



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO

GESTÃO E AVALIAÇÃO IMOBILIÁRIA

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

MODELOS DE AVALIAÇÃO EM MASSA: REDES NEURONAIAS
ARTIFICIAIS APLICADAS AO SECTOR IMOBILIÁRIO
RESIDENCIAL EM PORTUGAL. ESTUDO DE CASO NA CIDADE
DE LISBOA

JORGE FILIPE MONTEZ VAZ MONTEIRO AMARO

SETEMBRO – 2012



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO

GESTÃO E AVALIAÇÃO IMOBILIÁRIA

TRABALHO FINAL DE MESTRADO

DISSERTAÇÃO

**MODELOS DE AVALIAÇÃO EM MASSA: REDES NEURONAIAS
ARTIFICIAIS APLICADAS AO SECTOR IMOBILIÁRIO
RESIDENCIAL EM PORTUGAL. ESTUDO DE CASO NA CIDADE
DE LISBOA**

JORGE FILIPE MONTEZ VAZ MONTEIRO AMARO

ORIENTAÇÃO:

DOUTOR JOÃO MANUEL DE SOUSA ANDRADE E SILVA

SETEMBRO – 2012

Agradecimentos

Em primeiro lugar e sempre em primeiro lugar, o meu agradecimento à minha mulher e à minha filha pela paciência e incentivo mas sobretudo pelas horas que lhes roubei a minha companhia. É também por Vós.

Ao meu orientador, Prof. João Andrade e Silva pela ajuda e pelo conhecimento transmitido, mas sobretudo por nunca ter deixado de acreditar neste projecto.

À RE/MAX Portugal, em particular ao Dr. Manuel Alvarez, pela participação e disponibilização dos dados que tornaram este trabalho possível.

Por último e não menos importante, o meu agradecimento à EDP, a empresa onde me orgulho de trabalhar, pela oportunidade concedida e pelo incentivo dado na frequência deste curso de mestrado.

Resumo

As redes neuronais artificiais são uma metodologia alternativa aos modelos tradicionais de previsão. A sua utilização tem-se vindo a massificar, sobretudo nas áreas da medicina, finanças, indústria automóvel e, mais recentemente em modelos de avaliação em massa aplicados ao mercado imobiliário.

Este trabalho teve como objectivo fundamental a realização de experiências que utilizassem esta metodologia de previsão não paramétrica (não linear). Foi efectuada uma revisão da principal bibliografia que suporta todo o estudo de caso, com particular enfoque para a investigação que tem vindo a ser desenvolvida internacionalmente, nos domínios da inteligência artificial, aplicada ao sector imobiliário.

No estudo de caso apresentado, foram analisadas 2.013 transacções, que ocorreram entre 2007 e 2012, mediadas pela rede de franchisados RE/MAX, relativas ao segmento residencial na cidade de Lisboa. Com esta base de dados, e depois de encontrada a melhor rede neuronal, foi possível obter um erro médio percentual absoluto (MAPE) na ordem dos 19%, em que para cerca de 67% da amostra foi alcançado um erro de estimação abaixo dos 20%.

Utilizando esta metodologia, também foi observado que a rede neuronal funciona melhor se eliminados os *outliers* da amostra, aumentando a sua precisão. Foi ainda experimentada a introdução de variáveis temporais e de localização, tais como o ano de transacção de um determinado imóvel e a sua idade, ou ainda a freguesia onde está inserido, tendo sido obtidos comportamentos positivos no desempenho das redes.

Para além da originalidade do tema, é de destacar que foram utilizados para este trabalho, valores reais de transacção relativos ao sector residencial em Portugal, tendo sido verificada uma aproximação do comportamento e da tendência do valor de transacção estimado pela rede, aos verificados pelo mercado.

Palavras chave: redes neuronais artificiais, avaliação em massa, mercado imobiliário

Abstract

The artificial neural networks are an alternative approach to traditional forecasting. Its use has been largely used, particularly in the areas of medicine, finance, automotive, and more recently in mass valuation models applied to the real estate market.

This work had as main objective conducting experiments that used this methodology to forecast nonparametric (nonlinear). It was made a review of the primary literature that supports the entire case study, with particular focus on the research that has been developed internationally in the fields of artificial intelligence, applied to the real estate sector.

In the case study were analyzed 2013 transactions that occurred between 2007 and 2012, mediated by the network of franchisees RE / MAX, for the residential segment in Lisbon. With this database, and after found the best neural network, it was possible to obtain an mean absolute percent error (MAPE) of around 19%, in which to approximately 67% of the sample was reached estimation error below 20 %.

Using this methodology, it was also observed that the neural network works better if the outliers are removed from the sample, increasing its accuracy. It also experienced the introduction of temporal variables and location, such as the year of a particular property transaction and its age, or the parish where it is located, having obtained positive behaviors in network performance.

Beyond the originality of the subject, it is worth noting that was used for this work, actual transaction values for the residential sector in Portugal, having been verified an approximation of the behavior and trend of the transaction value estimated by the network, verified by the market .

Keywords: artificial neural networks, mass appraisal, real estate market

Índice geral

Resumo	4
Abstract	5
Índice geral.....	6
1. Introdução	7
2. Revisão da literatura.....	10
2.1. Enquadramento teórico	10
2.2. Modelo de análise	12
3. Metodologia e dados	17
3.1. Metodologia.....	17
3.2. Dados.....	20
4. Estudo de caso	24
4.1. Efeito de <i>outliers</i> e variáveis explicativas.....	25
4.2. Análise de sensibilidade	31
5. Conclusões	34
6. Investigação futura	36
7. Bibliografia	37
8. Anexos.....	39

1. Introdução

As redes neuronais artificiais (RNA) são estudadas neste trabalho, como uma das alternativas aos modelos de avaliação em massa. Desde os anos 90, que tem sido experimentada a sua aplicação por diversos autores, entre os quais se destacam Borst (1991), Do e Grudnitsky (1992), McCluskey (1996) ou Nguyen e Cripps (2001) e mais recentemente por Peterson e Flanagan (2009). Foi identificada uma tese de doutoramento (Canavarro, 2011), sobre o mercado residencial na cidade de Castelo Branco, mas efectuada na Universidade de Córdoba, em Espanha.

Numa altura em que o mercado imobiliário atravessa a pior crise de que há memória, os valores de avaliação e de transacção dos imóveis apresentam diferenças significativas que poderão originar variações patrimoniais importantes para os proprietários, pelo que metodologias que possam contribuir para ajustar de forma expedita o seu valor, serão uma mais-valia.

Em Portugal, o Código do Imposto Municipal sobre Imóveis (CIMI) veio introduzir a partir de 2003, e de acordo com o número 1 do artigo 38º do Decreto-Lei n.º 287/2003 de 12 de Novembro, uma metodologia de avaliação que através da aplicação de um modelo factorial aditivo devidamente identificado, com dados de mercado, de localização e do imóvel avaliado, produzirá o seu respectivo valor patrimonial tributável. Decorre actualmente (2012) a “Avaliação Geral de Imóveis”, cuja Lei n.º 60-A/2011, de 30 de Novembro, que define o regime de avaliação geral dos prédios urbanos que ainda não tenham sido avaliados de acordo com o CIMI. Esta iniciativa do estado português, dando cumprimento ao disposto no memorando de entendimento assinado com a “Troika” em Maio de 2011, terá que ficar concluída até final de 2012, constituindo muito provavelmente a maior operação de avaliação em massa de imóveis de que há memória em Portugal.

Outro fenómeno que está a levantar grandes problemas aos principais bancos a operar em Portugal, sejam públicos ou privados, é a reavaliação interna da carteira de imóveis que constituíram garantia hipotecária para o crédito concedido a empresas e particulares nos

últimos anos. Têm sido desenvolvidos modelos de avaliação em massa suportados em índices imobiliários tendo por base modelos hedónicos, que actualizam o valor de avaliação dos imóveis à data de hoje, informação relevante para efeitos de gestão de risco.

Por outro lado, as normas desenvolvidas pelo *International Valuation Standards Council* (IVSC) prevêm na sua *guidance note* n.º 13 (IVS 2005), que os sistemas de avaliação em massa deverão estar suportados em modelos estatísticos reconhecidos internacionalmente, como por exemplo em modelos de regressão linear múltipla.

As RNA poderão ser aplicadas na actualização em massa de valores de avaliação de imóveis constantes nas carteiras de activos provenientes de incumprimento de crédito hipotecário ou, por exemplo, para aferir o grau de rigor com que os serviços de finanças estão a proceder à reavaliação fiscal de todo o parque imobiliário do País. Também por isso, a utilização desta metodologia poderá ser relevante, não se pretendendo neste trabalho a sua comparação ou aferição de resultados com outras metodologias. Assim, os objectivos deste trabalho são:

- Aplicar o modelo em investigação, utilizando para o efeito uma base de dados de transacções, com preços de venda reais e atributos de oferta residencial na cidade de Lisboa;
- Comparar os resultados obtidos com os reais, analisando o erro de previsão e a forma de como este se comporta.

A crescente compreensão do funcionamento do cérebro humano, levou a que inúmeros investigadores tenham conseguido simular o seu comportamento com maior precisão. Mas será que se conseguirão resultados interessantes no reconhecimento de características dos imóveis e na obtenção de valores de avaliação quando se aplicam metodologias que simulam tal funcionamento? Pela análise da bibliografia consultada na elaboração deste trabalho, poder-se-á afirmar que sim e com algum grau de sucesso, demonstrado pelos resultados evidenciados em inúmeras experiências, as principais explicadas em detalhe mais adiante.

Os resultados esperados deste trabalho são: (i) revisão cuidada e sintética da literatura referente à aplicação de uma metodologia de avaliação em massa com base em redes neuronais artificiais (RNA), assim como a (ii) definição e estimação de uma RNA aplicada ao

segmento residencial no concelho de Lisboa. Pretende-se ainda, (iii) comparar a performance da RNA estimada, com os resultados de outras experiências desenvolvidas. Por último (iv), pretendem-se comparar os resultados obtidos com a utilização de novas variáveis, nomeadamente temporais.

A Secção 2 incide de forma sintética, na temática dos modelos de avaliação em massa, com particular enfoque na revisão bibliográfica que evidencie os resultados das principais experiências efectuadas por outros autores, no que diz respeito à utilização de redes neuronais artificiais e a sua aplicação à avaliação imobiliária.

Na Secção 3, metodologia e dados, será explicada de forma não exaustiva o funcionamento teórico de uma rede neuronal, assim como os algoritmos que a suportam, fundamentalmente as ligações e pesos entre neurónios artificiais e as funções de activação. Pretende-se sobretudo apresentar a fundamentação teórica e o suporte científico que baseará todo o estudo de caso. Ainda nesta secção, será apresentada de forma sumária, a base de dados recolhida, a forma como foi tratada, assim como alguma estatística descritiva e as variáveis chave do estudo de caso.

Na Secção 4, serão analisados os resultados do estudo de caso, com particular enfoque para as diferenças encontradas entre o valor real de transacção e o valor simulado pela rede neuronal, assim como a sua comparação com os resultados das várias experiências efectuadas.

Seguidamente, e na Secção 5, tentar-se-ão evidenciar as principais conclusões deste trabalho científico, reservando para a Secção 6 as principais linhas de orientação para investigação futura. Na Secção 7 será elencada a principal bibliografia consultada e em anexos (Secção 8) ficarão depositadas todas as referências à documentação considerada relevante no trabalho e que ajude à sua melhor compreensão.

2. Revisão da literatura

2.1. Enquadramento teórico

Modelos de avaliação em massa aplicados ao sector imobiliário são aqueles que permitem obter a melhor estimativa de valor de um determinado conjunto de imóveis, de forma sistemática, simultânea e repetível, usando para o efeito um conjunto de metodologias e normas padrão (d'Amato e Kauko, 2008).

