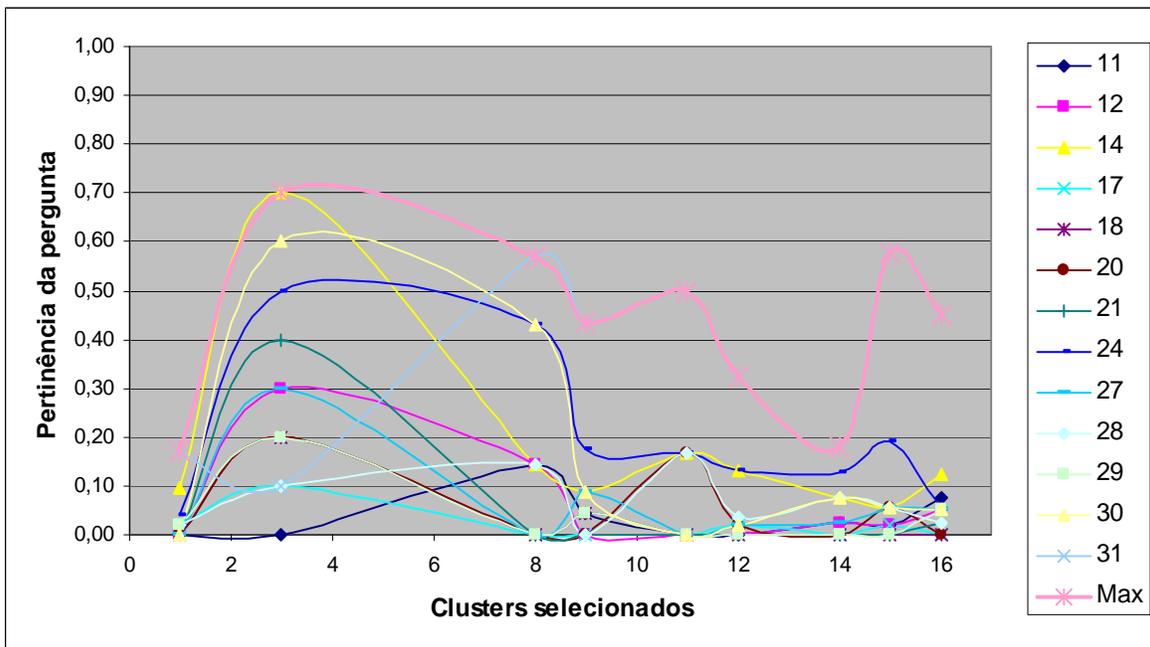
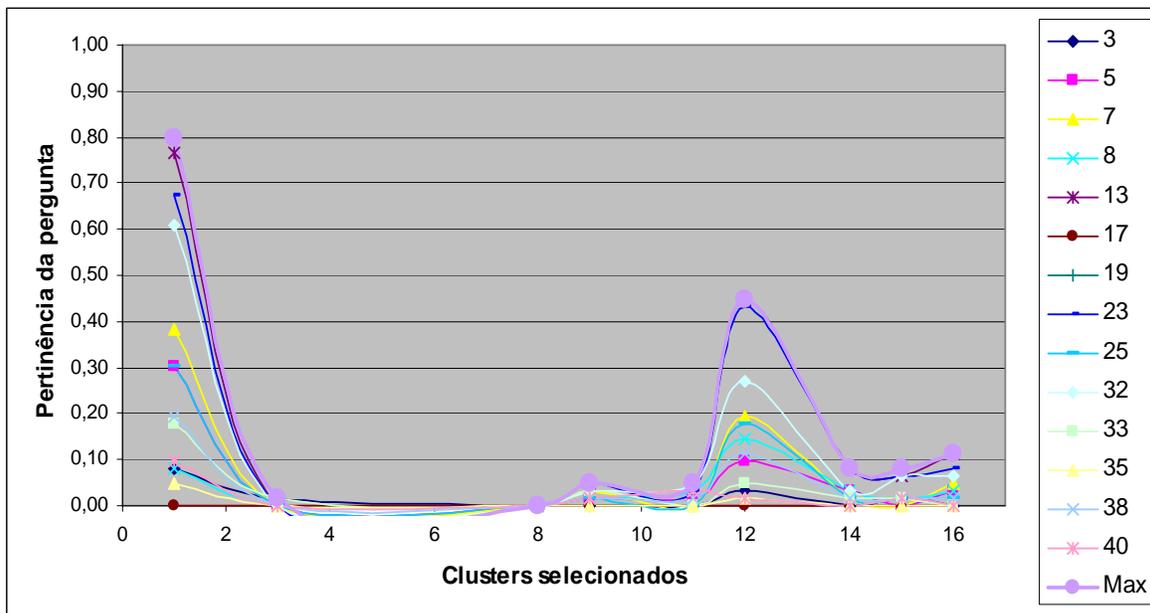