Existem duas linhas de investigação e aplicação de modelos de avaliação em massa (d'Amato e Kauko, 2008): As metodologias convencionais, são aquelas que utilizam modelos de regressão linear e que são, tipicamente, as mais usadas em avaliação em massa, com particular destaque para os modelos hedónicos. De entre as que são identificadas como não convencionais, poder-se-ão destacar, as redes neuronais artificiais, os algoritmos genéticos e mais recentemente os modelos “fuzzy”. Na Tabela 1, é efectuado um resumo das várias metodologias de avaliação em massa, assim como dos principais autores e investigadores nesta matéria.

Tabela 1 – Metodologias de avaliação em massa e principais autores

Metodologia	Principais autores
<i>Comparative method</i>	Wiltshaw (1991)
<i>Multiple regression and hedonic prices</i>	Rosen (1974), Meacham (1988), Mark and Goldberg (1988), Fibbens (1995)
<i>Adaptive estimation procedure</i>	Corbone e Longini (1977), Renwick e Flaherty (1996)
<i>Ruled based expert systems</i>	Scott and Gronow (1989), Nawawi and Gronow (1991), Nawawi et al. (1997)
<i>Case based reasoning</i>	Barletta (1991), O’Roarty et al. (1997)
<i>Artificial Neural Networks</i>	Borst (1991), Evans et al. (1992), Do and Grudnitsky (1992), McCluskey and Borst (1997), Worzala et al. (1995), Lenk et al. (1997), Rossini (2000), Nguyen and Cripps (2001), Patterson (2009)
<i>Abductive network models</i>	Borst and McCluskey (1996)
<i>Genetic algorithm</i>	Holland and Goldberg (1994), Cooley (1994)
<i>Fuzzy logic</i>	Lee (2003), Bagnoli and Smith (1998), Sui (1992), Soibelman (2002) e Gonzalez (2005)
<i>Rough set theory</i>	D’Amato (2002, 2003, 2007)
<i>Interviews and surveys</i>	Lentz and Wang (1998), Gartner and Chappelle (1996), Kauko (2002), Fischer (2003)

Fonte: Adaptado McCluskey e Anand, (1999) e Kauko e d’Amato (2008)

Quando se fala em modelos não convencionais, i.e., modelos não lineares ou não paramétricos, estar-se-ão sobretudo a estudar metodologias baseadas em inteligência artificial, que nos últimos anos têm tido inúmeras aplicações, entre as quais as aplicadas ao sector imobiliário residencial. Por outro lado, poder-se-ão encontrar outras metodologias de previsão, onde a regressão linear foi aquela em que maiores contributos foram dados. Na Tabela 2, poder-se-á verificar a ênfase que a maior parte dos investigadores têm dedicado a cada uma das técnicas. A regressão linear é aquela que foi mais vezes utilizada no sector imobiliário, muito por força da sua maior disseminação entre as aplicações que recorrentemente são utilizadas nos computadores e porque goza de um tipo de explicação matemática menos complexa e com uma apresentação menos “*black-box*”.

Como as redes neuronais artificiais são na grande maioria dos casos, a metodologia mais usada ligada ao domínio da inteligência artificial, os seus resultados de previsão são muitas das vezes comparados com os modelos de regressão, tendo tido em alguns casos, menor erro de previsão, como por exemplo o apontado por Borst (1991), Do e Grudnitsky (1992) ou McCluskey (1997) e mais recentemente por Nguyen e Cripps (2001), ou, pelo contrário, o apontado por Worzala, Lenk e Silva (1995) que não reconhecem às RNA o mérito de serem consideradas mais fiáveis que os MRL (Modelos de Regressão Linear).

Tabela 2 – Investigação para a comunidade científica por metodologia em percentagem

Metodologia	Contribuição para a comunidade científica desde 1970 (%)
<i>Multiple regression analysis</i>	35
<i>Models including spatial control</i>	10
<i>Model free regression models (non-parametric regression)</i>	15
<i>Artificial neural networks and genetic algorithm</i>	15
<i>Qualitative intelligent methods and rough set theory</i>	10
<i>Model based on judgmental data</i>	15

Fonte: Kauko e d’Amato (2008)

Em 2011, Maria Canavarro Teixeira, no âmbito da sua tese de doutoramento na Universidade de Córdoba, efectuou uma comparação do erro de previsão entre um modelo de regressão e

uma rede neuronal, efectuada para imóveis residenciais na cidade de Castelo Branco (Canavarro, 2011).

A aplicação de redes neuronais a vários sectores de actividade é largamente conhecida, sendo o seu maior contributo dado pela capacidade invulgar que tem de reconhecer padrões e de calcular de forma iterativa, o peso relativo que cada variável explicativa assume, de maneira a que se obtenha o menor erro na estimativa.

A primeira vez que foi simulado o funcionamento de um neurónio foi em 1943, quando McCulloch e Pitts criaram o primeiro modelo lógico-matemático que tentava simular o seu funcionamento (Cortez, 2000). Por outro lado, quando em 1958 vê publicado o seu livro intitulado *"The perceptron: a probabilistic model for information storage and organization in the brain"*, Rosenblatt longe estaria de imaginar que a sua investigação pudesse conduzir a avanços tão significativos no domínio da inteligência artificial. Em 1962, volta a publicar novo livro intitulado *"Principles of neurodynamics: Perceptrons and the theory of brain mechanisms"*, que defendia que o cérebro humano era constituído por nós de decisão ou de processamento, os neurónios, e que através de um conjunto repetido de estímulos ponderados e aleatórios (treino), conseguiria aprender uma determinada tarefa. Ao algoritmo que resulta da interpretação desta função não linear, Frank Rosenblatt deu o nome de *Perceptron*, ou na literatura anglo-saxónica mais recente, *Single Layer Perceptron*. (Rosenblatt, 1958, 1962)

2.2. Modelo de análise

O primeiro autor a conseguir simular com algum sucesso o comportamento do cérebro humano na avaliação de um conjunto de imóveis foi Richard Borst em 1991. Borst (1991), utilizou duas amostras para treinar a sua rede, a primeira com 22 e a segunda com 20 observações, tendo obtido um MAPE – Mean Absolute Percentage Error¹ de 8,7% e 12,4% respectivamente. Para a sua experiência utilizou 4 cenários de simulação: (1) Amostra de observações com tratamento estatístico e com transformação de dados, obtendo um resultado de 8,9% de erro percentual médio absoluto (MAPE); (2) Amostra de observações com

¹ Tradução livre para português: erro médio percentual absoluto

tratamento estatístico e sem transformação de dados, obtendo um resultado de 10,8%; (3) Amostra de observações sem tratamento estatístico e sem transformação de dados, obtendo um resultado de 12,4%; (4) Amostra de observações com tratamento estatístico e sem transformação de dados, com 16 intervalos de valor de transacção (*outputs*) obtendo um resultado de 8,7%; Podendo o MAPE (Mean Absolute Percentage Error) ser definido como:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_i - P_i}{A_i} \right|$$

n – número de observações; A_i – valor de avaliação ou transacção efectivo; P_i – Valor estimado pela rede neuronal

Este trabalho identificou factores que conferiram ou retiraram consistência à experiência, nomeadamente se a amostra é tratada estatisticamente (retirando *outliers*) ou se as variáveis são transformadas ou não. Paralelamente a este facto, defendeu que as redes neuronais têm uma melhor performance que os modelos de regressão.

Evans, James e Collins (1991) testaram o efeito dos *outliers* nos resultados obtidos pela rede neuronal, tendo concluído que os mesmos variavam cerca de 2% em termos de erro observado.

Em Dezembro de 1992, Do e Grudnitsky apresentaram as conclusões da sua experiência num artigo intitulado de “*A Neural Network Approach to Residential Property Appraisal*”, sendo ainda hoje uma das principais referências na utilização de redes neuronais na avaliação imobiliária. Segundo os autores, as redes neuronais são particularmente eficazes na avaliação de imóveis residenciais, num mercado caracterizado pela informação sobre as transacções muitas das vezes complexa, imprecisa e incompleta. A sua não linearidade, suporta bem muitas das ineficiências quer do mercado, quer dos *inputs* que são utilizados. A sua experiência foi efectuada tendo por base 105 transacções de casas vendidas entre Janeiro e Junho de 1991, de uma parte da cidade de San Diego nos EUA. A variável preço de transacção, ou seja, o que se pretende estimar, foi calculada através de uma rede neuronal de 8 entradas com uma camada escondida de 3 nodos. As variáveis explicativas identificadas foram a idade do imóvel, número de casas de banho, número de quartos, área e número de

pisos da casa, número de garagens, número de lareiras e tamanho do lote de terreno em m². Os autores usaram 58 observações para treinar a rede, testando-a com as outras 47 observações. Com estes parâmetros de estimação, obtiveram um MAPE de 6,9%. Utilizaram nas suas experiências um *software* que já tinha sido anteriormente utilizado por Eberhart e Dobbins (Do e Grudnitsky, 1992).

Também em 1992, Tay e Ho experimentaram a utilização de redes neuronais para estimar o preço de imóveis em Singapura, mas usando para o efeito, uma amostra maior que os seus antecessores (cerca de 1.055 observações). Constataram que a utilização de amostras maiores conduzia a menores erros de previsão, 3,9% com inclusão de *outliers* e 3,7% retirando os mesmos (Tay e Ho, 1992).

Worzala, Lenk e Silva, três investigadoras na área de avaliação em massa, elaboraram 2 artigos que se tornaram também referências nesta metodologia. Primeiro em 1995, basearam-se nas experiências de Borst (1991), Evans, James e Collins (1991) e Do e Grudnitsky (1992) chegando à conclusão que, em alguns casos, mesmo que marginalmente, os modelos de regressão obtêm melhores resultados (erros menores). A sua experiência baseou-se em três casos: (1) a totalidade da base de dados; (2) replicaram a experiência de Do e Grudnitsky e (3) utilizaram uma amostra mais homogénea e mais pequena de dados, trabalhados estatisticamente (Worzala, Lenk e Silva, 1995).

Em 1997, as autoras voltam a trabalhar sobre o tema fazendo uma experiência sobre modelos de avaliação em massa, em que comparam novamente os modelos de regressão linear com redes neuronais, com a particularidade de experimentarem vários *softwares*, provando não só, que existem resultados diferentes nas redes neuronais se a experiência se fizer com aplicações diferentes usando sempre a mesma amostra, assim como chegam também à conclusão, de que estes modelos de avaliação poderão até nem serem a melhor opção para a avaliação de imóveis residenciais. Fazem contudo uma nova avaliação no que à superioridade das redes neuronais relativamente aos modelos de regressão diz respeito, constatando que em determinadas circunstâncias elas estimam melhor, sobretudo se na condução da investigação se usar uma base de dados de amostra sem *outliers* (Lenk, Worzala

e Silva, 1997). Na Tabela 3, são apresentados os resultados destas experiências, que no decorrer do trabalho de investigação foram consideradas relevantes.

Tal como Evans, James e Collins em 1991, McCluskey (1997) concluiu que as redes neuronais obtiveram melhores resultados que os modelos tradicionais de regressão linear, numa experiência realizada na Irlanda do Norte e com base em 516 observações.

Tabela 3 – Resultados obtidos na modelação de dados por RNA por autor

Autores	Ano	País de teste	Amostra (n.º imóveis)	Resultado (MAPE)
Borst	1991	Estados Unidos América	Amostra de treino 1: 22 Amostra de treino 2: 20	8,7% (1ª amostra) 12,4% (2ª amostra)
Evans, James e Collins	1991	Inglaterra e País de Gales	34	13,48% (1º teste) 5,03% (2º teste)
Do e Grudnitsky	1992	Estados Unidos América	Treino: 58 Teste: 47	6,9%
Tay e Ho	1992	Singapura	Treino: 833 Teste: 222	3,9%
Worzala, Lenk e Silva	1995	Estados Unidos América	Caso 1 Treino: 217; Teste: 77 Caso 2 Treino: 137; Teste: 43 Caso 3 Treino: 83; Teste: 29	Caso 1: 13,2% 14,4% Caso 2: 10% 13,1% Caso 3: 11,6% 11,1%
McCluskey	1997	Irlanda do Norte	Treino: 378 Teste: 138	15,7% (1º teste) 7,75% (2º teste)
Worzala, Lenk e Silva	1997	Estados Unidos América	288 Observações (17 outliers), 204 observações de treino	18% de MAPE acima dos 15%
McGreal, Adair, McBurney e Patterson	1998	Irlanda do Norte	1.026 Observações	Apenas 80% das observações estavam abaixo dos 15% de MAPE
Nguyen e Cripps	2001	Estados Unidos da América	3.906 Observações (18 datasets de múltiplos de 200 observações crescentes)	À medida que a amostra de teste vai aumentando os resultados vão melhorando. Mínimo de 7.1% (T18;V18)
Peterson e Flanagan	2009	Estados Unidos América	46.467 Observações entre 1999-2005	Entre os 20% e os 25% em média
Canavarro	2011	Portugal	200 Observações	8,77%

Fonte: Lai e Fischer (2005) e adaptação própria

Outro dos trabalhos de referência relacionado com as redes neuronais artificiais, foi o desenvolvido por McGreal, Adair, McBurney e Patterson apresentado em 1998 à comunidade científica. Segundo os autores (McGreal *et al*, 1998), os três pontos-chave de uma rede neuronal são: (i) os nodos e as ligações entre eles, (ii) o algoritmo para o treino da rede e por fim, (iii) a metodologia de interpretação de resultados na fase de teste. À semelhança do estudo de caso apresentado adiante, estes autores também usaram uma base de dados de transacções. Usaram 1.026 observações de compra e venda para o mercado de Belfast, na Irlanda do Norte. Concluíram que 80% do valor dos imóveis conseguiam estar abaixo dos 15% de MAPE.

Em 2001, Nghiep Nguyen e Al Cripps, realizaram uma experiência (Nguyen e Cripps, 2001) com vários tamanhos de amostra entre as 306 observações e as 3.906 observações. Com a construção de vários *datasets*, os autores demonstraram que à medida que a amostra de treino vai sendo maior, o MAPE vai diminuindo, assumindo-se os modelos baseados em redes neuronais superiores aos de regressão.

De entre as experiências mais recentes, Peterson e Flanagan, apresentaram em 2009, os resultados da sua investigação (Peterson e Flanagan, 2009), utilizando uma base dados de 46.497 observações relativas a uma localidade nos Estados Unidos da América. Concluíram que as redes neuronais têm uma melhor performance de previsão que os modelos de regressão, tratam melhor os *outliers* e conseguem extrapolar melhor os valores de previsão com uma amostra mais volátil. Defenderam ainda, que redes *multi layer* têm melhor performance que *single layer*. Concluem também que as redes neuronais, são mais apropriadas na construção de um modelo hedónico, pois lidam melhor com muitas variáveis *dummy*.

Canavarro, apresentou em 2011 os resultados da sua tese de doutoramento, tendo por base uma amostra de 200 observações de transacções de apartamentos entre 2005 e 2009, na cidade de Castelo Branco em Portugal. Faz, à semelhança de outros autores, a comparação da performance de previsão das redes neuronais com os modelos de regressão linear, chegando a um MAPE de 8,77% para uma metodologia e a 10,97% para a outra, respectivamente

(Canavarro, 2011). Em suma, e de acordo com o relato das várias experiências em torno da utilização das redes neuronais, apresentam-se seguidamente os seus aspectos mais relevantes retirados da revisão bibliográfica:

1. A utilização de redes neuronais como metodologia não paramétrica de estimação pode ser utilizada, de per si ou combinada com outra, seja linear ou não linear (e.g. *Fuzzy* ou modelos Hedónicos);
2. Muitos autores testaram os limites da sua utilização, combinando vários factores que poderiam alterar os resultados obtidos por esta metodologia, de entre os quais se destacam: dimensão da amostra, utilização de observações de treino e de teste (com diferentes tamanhos de amostras de treino e de teste), utilização/identificação de *outliers*, introdução/redução de variáveis explicativas do modelo (semelhante a uma metodologia *stepwise*), utilização de observações de dimensão plurianual, utilização de redes multicamada ou de camada única, utilização de variáveis alteradas/modificadas ou ainda os resultados obtidos por diferentes *softwares* para a mesma amostra.

3. Metodologia e dados

3.1. Metodologia

As redes neuronais artificiais pretendem replicar o funcionamento do cérebro e da sua capacidade de aprender.

As mais conhecidas e utilizadas pelos investigadores são as redes multicamada em que o estudo de caso apresentado adiante recairá, podendo ter uma organização e fluxo de informação do tipo “*feed-forward*”, “*feed-back*” ou “*competitive*” (Kauko e d’Amato, 2008):

As redes neuronais do tipo “*feed-forward*” permitem que a informação apenas circule num único sentido. Isto quer dizer que o resultado de cada nodo, não influenciará o resultado de outro nodo da mesma camada. São um tipo de rede muito usada em reconhecimento de padrões.

As redes do tipo “*feed-back*” podem ter os sinais a percorrer os nodos em ambos os sentidos.

É um tipo de organização da rede que procura um ponto de equilíbrio em permanência.

Existem ainda as denominadas “*competitive networks*” que são um tipo de rede com algoritmo de aprendizagem não supervisionada, ou seja, todos os elementos da rede competem entre eles pelo direito de dar a resposta ao próximo nodo, que conduzirá à acção seguinte.

Para além do tipo de organização de rede, existem ainda vários tipos de algoritmos de aprendizagem da rede, que poderão ser do tipo supervisionado e não supervisionado. Se o método de aprendizagem for supervisionado, então poderemos ter ainda os modelos de aprendizagem através de retro propagação ou ainda os de correcção (Rojas, 1996).

Os modelos de redes neuronais mais usados na investigação ligada à avaliação imobiliária, são as do tipo “*feed-forward*”, cujo treino é efectuado com um algoritmo de retro propagação (*backpropagation algorithm*), em que o objectivo é o de minimizar a soma do quadrado dos erros, através da combinação dos pesos nos nodos através de um processo iterativo, que é despoletado a partir de um determinado limiar (*threshold*) de uma função de activação de decisão (Rojas, 1996). As funções de activação são apresentadas em detalhe no Anexo 1.

Tal como é percebido, toda esta teoria computacional é embebida de um elevado nível de aprendizagem da rede, onde o treino terá papel primordial. Assim, e para um conjunto alargado de observações, importa perceber quantas serão de teste e quantas servirão para o treino (aprendizagem) da rede.

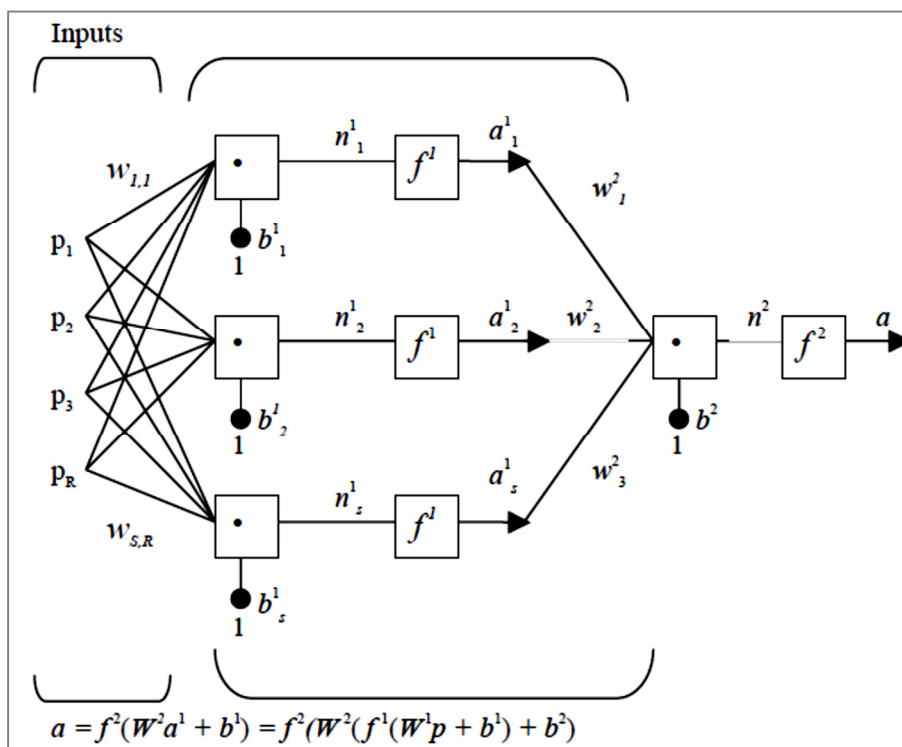
Segundo Nguyen e Cripps (2001), poderão existir muitas decisões metodológicas que importam acautelar, como por exemplo, quantas camadas terá a rede ou quantos nodos terá cada camada. Defendem ainda que o treino e o número de nodos na camada intermédia afectam largamente a capacidade de aprendizagem, em que deverá ser acautelando o “*overtraining*”.

O desenho da rede neuronal que servirá de modelo para o que será efectuado adiante, no estudo de caso, poderá ser sumariamente explicado, com o auxílio da Figura 1: uma rede neuronal, com uma camada oculta, treinada pela retro propagação do erro (rede supervisionada), ajustando os pesos das ligações aos nodos, minimizando o soma do quadrado dos erros obtido entre o valor a explicar e o valor estimado (Peterson e Flanagan,

2009). Com a letra “*p*” são identificados os diversos atributos da variável que se pretende explicar (no estudo de caso foi o “valor de transacção”). A letra “*w*” representa os pesos que cada atributo tem na rede neuronal e a letra “*b*” representa o “bias” ou termo constante da equação, em tudo semelhante ao “*intercept*” de um modelo de regressão. O *output* de cada nodo (neurónio artificial), é a soma ponderada dos vectores de *input*. Depois de pesados os *inputs*, estes terão que ser transformados por uma função de activação.

O Anexo 1 apresenta as principais funções de activação que depois de transformarem a variável a encaminham para a camada seguinte. A função de activação é importante pois define a intensidade do sinal que é transmitido dos nodos da primeira camada para a segunda.

Figura 1 – Esquema de uma rede neuronal do tipo “*feed-forward*”



Fonte: Peterson e Flanagan (2009)

De acordo com os mesmos autores (Peterson e Flanagan, 2009, referindo, Haykin, 2003), a camada oculta serve fundamentalmente para detectar características específicas de cada observação.

Com este processo, os pesos da rede neuronal são treinados de forma iterativa, até que seja minimizada a função da soma dos quadrados dos erros, comparando até um determinado nível de aceitação o valor observado com o valor estimado pela rede. A função mencionada poderá ser dada pela seguinte fórmula (Peterson e Flanagan, 2009):

$$F = w \min \sum (t - a)^2$$

Sendo “ t ”, o valor observado e “ a ” o valor estimado. Para que se minimize o erro, é necessário treinar a rede com recurso a um algoritmo, em que por norma é utilizado o gradiente descendente, fazendo com que o erro ponderado altere o peso dos nodos das camadas anteriores da seguinte forma:

$$w_{i,j}^s(k+1) = w_{i,j}^s(k) - \alpha \frac{\partial F}{\partial w_{i,j}^s}$$

Representando “ α ”, a taxa de aprendizagem da rede, “ ∂F ”, a derivada da função de activação e “ $\partial w_{i,j}^s$ ” a derivada do peso ponderado na camada “ s ”.

A todo este processo deu-se o nome de “*backpropagation*”, que serve de base à maior parte da investigação em redes neuronais, independentemente da sua utilização.

3.2. Dados

A base de dados utilizada para a realização deste trabalho, foi disponibilizada pelo “*Master Franchise*” da RE/MAX Portugal, mediante acordo de confidencialidade (Anexo 2). Os dados foram extraídos e disponibilizados no dia 29 de Fevereiro de 2012 e incluíam 30.550 observações, abrangendo transacções em todos os concelhos do distrito de Lisboa em que foram registados todos os contratos de compra e venda e de arrendamento, firmados entre Fevereiro de 2001 e Fevereiro de 2012 e que estivessem carregados informaticamente na Base de Dados Própria da RE/MAX Portugal. A base de dados original contou com 19 campos de caracterização da amostra (Anexo 3).