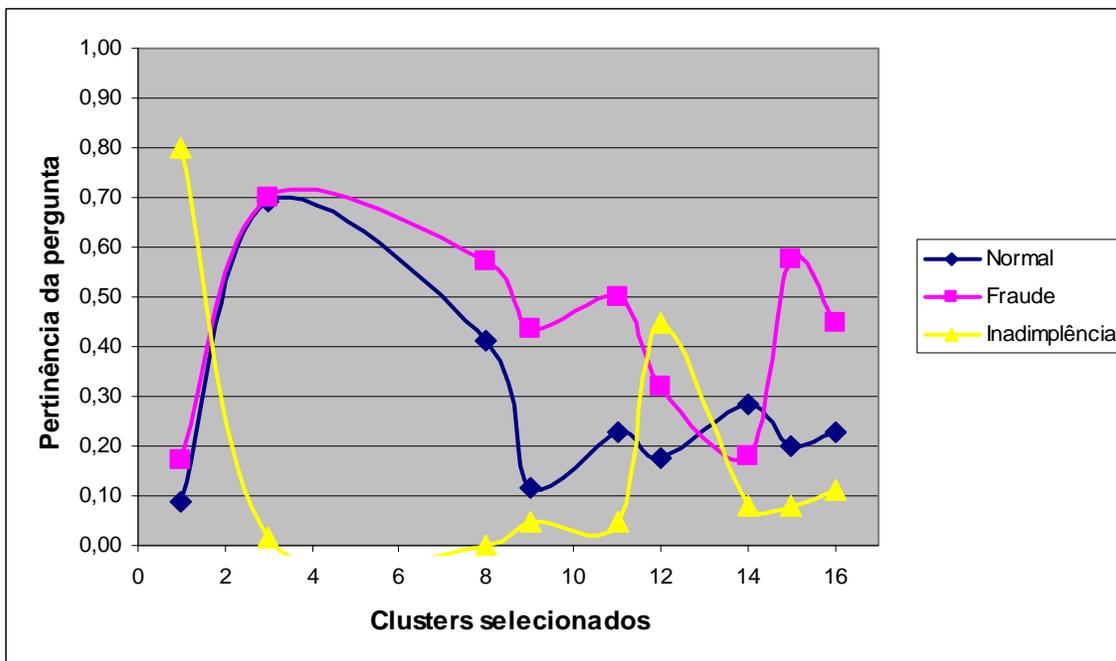