3.2.1. Metodologia de tratamento e depuração da base de dados

Sobre a base de dados original foi efectuada uma primeira segmentação de acordo com o objecto de estudo (concelho de Lisboa e de negócio de compra e venda). Seguidamente foi depurada (com eliminação de erros) e harmonizada de maneira a permitir o seu tratamento estatístico. Daqui resultou a base de dados final para aplicação da metodologia.

3.2.2. Análise de dados e desenho da amostra

A primeira operação de segmentação efectuada, foi seleccionar só o concelho de Lisboa e os imóveis vendidos. A base de dados inicial continha 19.365 observações correspondentes a transacções (de compra e venda ou arrendamento) realizadas fora do concelho de Lisboa. Com esta operação reduziu-se a base de dados global para cerca de 1/3, ficando com 11.185 observações. Seguidamente foram também isoladas apenas as transacções que correspondiam a contratos de compra e venda, visto o presente trabalho incidir apenas neste segmento. Daqui resultaram 5.646 observações. Estando este trabalho focado no sector imobiliário residencial, foram também identificadas e eliminadas todas as observações diferentes de “Moradia”, “Apartamento”, “Duplex”, “Triplex” e “Prédio”. Com mais esta operação registaram-se 5.223. Foram também retiradas as observações cujos resultados constantes nos campos “Área” e “Valor de venda” fossem “0” ou “NULL”. Como resultado, chegaram-se às 5.164 observações. Como o objecto deste trabalho é um estudo de caso sobre o segmento residencial, e em virtude da maior parte das observações serem sobre transacções de apartamentos, optou-se por incidir a investigação apenas sobre esta categoria de oferta residencial, contando agora com 4.873 observações, incluindo “Apartamento”, “Duplex” e “Triplex”.

A partir daqui deu-se início ao trabalho de depuração da base de dados, que previa a construção de campos auxiliares ou a limpeza de informação não normalizada e cujo sequente tratamento estatístico fosse afectado. Para o efeito foram utilizadas duas metodologias, sobretudo de captação de inconformidades da amostra relativamente ao valor mínimo por m² que seria admissível (foi construída uma nova variável: eur/m²). Foram apuradas 48

inconsistências entre campos, sobretudo naqueles, cujos valores de idade, estado, de transacção, área ou o quociente entre os dois anteriores, não fosse plausível. Por outro lado, a Portaria n.º 291/2011, que define o valor mínimo de construção por m², incluindo terreno, serve de base ao cálculo de uma renda condicionada em Lisboa, cifrando-se em 767,42 Euros para o ano de 2012. Tendo por base este valor foram eliminadas mais 91 observações.

Depois de todo o tratamento de segmentação e de depuração da base de dados original, foi considerada como amostra base para o trabalho que se seguiu, cerca de 4.734 observações, incluído *outliers*. Como a variável idade é determinante para a correcta análise dos resultados da simulação, só foram consideradas 2.013 observações. As remanescentes não tinham a variável idade do imóvel (originalmente “ano de construção”) devidamente preenchida. A variável “idade” foi construída tendo por base a diferença entre o ano de transacção e ano de construção.

3.2.3. Análise e consistência da amostra

A amostra de 2.013 observações poderá ser resumidamente caracterizada de acordo com a Tabela 4:

Tabela 4 - Estatística descritiva da amostra total (com outliers) - variáveis quantitativas

Variável	Unid.	Min	Max	Média	Desvio Padrão
Valor de Transacção	Euros	29.980	975.000	197.107,01	115.524,533
Área	m ²	16	412	103,75	45,872
Idade	Anos	0	212	34,68	28,799
N.º de quartos	#	0	8	2,46	1,095
Estacionamento	S/N	0	1	0,34	0,474

Fonte: RE/MAX Portugal, análise do autor; Output: SPSS v. 19

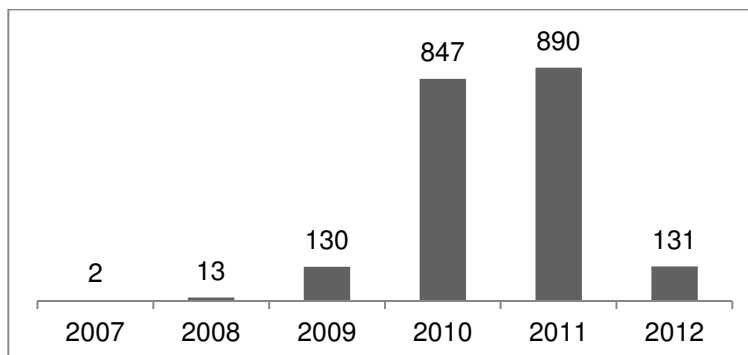
Os *outliers* identificados, são os que dizem respeito à variável “valor de transacção”, quando analisada isoladamente. Foram considerados *outliers*, as observações cujo valor estivesse contido no intervalo inferior ou superior, dado, respectivamente, pelas seguintes fórmulas:

- Limite inferior = 1º Quartil – 1,5 X AIQ² (outlier inferior)
- Limite superior = 3º Quartil + 1,5 X AIQ (outlier superior)

² AIQ: Amplitude interquartil = 3º Quartil da amostra – 1º Quartil da amostra

Como resultado, foram identificados 77 *outliers* superiores (valor de transacção > 452.500 euros). A distribuição das observações ao longo do tempo é a constante na Figura seguinte:

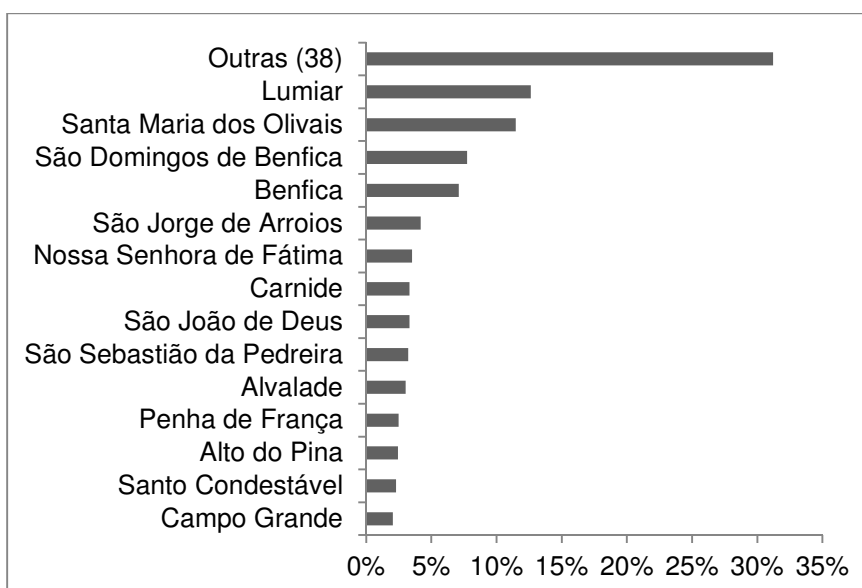
Figura 2 – Distribuição de observações ao longo do tempo



Fonte: RE/MAX Portugal, análise do autor; Output: Microsoft Excel

Por fim, importou caracterizar a amostra no que diz respeito à sua divisão geográfica, i.e., por freguesia de concelho de Lisboa. Foram verificadas transacções para 52 freguesias de Lisboa (de 54), sendo que as mais representativas são as freguesias de Santa Maria dos Olivais, Lumiar e São Domingos de Benfica com cerca de 32% do total de transacções (Figura 3).

Figura 3 – Divisão geográfica da amostra em percentagem



Fonte: RE/MAX Portugal, análise do autor; Output: Microsoft Excel

Associado à localização do apartamento (freguesia), foram introduzidas duas variáveis auxiliares, que ajudaram a explicar o efeito da freguesia da amostra. Foram adicionados, os coeficientes de localização máxima e mínimo, utilizados no cálculo do valor patrimonial tributário de acordo com o CIMI (Anexo 4), de maneira a atribuir uma importância diferente a cada freguesia em função do coeficiente (que influencia o valor patrimonial tributável).

3.2.4. Definição das variáveis chave

À semelhança dos casos de estudo analisados no decurso deste trabalho, as variáveis chave da amostra são, para além do valor de transacção (variável dependente): a área, o número de quartos, idade do imóvel, assim como a freguesia com os respectivos coeficientes de localização. Adicionalmente serão estudados os efeitos do ano de transacção, da classe de área e da zona onde está o apartamento e que impacto tiveram no resultado do MAPE. A classe de área identifica em que intervalo de área está inserido o imóvel e a zona resulta do agrupamento de freguesias. Ambas serviram para melhorar o reconhecimento de padrões da rede. Foi ainda testado o efeito da introdução da variável “elevador” (existência ou não).

4. Estudo de caso

No estudo de caso foram efectuadas duas experiências, resultantes da aplicação de redes neuronais artificiais:

(4.1) O efeito que os *outliers* e a introdução de variáveis explicativas têm no erro percentual médio absoluto (MAPE). Tentou-se replicar o que Nguyen e Cripps (2001) e Worzala, Lenk e Silva (1997) fizeram, utilizando para o efeito a mesma amostra de treino e de teste;

(4.2) Utilizar os algoritmos que serviram para o cálculo dos resultados da melhor rede, para verificar se esta consegue captar os efeitos de alterações das variáveis explicativas. Foi estudada a sensibilidade que a rede tem ao ano de transacção, com a identificação da tendência de valorização ou desvalorização de um apartamento tipo.

Para a realização destas experiências, foi utilizado o IBM SPSS, versão 19 quer para a

construção das várias redes neuronais, quer para a análise dos seus resultados. Este *software*, utiliza para a simulação de uma rede neuronal, uma arquitectura de rede multi-camada do tipo “*feed-forward*”, com um algoritmo de treino “*backpropagation*” e três tipos de função de activação. Não se efectuaram nenhum tipo de selecção de desenho da rede, função de activação ou de número de nodos, deixando que fosse sempre a aplicação a seleccionar automaticamente o desenho e principais características da rede. Foram usadas 30% de observações para teste e 70% para treino. Tal como identificado por Lenk, Worzala e Silva (1995), as amostras foram trabalhadas estatisticamente de maneira a que se pudessem obter melhores resultados. O objectivo foi o de tentar obter o melhor desempenho possível com a base de dados disponibilizada, tentando assim tirar o melhor partido na estimação em massa de valores de transacção.

4.1. Efeito de *outliers* e variáveis explicativas

No decurso deste trabalho, estudou-se o efeito que os *outliers* tinham no comportamento da rede, o impacto que o número de camadas ocultas e nodos por camada tinham no erro, assim como, se o tamanho da amostra de treino ou de teste tinha algum tipo de relação de causa – efeito com os resultados finais.

As redes neuronais funcionam por reconhecimento de padrões, ou seja, se a amostra contiver um conjunto de “marcadores” que permitam que a rede consiga identificá-los, os resultados são melhores. Foi assim testada a influência que os *outliers* tiveram nos resultados finais, tendo-se verificado que a rede conseguiu melhores resultados sem a utilização de *outliers*. Paralelamente foi ainda testado o comportamento da rede quando se introduziram ou retiraram variáveis explicativas ao modelo. Para tal foram construídos 9 cenários de teste e analisados os seus resultados. A amostra de treino foi constituída previamente (e usada sempre a mesma em todos os cenários), em que 70% das observações foram seleccionadas e marcadas de forma aleatória. O remanescente (30%), serviu para teste da rede. Para a construção da amostra de treino e de teste foram apenas levados em consideração dois pressupostos que influenciaram bastante o seu desempenho:

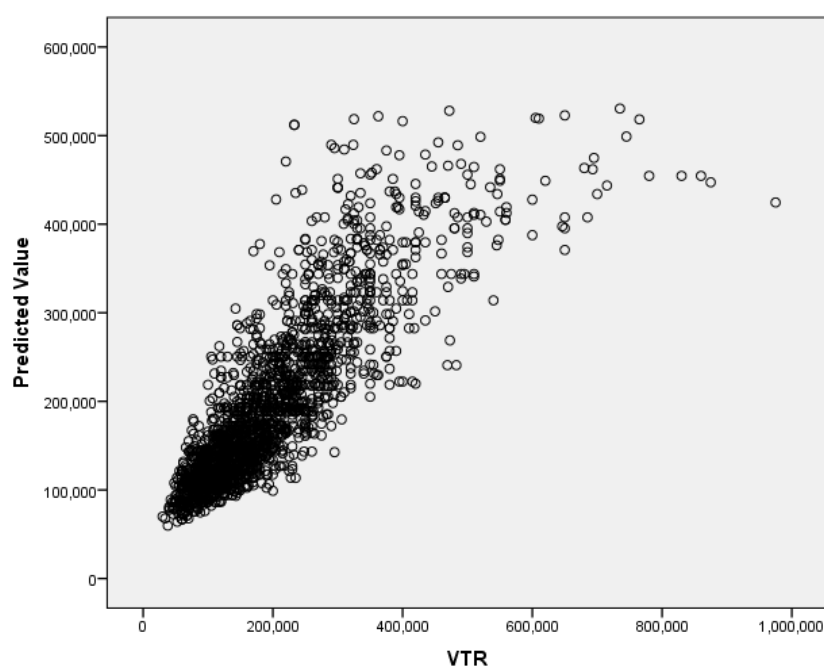
- O intervalo de valor de transacção na amostra de teste estava contido na amostra de treino;
- Todas as freguesias tinham a mesma proporção aproximada de observações, quer na amostra de treino (70%), quer na amostra de teste (30%).

Foram usadas todas as observações disponíveis na amostra (2.013), e incluídos inicialmente os 77 *outliers* identificados. Os cenários construídos para a execução desta primeira experiência, tiveram como objectivo a introdução ou remoção de variáveis explicativas que pudessem originar a melhor estimacção possível do valor de transacção de um apartamento na cidade de Lisboa. Assim, e para explicar a variável dependente (valor de transacção) foram utilizadas as seguintes variáveis explicativas, construindo-se 9 cenários:

1. Área (variável quantitativa) + número de quartos (variável qualitativa);

Conforme exposto na Figura 4, conseguiu-se perceber que a rede era sensível à área do imóvel e ao número de quartos, apresentando contudo um défice de reconhecimento de observações que estivessem mais próximas dos extremos da amostra. Para um valor acima dos 500.000 euros a rede não conseguiu estimar com precisão:

Figura 4 – Desvio entre o valor estimado e o valor observado (euros), 1º cenário



Fonte: RE/MAX Portugal, análise do autor; Output: SPSS v. 19

Esta primeira rede tinha uma camada oculta com 8 nodos (com 10 nodos na camada inicial - excluindo o termo independente), utilizando como função de activação a tangente hiperbólica.

2. Área + número de quartos + freguesia (variável qualitativa);
3. Área + número de quartos + freguesia + coeficientes de localização máximos e mínimos (variáveis qualitativas);
4. Área + número de quartos + freguesia + coeficientes de localização máximos e mínimos + zonas (variáveis qualitativas); As zonas foram definidas pelo agrupamento de freguesias de acordo com a definição do INE, conforme Anexo 5.
5. Área + número de quartos + freguesia + coeficientes de localização máximos e mínimos + zonas + ano da transacção (variável qualitativa);
6. Área + número de quartos + freguesia + coeficientes de localização máximos e mínimos + zonas + ano da transacção + classe de área (variável qualitativa); As classes de área, foram definidas de acordo com os intervalos identificados na Tabela 5, sobretudo para que ajudasse a rede a distinguir apartamentos com áreas típicas daqueles que têm áreas atípicas:

Tabela 5 – Classes de área em m² (n.º observações)

Classe Área	Limite	Limite	Frequência
	Inferior	Superior	
1		79	689
2	79	158	1.071
3	158	238	231
4	238	317	20
5	317		2

Fonte: RE/MAX Portugal, análise do autor; Output: Microsoft Excel

7. Área + número de quartos + freguesia + coeficientes de localização máximos e mínimos + zonas + ano da transacção + classe de área + idade do apartamento (variável quantitativa);
8. Foram retiradas do cenário 8 as variáveis “Ano de Transacção”, “Classe de área” e “Zona”, para estudar o comportamento da rede;

9. Por último foi adicionada uma variável de caracterização, p.ex., existência de “elevador” no prédio, não se revelando determinante na diferenciação do valor estimado de transacção.

Estes 9 cenários de teste foram construídos para que se encontrasse a melhor rede neuronal possível com ou sem *outliers*, de maneira a minimizar o MAPE. Foi ainda apurada a percentagem de observações cujo MAPE se apresentou inferior ou igual a 10%. A Tabela 6, sintetiza os resultados desta experiência:

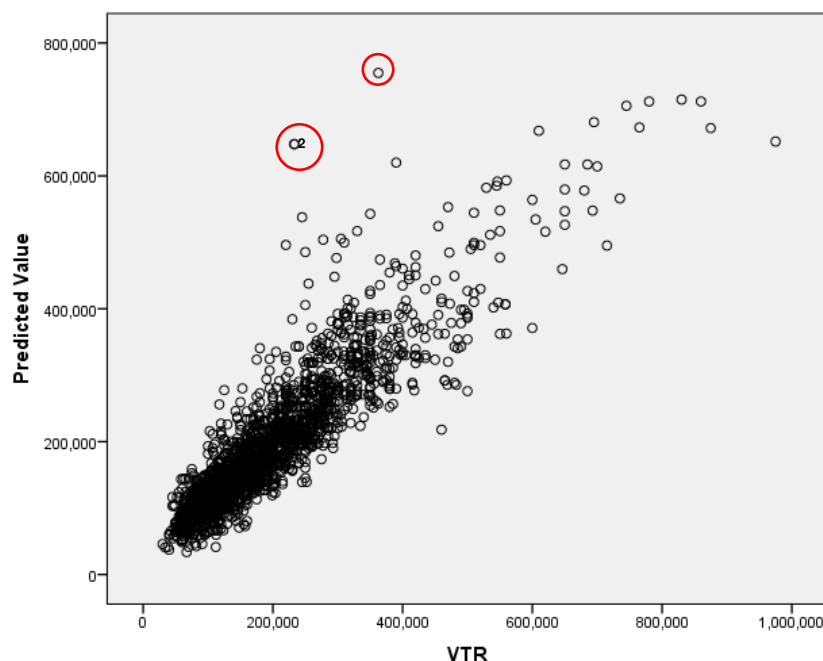
Tabela 6 – Resumo dos resultados das experiências por cenário

Cenário	Com outliers			Sem outliers		
	Média	MAPE	% ≤ 10%	Média	MAPE	% ≤ 10%
1	197.340	24,30%	28,50%	181.775	23,00%	30,10%
2	198.347	23,20%	31,80%	183.357	21,40%	34,20%
3	193.337	23,20%	30,20%	183.000	21,80%	33,90%
4	197.392	21,60%	33,70%	182.085	20,70%	34,80%
5	197.441	22,90%	31,40%	183.836	21,00%	35,90%
6	198.994	23,80%	31,20%	180.969	20,80%	33,20%
7	201.958	21,50%	35,50%	182.437	18,60%	39,70%
8	196.802	19,70%	37,00%	183.803	19,60%	38,00%
9	195.813	20,20%	35,20%	183.880	19,50%	36,90%

Fonte: RE/MAX Portugal, análise do autor; Output: Microsoft Excel

Através da Figura 5, verificou-se que com os pressupostos atrás definidos, a rede consegue, utilizando o cenário 8 (com *outliers*), reconhecer as observações atípicas com algumas limitações. A melhor rede concebida pela aplicação, ou seja, a que conduziu a erros de estimação menores, foi a que estava suportada no cenário 7, sem *outliers*. Os *outliers* têm influência no desempenho da rede, contribuindo negativamente para o erro de estimação em sensivelmente 3%, valor próximo do que Evans, James e Collins (1991) tinham alcançado.

Figura 5 – Relação entre o valor estimado e o valor observado (euros), cenário 8 (*com outliers*)

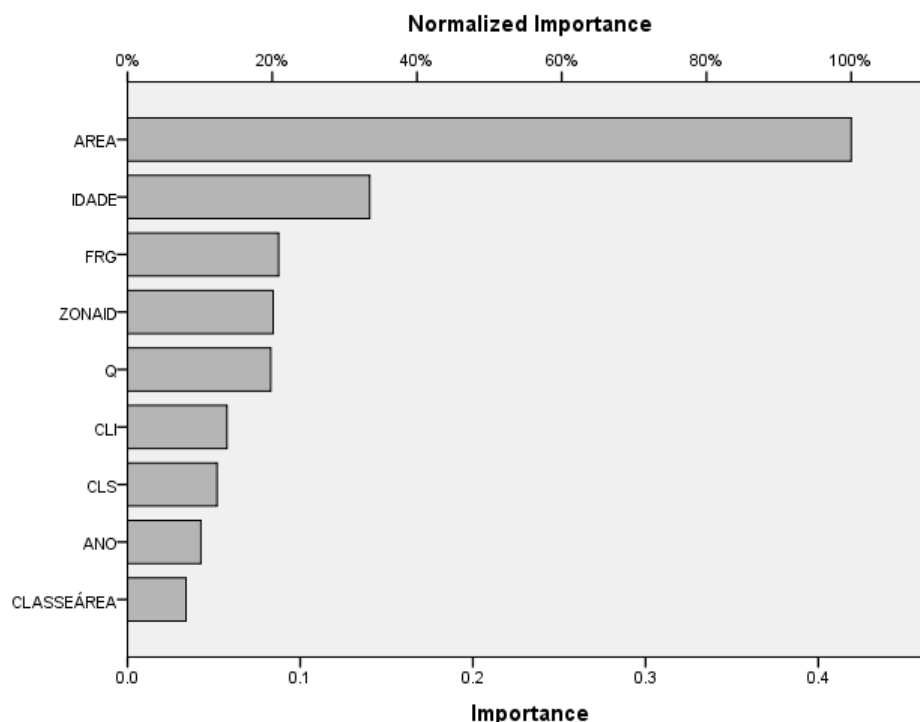


Fonte: RE/MAX Portugal, análise do autor; Output: SPSS v. 19

As 3 observações assinaladas na Figura 5, corresponderam a observações que provavelmente foram mal inseridas na base de dados, ou para as quais as variáveis do modelo não têm capacidade explicativa, ou cujo valor por m² esteja muito desfasado da amostra. Não são *outliers*, mas influenciam também o resultado final.

A Figura 6, ilustra a capacidade explicativa das variáveis da melhor rede, suportada nos pressupostos do cenário 7, sem *outliers*. Neste caso, também se verificou que as variáveis “ano de transacção”, “classe de área” e “zona”, ganham outra relevância que, com a inclusão dos *outliers* não foi possível observar. Sem *outliers*, só foram analisadas 1.935 observações, a que correspondem a 2.013 observações da amostra base, subtraídas dos 77 *outliers* e de mais uma observação que correspondia à freguesia da Madalena e que era simultaneamente *outlier* na amostra de treino.

Figura 6 - Capacidade explicativa das variáveis, cenário 7 (sem outliers)



Fonte: RE/MAX Portugal, análise do autor; Output: SPSS v. 19

Importa por último referir que do primeiro para o segundo teste, com e sem *outliers* respectivamente, a rede melhorou significativamente o seu desempenho no que à capacidade de estimar o valor de transacção diz respeito. Pela análise das Tabelas 7 e 8, constatou-se a melhoria significativa desde o primeiro cenário até à melhor rede desenhada.

Tabela 7 – % de erro da pior rede (cenário 1 – com outliers)

de	até	Frequência	Frequência Σ	% amostra Σ
	5%	317	317	15,7%
5%	10%	257	574	28,5%
10%	15%	293	867	43,1%
15%	20%	249	1.116	55,4%
20%	25%	204	1.320	65,6%
25%	30%	151	1.471	73,1%
30%	40%	214	1.685	83,7%
40%	60%	178	1.863	92,5%
60%	80%	77	1.940	96,4%
80%	100%	38	1.978	98,3%
100%	Mais	35	2.013	100,0%

Fonte: RE/MAX Portugal, análise do autor; Output adaptado: Microsoft Excel

Da leitura da Tabela 7, verificou-se que esta rede (pior cenário) estimou, que cerca de 55% da amostra, continha no máximo, cerca de 20% de desvio face aos valores de transacção.