Figura 36: Curvas de pertinência de clientes fraudulentos de Guarujá



**Figura 37: Curvas de pertinência de clientes inadimplentes de Guarujá**

Geradas as curvas de pertinência e definidas as curvas de máximo de cada um dos três segmentos de clientes, pôde-se então fazer a análise final destas curvas que é mostrada na figura que se segue. Nesta figura, constam as curvas de máximo obtidas a partir das curvas de pertinências dos clientes normais fraudulentos e inadimplentes mostradas nas figuras 35, 36 e 37 anteriormente apresentadas.



**Figura 38: Resultado final da classificação fuzzy dos *clusters* para metade de Guarujá**

Quando se analisa a figura 38, nota-se que os *clusters* 1 e 12 apresentam o maior número de clientes inadimplentes e que a existência de fraude nestes só é significativa no grupo 12. Analisando-se a fraude, grande parte dos *clusters* a possui. No entanto, os grupos com maior incidência deste “tipo” de cliente são: 3, 8, 11, e 15.

Deve-se atentar que em uma análise mais acurada de fraudes que os *clusters* 3 e 8 não devem ser considerados por possuírem um número muito grande de clientes normais. No entanto, o grupo 3 é aquele que tem o maior grau de respostas dos clientes fraudulentos, o que nos leva a ficar em dúvida quanto a tomar ou não este *cluster* para uma análise. É possível especular que neste grupo muitos destes clientes adimplentes podem estar “camuflados” aos clientes fraudadores.

## 5.4 ANÁLISE FINAL DOS RESULTADOS

### 5.4.1 Introdução

Para validação do modelo duas análises foram feitas: a primeira foi verificar para os clientes de metade de Guarujá, que foram classificados pelo processo fuzzy como fraudulentos ou inadimplentes, quais estavam nos *clusters* de inadimplência e fraude para toda Guarujá; em seguida, fez-se uma análise de quantos clientes que foram identificados pela ELEKTRO como reais inadimplentes ou reais fraudadores estavam inseridos nos *clusters* mais significativos quanto à fraude e à inadimplência no município de Guarujá.

### 5.4.2 Comparação entre Metade e Todos os Clientes de Guarujá

Esta comparação foi feita primeiramente para os clientes fraudulentos e em seguida para os clientes inadimplentes.

#### 5.4.2.1 Clientes Fraudulentos

Para os clientes fraudadores, quando se considerou metades dos clientes de Guarujá, foram obtidos os *clusters* 9, 15 e 16 (figura 37) como os de maiores incidência de respostas à fraude. Quando se considerou todos os clientes deste município, a maior incidência de fraudes ocorreu nos

*clusters* 3, 8, 11 e 15 (figura 38). No entanto, o *cluster* 8 não foi considerado para esta análise por ter uma pertinência para os clientes normais muito parecida com os clientes de fraude. Seguindo este último critério, o *cluster* 3 também deveria ser descartado. Entretanto, este grupo não o foi pelo fato de ser o grupo com maior incidência de fraudes.

A seguir, tem-se a tabela que mostra a quantidade de clientes em cada um dos *clusters* mais significativos quanto a fraude para os clientes de metade Guarujá. Consta também nesta mesma tabela o número de clientes coincidentes, ou seja, aqueles que estavam nos *clusters* 3, 11 e 15 de toda Guarujá (total de 23.004 clientes) e que se encontravam nos *clusters* 9, 15 e 16 de metade de Guarujá (13.080 consumidores).

**Tabela 9: Comparação entre os clientes fraudulentos de metade e toda Guarujá.**

Metade de Guarujá		Clientes Coincidentes
Cluster	Clientes	
9	1.209	852
15	5.398	3.783
16	6.473	4.571
Total	<b>13.080</b>	<b>9.206</b>

Nota-se que dos 13.080 clientes de metade de Guarujá, 9.206 foram encontrados em toda Guarujá. Tem-se assim, um percentual de 70,38% de clientes encontrados, o que é um resultado muito bom e que comprova a eficácia do modelo para a busca de clientes com algum tipo de fraude.

### 5.4.2.2 Clientes Inadimplentes

Quando se fez a análise fuzzy para os 50.150 clientes de Guarujá, viu-se que os *clusters* que mais tinham incidência de clientes inadimplentes eram 3, 10 e 11 (9.960 consumidores), figura 37. Já para todos, os 92.440 clientes desta mesma localidade, o número de *clusters* selecionados com maiores incidências de inadimplência foram 1 e 12 (11.461 consumidores), figura 38.

A tabela 10 mostra a quantidade de clientes em cada um dos *clusters* mais significativos quanto a inadimplência para metade de Guarujá. Além disso, nesta tabela consta o número de clientes coincidentes que estavam nos clusters de metade de Guarujá e que também se encontravam nos clusters de toda Guarujá.