Tabela 8 – % de erro da melhor rede (cenário 7 – sem *outliers*)

<i>de</i>	<i>até</i>	<i>Frequência</i>	<i>Frequência Σ</i>	<i>% amostra Σ</i>
	5%	434	434	22,4%
5%	10%	335	769	39,7%
10%	15%	288	1.057	54,6%
15%	20%	244	1.301	67,2%
20%	25%	199	1.500	77,5%
25%	30%	121	1.621	83,8%
30%	40%	125	1.746	90,2%
40%	60%	110	1.856	95,9%
60%	80%	52	1.908	98,6%
80%	100%	14	1.922	99,3%
100%	Mais	13	1.935	100,0%

Fonte: RE/MAX Portugal, análise do autor; Output adaptado: Microsoft Excel

Neste cenário, a rede estimou que para os mesmos 20% de erro máximo, o número de observações cresceu para os 67%, variação muito significativa do poder explicativo da rede.

4.2. Análise de sensibilidade

Depois de se ter encontrado a melhor rede disponível com os dados disponibilizados pela RE/MAX, foi testado o poder explicativo da rede, quando se faz variar o ano de transacção. Para tal, foi construído a partir do quadro com as estimativas dos pesos para a melhor rede (Anexo 6), um modelo em Microsoft Excel para que se pudessem simular valores de transacção usando o ano de transacção do imóvel, entre 2008 e 2012, “*ceteris paribus*”. Todo o simulador foi construído tendo por base o apresentado no ponto 3.1 – Metodologia, e serviu para efectuar a análise de sensibilidade ao ano de transacção do imóvel, aferindo a sua razoabilidade com o mercado.

As variáveis quantitativas, “área” e “idade”, foram standardizadas utilizando para o efeito a média e o desvio padrão da amostra de treino (1.352 observações), constantes na Tabela 9. As variáveis qualitativas “n.º quartos”, “freguesia”, “coeficientes de localização”, “zona”, “ano de transacção” e “classe de área”, foram introduzidas através das variáveis “*dummy*” adequadas.

Tabela 9 – estatística descritiva da amostra de treino (1.352 Obs.)

	Min	Max	Mean	Std. Deviation
Valor de transacção (eur)	29.980,00	450.000,00	180.195,72	88.663,58
Área (m ²)	20,00	412,00	99,01	41,66
Idade (anos)	0,00	211,00	34,95	28,27

Fonte: RE/MAX Portugal, análise do autor; Output adaptado: SPSS v. 19

A Tabela 10, ilustra a informação principal da melhor rede neuronal construída com a base de dados disponível (cenário 7 retirados os *outliers*), assim como o número de nodos, variáveis utilizadas e função de activação, necessários para a construção do simulador.

Tabela 10 – Resumo das características da melhor rede neuronal

Input Layer	Factors	1	Q
		2	FRG
		3	CLI
		4	CLS
		5	ZONA ID
		6	ANO
		7	CLASSE ÁREA
	Covariates	1	AREA
		2	IDADE
	Number of Units ^a		97
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 ^a		12
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	VTR
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

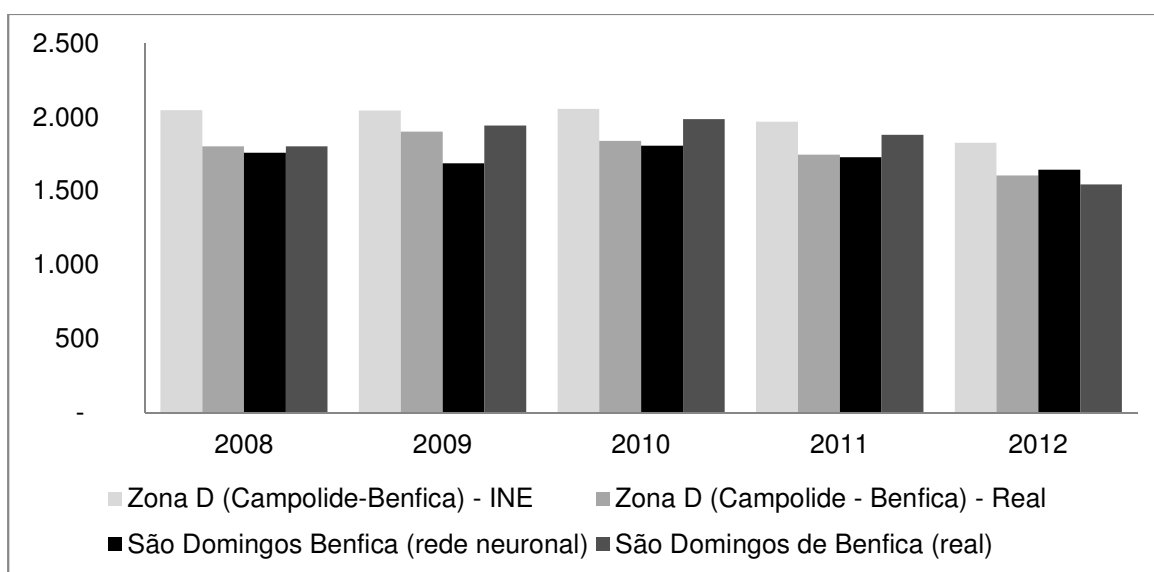
Fonte: RE/MAX Portugal, análise do autor; Output adaptado: SPSS v. 19

Depois de multiplicados os valores dos nodos pelos respectivos pesos, os parâmetros foram sujeitos à sua função de activação (tangente hiperbólica – ver Anexo 1), tendo-se obtido o valor

estandardizado dos parâmetros de output da camada oculta de 12 nodos. Esses valores foram novamente multiplicados pelos pesos de maneira a que se obtivesse o valor estimado de transacção. A esse valor foi adicionada a média do valor de transacção da amostra de treino e multiplicada pelo seu desvio padrão, de maneira a ter-se o valor de transacção sem estar estandardizado.

Foi testada a consistência do simulador, através da confirmação do valor de transacção da base de dados, no caso, um apartamento com 59 anos, 100 m², T3, freguesia de Campolide, adquirido em 2010 (observação ref. 13.473). O resultado foi exactamente igual ao obtido pelo SPSS. Elaborado com sucesso o simulador que calcula o valor estimado de transacção com base na melhor rede neuronal, foi efectuada uma análise de sensibilidade, utilizando uma observação fictícia e que não estivesse constante na base de dados, de maneira a ser testado o poder preditivo da rede. Assim, foi construída uma nova observação, correspondente a um apartamento novo, com 110 m², T2, freguesia de São Domingos de Benfica, fazendo então variar o ano de transacção entre 2008 e 2012 de maneira a identificar os valores estimados pela rede neuronal, quando modificada a variável explicativa “ano de transacção”. De maneira a verificar a consistência dos resultados obtidos, foi comparado o valor de transacção obtido pela rede, com o valor por m² tendo por base valores reais de transacção constantes na base de dados que serviu para o estudo de caso, assim como comparada a sua tendência de valorização ou desvalorização, obtida através do valor médio de avaliação bancária, retirado do Inquérito à Avaliação Bancária na Habitação publicado mensalmente pelo INE, cuja última referência reporta a Julho de 2012 (INE, 2012). Foi utilizado neste último caso o valor médio mensal de habitação dos meses analisados, com referência ao ano.

Figura 7 – Comparação dos valores simulados (eur/m²) com outras fontes (análise de tendência)



Fonte: RE/MAX Portugal, INE; análise do autor; Output: Microsoft Excel

Da leitura da Figura 7, retirou-se a conclusão que a rede construída acompanha relativamente bem a tendência de valorização ou desvalorização do imóvel simulado. Comparando a tendência e grandeza de valores do valor simulado, quer com o valor real, quer sobretudo, com o valor médio por m² por zona apurado pelo INE, verificou-se que a tendência dos valores simulados pela rede é consistente com os observados por outras fontes de informação disponíveis no mercado. Significa assim, que esta rede captou relativamente bem o efeito que o ano de transacção teve no valor estimado e que acompanhou relativamente bem a tendência de valorização e desvalorização do mercado. Ressalva-se contudo, que a rede teve tendência a subvalorizar os imóveis, mesmo que de forma pouco significativa.

5. Conclusões

Foi utilizada uma amostra com um conjunto considerável de observações (2.013) e com uma dispersão de valores bastante elevada, de maneira a que se testassem as reais capacidades do poder de estimação das redes neuronais artificiais. Apenas se utilizou uma parte bastante reduzida das suas reais capacidades, dado que este trabalho pretendeu ser mais um percurso

de outros trabalhos de investigação em Portugal e identificar alguns caminhos que pudessem servir de base à diminuição do erro médio percentual absoluto (MAPE). Foram identificadas duas abordagens que influenciaram positivamente o comportamento das redes neuronais:

- (1) A utilização de uma base de dados que foi tratada estatisticamente, paralelamente ao facto de se excluírem os *outliers* da amostra, melhora substancialmente o desempenho da rede, medida pelo MAPE, confirmando as investigações de Borst (1991), Tay e Ho (1992), Evans, James e Collins (1991) e mais recentemente por Nguyen e Cripps (2001) ou Paterson e Flanagan (2009).
- (2) As variáveis com poder explicativo de dimensão temporal e geográfica, trouxeram melhorias no desempenho da rede neuronal, sobretudo aquando da sua utilização conjunta. O ano em que um determinado apartamento foi transaccionado, a par da sua idade e localização (freguesia), confere um menor erro de estimação ao modelo e aumentam o número de observações com um MAPE inferior a 10%.

O conceito de “*black-box*” encontrado em trabalhos de investigação, afasta muitas vezes a utilização desta metodologia, fundamentalmente porque:

- É um processo de estimação mais complexo e que tem origem numa atribuição aleatória dos pesos iniciais. Neste sentido, duas experiências sobre a mesma amostra, podem conduzir a resultados semelhantes mas não iguais;
- A interpretação dos resultados e dos efeitos das variáveis explicativas, não deixa de ser complexa, quando comparado com a dos modelos de regressão. Uma abordagem semelhante àquela que se fez com o modelo de Excel, pode no entanto ajudar a isolar os efeitos parciais das variáveis explicativas e ajudar a perceber melhor o comportamento das RNA.

Outra das conclusões imediatas que foi possível retirar, é que apesar da base de dados disponibilizada ser muito relevante no que diz respeito à estimação com dados reais de transacção, a mesma também pecou pelo falta de qualidade da informação nela constante, cujos problemas foram a todo o custo eliminados de maneira a comprometerem o menos possível a capacidade preditiva da rede.

Apesar dos resultados de MAPE obtidos, é fundamental não esquecer que o critério de optimização do poder preditivo das redes neuronais modeladas foi sempre tendo em conta a soma do quadrado dos erros e não sobre o MAPE, podendo esse facto ter influência nos resultados finais.

Por último, importa referir que as redes neuronais artificiais são uma séria alternativa a considerar nos modelos de avaliação em massa, e que a sua utilização, depende, à semelhança dos modelos suportados em regressão linear, não só de boa qualidade de informação, como de variáveis que objectivamente se constituam como padrões de comportamento que as redes são exímias a captar.

6. Investigação futura

Para além da comparação óbvia dos resultados obtidos com modelos de regressão linear, importará ainda, testar estes modelos com outras bases de dados em que seja possível, a introdução de mais variáveis explicativas, de maneira a tentativamente afinar melhor o modelo de estimação. Com a introdução de outras variáveis explicativas do valor do imóvel, é provável que o tratamento estatístico da amostra não seja tão determinante e que a identificação dos *outliers*, seja mais facilmente efectuada com o recurso a redes neuronais artificiais.

Pelo comportamento que as variáveis temporais e de localização tiveram na rede, suspeita-se que o seu potencial ficará maximizado se forem utilizadas em conjunto com a modelação geográfica de dados, resultando, muito provavelmente, em sistemas de informação geográficos com uma excelente capacidade preditiva. A sua não linearidade, poderá ser muito vantajosa, não só neste contexto, como também na construção de modelos hedónicos suportados em redes neuronais.

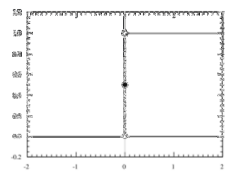
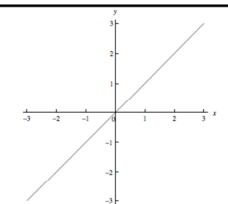
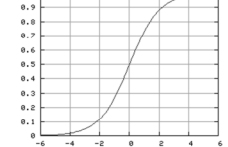
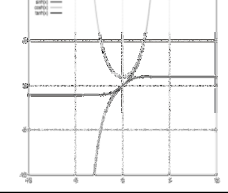
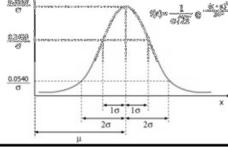
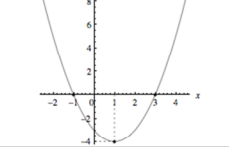

7. Bibliografia

1. Borst, R. (1991). Artificial Neural Networks: The Next Modeling/Calibration Technology for the Assessment Community? *Property Tax Journal*. 10(1), 69–94.
2. Borst, R. (1995). Artificial neural networks in mass appraisal. *Journal of Property Tax Assessment & Administration*. 1(2), 5-15.
3. Borst, R. e McCluskey, W. (1996). The Role of Artificial Neural Networks in the Mass Appraisal of Real Estate. Paper presented to the Third European Real Estate Society Conference, Belfast, June 26-28.
4. Canavarro, M. (2011). Modelos Econométricos para el precio de los inmuebles: un caso de estudio en Portugal. Ed. Universidade de Córdoba.
5. Cortez, P. (1997). Algoritmos Genéticos e Redes Neuronais na Previsão de Séries Temporais. Tese de Mestrado em Informática, Universidade do Minho.
6. Cortez, P. e Neves, J. (2000). “Redes neuronais artificiais”, Unidade de ensino, Escola de Engenharia da Universidade do Minho, Departamento de informática.
7. Do, Q. e Grudnitsky, G. (1992). A Neural Network Approach to Residential Property Appraisal, *The Real Estate Appraiser*. 58, 38-45.
8. Evans, A., James, H. e Collins, A., 1992. “Artificial neural networks: an application to residential valuation in the UK”. *Journal of Property Valuation & Investment* 11, 195-204.
9. I.V.S.C. (2005) International Valuation Standards. International Valuation Standards Committee.
10. INE, (2012), Inquérito à avaliação bancária na habitação, ed. Julho 2012
11. Kathman, R. (1993) Neural Networks for the Mass Appraisal of Real Estate. *Computer Environment and Urban Systems*, 17: 373–384.
12. Kauko, T. e d’Amato, M. (2008a) Property market classification and mass appraisal methodology, Ch 13, in *Mass Appraisal Methods: An International Perspective for Property Valuers* (eds T. Kauko and M. d’Amato), Wiley-Blackwell, Oxford, UK.
13. Kauko, T. e d’Amato, M. (2008b) Suitability issues in mass appraisal methodology, Ch 1, in *Mass Appraisal Methods: An International Perspective for Property Valuers* (eds T. Kauko and M. d’Amato), Wiley-Blackwell, Oxford, UK.
14. Lai, P. e Fischer, D., (2006), “Artificial Neural networks and Computer Assisted Mass Appraisal”, 12th Annual Conference of the Pacific Rim Real Estate Society, January 22 to 25, 2006

15. Lenk, M., Worzala, E. and Silva, A. (1997) High-tech Valuation: Should Artificial Neural Networks Bypass the Human Valuer? *Journal of Property Valuation and Investment*, 15(1), 8–26.
16. Manual do IBM SPSS, Neural Networks, 19. (2010), IBM Publicatons
17. McCluskey, W. (1996) Predictive Accuracy of Machine Learning Models for Mass Appraisal of Residential Property. *New Zealand Valuer's Journal*, July: 41–47.
18. McCluskey, W. e Anand, S., (1999) "The application of intelligent hybrid techniques for the mass appraisal of residential properties", *Journal of Property Investment & Finance*, Vol. 17 Iss: 3, 218 – 239
19. McCluskey, W. e Borst, R. (1997): An evaluation of MRA, comparable sale analysis, and ANNs for the mass appraisal of residential properties in North Ireland. *Assess. J.* 4 (1), 47-55.
20. McCulloch, W. e Pitts, W. (1943): A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophycis*, 5, 115-133.
21. McGreal, S., Adair, A. McBurney, D. e Patterson, D. (1998): Neural Networks:the prediction of residential values. *Journal of Property Valuation & Investment* 16(1), 57-70.
22. Nguyen, N. e Cripps, A. (2001): Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 314-336.
23. Peterson, S. e Flanagan, A. (2009). Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research (JRER)*, 31 (2),147-164
24. Rojas, R. (1996): Neural Networks – A systematic Introduction. Springer-Verlag, Germany.
25. Rosenblatt, R. (1959): Principles of Neurodynamics. Spartan Books. New York.
26. Rossini, P. (1997) Application of Artificial Neural Networks to the Valuation of Residential Property. Third Annual Pacific-Rim Real Estate Society Conference. Palmerston North, New Zealand, January.
27. Rossini, P. (1997): Artificial Neural Networks versus Multiple Regression in the Valuation of Residential Property. *Australian Land Economics Review*. V: 3 N^o1.
28. Tay, D. e Ho, D. (1992): Artificial Intelligence and the Mass Appraisal of Residential Apartments. *Journal of Property Valuation and Investment*. 10:2, 525-40.
29. Tay, D. e Ho, D. (1994) Intelligent Mass Appraisal. *Journal of Property Tax Assessment and Administration*, 1(1): 5–25.
30. Worzala, E., Lenk, M. e Silva, A. (1995). An Exploration of Neural Networks and Its Application to Real Estate Valuation. *Journal of Real Estate Research*, 10(2), 185–201.

8. Anexos

Anexo 1 – Funções de activação de uma rede neuronal artificial

Nome	Função de activação	Contradomínio	Representação gráfica
Limiar (ou de <i>heaviside</i>)	$F(a) = \begin{cases} 1, & u_i \geq 0 \\ 0, & u_i < 0 \end{cases}$	[0,1]	
Por troços	$\begin{cases} 1, & u_i \geq 0.5 \\ ku_i, & -0.5 < u_i < 0.5 \\ 0, & u_i \leq -0.5 \end{cases}$	[0,1]	
Linear	u_i	$]-\infty, +\infty[$	
Logística (ou sigmóide)	$\frac{1}{1+e^{-ku_i}}$	[0,1]	
Tangente hiperbólica	$\tanh(ku_i)$	[-1,1]	
Seno	$\sin(u_i \bmod 2\pi)$	[-1,1]	
Co-seno	$\cos(u_i \bmod 2\pi)$	[-1,1]	
Gaussiana	$\frac{-u_i^2}{e^{\frac{u_i^2}{2k^2}}}$	[-1,1]	
Quadrada	$\text{sign}(u_i) u_i^2$	$]-\infty, +\infty[$	

Anexo 2 – Acordo de confidencialidade

Acordo de Confidencialidade

1. O presente Acordo é estabelecido entre Jorge Filipe Montez Vaz Monteiro Amaro, doravante designado como Mestrando, Cartão de Cidadão número 11031475 1220 e Casax Consultoria de Gestão, Lda (RE/MAX Portugal), doravante designada por Empresa, NIPC 504 689 371, representada pelo seu sócio gerente Manuel Alvarez e com poderes para o acto, em 29/02/2012, no âmbito da elaboração do trabalho final de mestrado em gestão e avaliação imobiliária leccionado no Instituto Superior de Economia e Gestão (ISEG) e subordinada ao tema: "Modelos de avaliação em massa: redes neuronais artificiais aplicadas ao sector imobiliário residencial em Portugal - Estudo de caso na cidade de Lisboa".

2. Com este acordo a Empresa disponibiliza ao Mestrando uma base de dados em formato digital com informação considerada pela Empresa como confidencial e que se anexa, que é sua propriedade e que é considerada pelo Mestrando como indispensável à realização do referido trabalho.

3. O Mestrando, assim como todos os envolvidos na elaboração ou avaliação do trabalho final de mestrado objecto deste acordo, comprometem-se a não revelar ou transmitir a terceiros, firmas concorrentes ou não, qualquer informação considerada e assinalada como confidencial pela Empresa e da propriedade exclusiva desta última, abstendo-se de proceder à sua divulgação independentemente do meio usado e/ou do local da divulgação - assegurando, assim, a máxima confidencialidade da mesma.

4. Qualquer publicação, com origem na informação agora disponibilizada pela Empresa no âmbito da execução do trabalho final de mestrado objecto do presente Acordo, sob a forma de artigos em revistas, congressos, seminários, ou outra, dependerá de acordo prévio entre as partes.

De como assim o disseram e reciprocamente aceitaram, vai o presente Acordo de Confidencialidade ser assinado, destinando-se uma via a cada uma das partes.

Lisboa, 29 de Fevereiro de 2012

A Empresa

RE/MAX PORTUGAL
CASAX - CONS. DE GESTÃO, L.D/
NIPC nº 504 689 371
Rua da Quinta Verde, 10, 1.º andar, E. 12, 1250-149 Lisboa
29/02/2012 - E. 12, 1250-149 Lisboa

Manuel Alvarez

O/a Mestrando(a)


Jorge Filipe Montez Vaz Monteiro Amaro

Anexo 3 – campos da base de dados fornecida pela RE/MAX

1. Ano de construção
2. Área Útil
3. Concelho
4. Distrito
5. Freguesia
6. Zona
7. Código postal
8. Data de transacção
9. Estado de angariação
10. Estado de conservação
11. Frentes
12. Garagem
13. Número de quartos
14. Número de WC
15. Orientação solar
16. Piso
17. Tipo de imóvel
18. Valor de venda
19. Características do imóvel (com cerca de 160 características adicionais, maioritariamente de áreas acessórias, conforto ou da envolvente)

Anexo 4 – Coeficientes de localização de acordo com o CIMI

Freguesia	Bairro	COEF_MIN	COEF_MAX
Ajuda	7º Bº	1,70	2,80
Alcântara	6º Bº	1,70	2,80
Alto do Pina	5º Bº	1,80	2,50
Alvalade	8º Bº	2,20	2,90
Ameixoeira	11º Bº	1,40	2,30
Anjos	4º Bº	2,20	3,00
Beato	9º Bº	1,70	2,00
Benfica	13º Bº	1,60	2,20
Campo Grande	8º Bº	2,20	2,90
Campolide	10º Bº	1,70	3,00
Carnide	12º Bº	1,70	2,45
Castelo	1º Bº	1,20	2,60
Charneca	11º Bº	1,40	2,30
Coração de Jesus	2º Bº	1,80	3,00
Encarnação	3º Bº	2,40	3,00
Graça	1º Bº	1,20	2,60
Lapa	2º Bº	1,80	3,00
Lumiar	11º Bº	1,40	2,30
Madalena	3º Bº	2,40	3,00
Mártires	3º Bº	2,40	3,00
Marvila	9º Bº	1,70	2,00
Mercês	2º Bº	1,80	3,00
Nossa Senhora de Fátima	8º Bº	2,20	2,90
Pena	3º Bº	2,40	3,00
Penha de França	1º Bº	1,20	2,60
Prazeres	6º Bº	1,70	2,80
Sacramento	3º Bº	2,40	3,00
Santa Catarina	3º Bº	2,40	3,00
Santa Engrácia	1º Bº	1,20	2,60
Santa Isabel	2º Bº	1,80	3,00
Santa Justa	3º Bº	2,40	3,00
Santa Maria de Belém	7º Bº	1,70	2,80
Santa Maria dos Olivais	14º Bº	1,60	3,00
Santiago	1º Bº	1,20	2,60
Santo Condestável	2º Bº	1,80	3,00
Santo Estêvão	1º Bº	1,20	2,60
Santos-o-Velho	6º Bº	1,70	2,80
São Cristóvão	1º Bº	1,20	2,60
São Domingos de Benfica	12º Bº	1,70	2,45
São Francisco Xavier	7º Bº	1,70	2,80
São João	1º Bº	1,20	2,60

(cont.)