**Tabela 10: Comparação entre os clientes inadimplentes de metade e toda Guarujá.**

Metade de Guarujá		Clientes Coincidentes
Cluster	Clientes	
3	3.128	1.520
10	2.769	854
11	4.063	2.358
Total	<b>9.960</b>	<b>4.732</b>

Pode-se notar que foram encontrados 4.732 clientes em metade de Guarujá dos 9.960 consumidores que constavam na amostra de todos os clientes desta localidade. Com isto tem-se um percentual de 47,51% de clientes encontrados.

Nota-se que o resultado para este “tipo” de cliente não foi muito bom. Este fato pode ser explicado pelo tamanho da amostra para este segmento. Esta amostra constava com apenas 102 pesquisas de posses e hábitos. Este é um valor de pouca significância estatística, o que induz a um maior valor de erro do modelo.

### **5.4.3 Procura dos Dados Reais nos *Clusters* Gerados**

Neste tópico, tem-se por objetivo a verificação se estão inseridos nos *clusters* caracterizados como de consumidores inadimplentes ou de fraudulentos no município de Guarujá pelo método proposto neste trabalho, os clientes já detectados por fraude ou inadimplência pela ELEKTRO neste mesmo município. Estes clientes já detectados se encontravam em dois bancos de dados que foram cedidos pela Empresa e que serão utilizados para as análises a seguir. A primeira análise trata da verificação de clientes fraudulentos e a segunda dos clientes inadimplentes.

#### **5.4.3.1 Clientes Fraudulentos**

Para os clientes fraudulentos, a ELEKTRO disponibilizou um banco de dados com 473 clientes já detectados com algum tipo de fraude para os anos de 2002 e 2003. Procurou-se, então, nos três *clusters* mais significativos quanto a fraude (3, 11 e 15), para toda Guarujá, estes clientes já identificados pela empresa.

Na tabela a seguir tem-se o número de clientes encontrados em cada um dos clusters considerados como de consumidores fraudadores pelo modelo em questão e o total destes clientes.

**Tabela 11: Verificação dos clientes fraudulentos reais encontrados nos clusters de Guarujá**

Cluster	Cientes de Guarujá	Cientes Encontrados
3	5.150	76
11	4.063	112
15	13.791	138
Total	<b>23.004</b>	<b>326</b>

Com estes dados pôde-se obter o percentual de clientes encontrados fazendo a razão entre este número, 326, e o número de clientes detectados pela empresa como fraudulentos, que é 473. Assim, tem-se um percentual de 68,92% de clientes encontrados. Este é um número bastante significativo e que indica que a maior parte dos clientes nestes *clusters* é de consumidores fraudulentos.

#### 5.4.3.2 Clientes Inadimplentes

Já para os clientes inadimplentes foi disponibilizado pela Empresa um banco de dados com 1603 clientes com algum tipo de inadimplência, durante o período de janeiro de 2002 a janeiro de 2004. Da mesma forma que para o banco de dados dos clientes fraudadores, procurou-se nos *clusters* mais significativos quanto a inadimplência (1 e 12) os clientes que estavam inseridos neste banco de dados disponibilizado pela ELEKTRO.

A seguir, tem-se a tabela que mostra o número de clientes encontrados em cada um dos clusters selecionados pelo modelo e o total dos clientes encontrados.

**Tabela 12: Verificação dos clientes inadimplentes reais encontrados nos clusters de Guarujá.**

<b>Cluster</b>	<b>Clientes de Guarujá</b>	<b>Clientes Encontrados</b>
1	6.106	461
12	5.365	228
<b>Total</b>	<b>11.471</b>	<b>689</b>

Foram encontrados 42,98% de clientes, ou seja, encontrou-se 689 clientes dos 1603 que já haviam sido identificados pela Empresa como inadimplentes. Deve-se frisar novamente que este resultado não foi tão bom quanto para o encontrado para os clientes com fraudes devido ao número de amostras dos inadimplentes ser um tanto quanto baixo.