(cont.)

Freguesia	Bairro	COEF_MIN	COEF_MAX
São João de Brito	8º Bº	2,20	2,90
São João de Deus	5º Bº	1,80	2,50
São Jorge de Arroios	4º Bº	2,20	3,00
São José	3º Bº	2,40	3,00
São Mamede	2º Bº	1,80	3,00
São Miguel	1º Bº	1,20	2,60
São Nicolau	3º Bº	2,40	3,00
São Paulo	3º Bº	2,40	3,00
São Sebastião da Pedreira	10º Bº	1,70	3,00
São Vicente de Fora	1º Bº	1,20	2,60
Sé	3º Bº	2,40	3,00
Socorro	1º Bº	1,20	2,60

Fonte: <http://www.e-financas.gov.pt/SIGIMI/portarias.jsp>

Anexo 5 – Zonas da cidade de Lisboa de acordo com o INE – Inquérito à Avaliação Bancária na Habitação

Freguesia	Zona ID	Zona
Ajuda	A	Alcântara-Belém
Alcântara	A	Alcântara-Belém
Alto do Pina	H	Zona Oriental
Alvalade	B	Avenidas
Ameixoeira	G	Zona Norte
Anjos	F	Zona Histórica
Beato	H	Zona Oriental
Benfica	D	Campolide-Benfica
Campo Grande	B	Avenidas
Campolide	D	Campolide-Benfica
Carnide	G	Zona Norte
Castelo	F	Zona Histórica
Charneca	G	Zona Norte
Coração de Jesus	B	Avenidas
Encarnação	F	Zona Histórica
Graça	F	Zona Histórica
Lapa	E	Lapa-Amoreiras-Campo de Ourique
Lumiar	G	Zona Norte
Madalena	C	Baixa
Mártires	C	Baixa
Marvila	H	Zona Oriental
Mercês	F	Zona Histórica
Nossa Senhora de Fátima	B	Avenidas
Pena	F	Zona Histórica
Penha de França	H	Zona Oriental
Prazeres	F	Zona Histórica
Sacramento	C	Baixa
Santa Catarina	F	Zona Histórica
Santa Engrácia	F	Zona Histórica
Santa Isabel	E	Lapa-Amoreiras-Campo de Ourique
Santa Justa	C	Baixa
Santa Maria de Belém	A	Alcântara-Belém
Santa Maria dos Olivais	H	Zona Oriental
Santiago	F	Zona Histórica
Santo Condestável	E	Lapa-Amoreiras-Campo de Ourique
Santo Estevão	F	Zona Histórica
Santos-o-Velho	F	Zona Histórica
São Cristóvão	F	Zona Histórica
São Domingos de Benfica	D	Campolide-Benfica

(cont.)

(cont.)

Freguesia	Zona ID	Zona
São Francisco Xavier	A	Alcântara-Belém
São João	H	Zona Oriental
São João de Brito	B	Avenidas
São João de Deus	B	Avenidas
São Jorge de Arroios	B	Avenidas
São Jorge de Arroios	B	Avenidas
São José	F	Zona Histórica
São Mamede	F	Zona Histórica
São Miguel	F	Zona Histórica
São Nicolau	C	Baixa
São Paulo	F	Zona Histórica
São Sebastião da Pedreira	B	Avenidas
São Vicente de Fora	F	Zona Histórica
Sé	F	Zona Histórica
Socorro	F	Zona Histórica

Fonte:

http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=133293641&DESTAQUESmodo=2

Anexo 6 – Output SPSS – Pesos nos nodos e parâmetros de estimação

Predictor		Parameter Estimates													VTR
		Predicted													
		Hidden Layer 1											Output Layer		
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	H(1:10)	H(1:11)	H(1:12)		
Input Layer	[Bias]	.493	-.170	-.262	.410	.279	-.139	.505	.057	.204	-.308	-.155	-.591		
	[Q=0]	-.487	-.386	.503	.221	.444	-.300	-.046	.033	.203	.374	.200	.461		
	[Q=1]	.139	-.244	-.531	-.186	-.126	.024	-.135	.249	.364	-.037	.192	-.127		
	[Q=2]	.444	.106	-.457	-.151	-.154	.190	.066	.275	.365	.517	.203	-.332		
	[Q=3]	-.080	-.367	.401	.522	.004	-.265	.294	.160	-.162	.226	-.358	.095		
	[Q=4]	-.092	-.129	-.098	.521	-.178	-.268	-.265	.200	.372	.265	-.316	-.068		
	[Q=5]	-.307	-.135	.046	.476	.102	.204	.096	.009	.209	-.259	-.118	-.480		
	[Q=6]	-.066	-.559	.065	.604	.205	-.434	.393	-.232	.192	.149	-.211	.200		
	[Q=8]	-.008	.422	-.126	-.250	.260	.143	.462	.449	-.322	-.324	-.318	.331		
	[FRG=Aluda]	-.019	-.424	.138	-.051	.340	-.369	-.298	-.170	.175	.294	-.190	.275		
	[FRG=Alcintara]	-.298	.242	.040	-.276	-.092	-.081	.350	-.060	.198	.420	-.467	.483		
	[FRG=Alto do Pina]	.421	-.073	.471	.194	-.066	-.300	.001	.124	-.173	.334	-.388	-.306		
	[FRG=Alvalade]	-.075	.267	.055	.161	-.094	-.157	.176	.067	.043	.116	-.126	-.373		
	[FRG=Ameixoeira]	-.419	-.357	.079	.549	.479	.066	.121	.405	.250	-.313	.413	-.076		
	[FRG=Anjos]	.488	.353	.222	.017	-.270	-.412	.434	.321	.289	.359	-.084	-.216		
	[FRG=Beato]	.369	-.332	.055	.068	-.511	.179	.440	.108	.198	-.312	.215	-.184		
	[FRG=Benfica]	.329	-.274	.430	-.248	.099	.181	.404	-.382	.056	.477	.149	-.274		
	[FRG=Campo Grande]	.122	.502	.206	-.425	-.436	.191	-.207	-.061	.279	.413	.244	-.342		
	[FRG=Campolide]	.271	-.521	.249	.615	-.168	-.151	-.248	-.503	-.353	.341	.341	.111		
	[FRG=Carnide]	-.299	-.140	-.205	.391	-.333	-.308	-.100	.197	.070	-.342	.435	.174		
	[FRG=Castelo]	-.131	-.271	.262	-.202	-.338	-.294	.199	.440	-.308	.031	.341	.085		
	[FRG=Coração de Jesus]	.451	-.548	.491	.119	-.515	-.331	.019	.419	-.473	.482	-.289	-.520		
	[FRG=Encarnação]	.406	.509	.063	.176	.434	.121	.499	-.317	-.068	.053	-.229	.063		
	[FRG=Graca]	-.224	-.351	.325	-.601	.110	.213	.502	.224	-.149	-.364	.414	.147		
	[FRG=Lapa]	-.159	.157	-.090	-.200	.003	.496	-.387	.317	-.064	.162	-.380	.428		
	[FRG=Lumiar]	.148	-.386	-.263	-.058	.327	-.144	.195	-.119	-.018	-.250	-.097	.390		
	[FRG=Mártires]	.284	.259	.325	.260	.008	.256	.497	-.134	-.214	-.384	-.142	.186		
	[FRG=Marvila]	.100	-.189	-.145	.625	-.155	.417	.417	.302	-.068	.130	.523	-.068		
	[FRG=Merçês]	.191	.464	-.460	-.019	-.347	-.487	-.010	.110	.191	.058	-.323	-.268		
	[FRG=Nossa Senhora de Fátima]	.216	.359	-.109	.068	-.384	.360	.037	.284	.143	-.117	.048	.349		
	[FRG=Pena]	.378	-.287	.389	-.067	-.263	.358	.481	-.315	.433	.314	.205	.320		
	[FRG=Penha de França]	-.193	.298	.199	-.047	-.015	-.183	.484	-.383	.353	.431	-.446	.065		
	[FRG=Prazeres]	-.108	.293	-.162	.382	.295	.038	.417	.303	-.444	-.137	.086	.386		
	[FRG=Sacramento]	-.007	-.337	.338	-.128	.120	-.306	.400	.413	-.200	.179	.273	.113		
	[FRG=Santa Catarina]	.330	.353	-.511	.342	-.051	-.165	.450	.299	-.081	-.356	-.453	.144		
	[FRG=Santa Engrácia]	.428	-.264	-.448	.483	.227	.094	.292	.355	-.401	.416	-.442	-.174		
	[FRG=Santa Isabel]	.053	-.340	.008	.330	.459	-.219	.294	-.198	.336	-.163	.365	.225		
	[FRG=Santa Justa]	-.166	.177	.229	.590	-.088	-.038	-.060	.263	-.113	-.067	.321	-.054		
	[FRG=Santa Maria de Belém]	.163	-.386	-.108	-.383	.070	.328	.180	-.038	-.176	-.138	-.186	.193		
	[FRG=Santa Maria dos Olivais]	-.435	.290	-.134	-.293	-.269	.315	-.559	.489	.125	.312	.012	.384		
	[FRG=Santiago]	-.312	-.412	-.386	-.330	-.407	.223	.465	.092	-.518	.073	.148	.177		
	[FRG=Santo Condestável]	.367	-.178	.029	-.362	.241	.349	-.377	-.104	-.444	-.388	.471	-.363		
	[FRG=Santo Estevão]	-.219	-.488	.057	.068	.144	.006	.130	-.355	.416	.013	.333	-.327		
	[FRG=Santos-o-Velho]	.146	.198	-.059	.205	.444	-.130	-.476	.078	-.249	.397	.081	.361		
	[FRG=São Cristóvão]	.469	.019	-.135	.470	.007	-.405	.009	.124	.065	-.375	-.089	-.449		
	[FRG=São Domingos de Benfica]	-.265	.153	-.226	.010	-.226	.330	.456	-.246	.247	.486	.178	.141		
	[FRG=São Francisco Xavier]	-.242	.438	-.459	.014	.064	-.245	-.299	-.154	-.065	.262	-.362	.311		
	[FRG=São João]	-.230	.425	-.303	.147	.468	.041	.057	.028	.042	-.008	.236	.348		
	[FRG=São João de Brito]	.328	-.314	-.305	.104	-.261	.184	-.303	.405	.249	.212	.441	.423		
	[FRG=São João de Deus]	-.236	.206	-.296	.330	-.009	-.278	-.391	.055	.360	.529	-.215	.035		
	[FRG=São Jorge de Arroios]	.442	.205	-.307	.381	.201	.186	.420	.529	.371	-.239	-.173	.301		
	[FRG=São José]	.029	-.500	.380	.054	.487	.286	-.024	.499	.507	.424	.048	-.125		
	[FRG=São Mamede]	-.009	-.432	-.498	-.194	.123	.165	.329	-.119	.398	.115	-.135	.187		
	[FRG=São Miguel]	.495	-.254	.134	.443	.292	.146	-.076	-.042	-.439	-.316	-.108	-.383		
	[FRG=São Nicolau]	.092	.508	-.227	.019	.519	.189	-.068	-.435	-.305	-.006	-.163	-.049		
	[FRG=São Paulo]	-.433	.187	.333	.356	-.049	-.267	.458	.429	.470	-.108	.487	.451		
	[FRG=São Sebastião da Pedreira]	.125	-.237	-.254	-.387	-.038	.289	.431	-.344	-.343	.434	-.465	.047		
	[FRG=São Vicente de Fora]	-.087	-.326	.355	.026	-.197	.154	.262	.364	.451	-.165	-.015	-.448		
	[FRG=Sá]	-.461	.209	-.083	.080	-.215	.316	-.381	-.166	.288	.398	.206	.313		
	[FRG=Socorro]	.134	-.264	.407	.191	.094	-.242	.206	.293	.016	-.137	.005	.314		
	[CLI=1.2]	.341	.198	-.238	.348	.336	.520	.163	-.153	.357	.451	.261	.204		
	[CLI=1.4]	-.319	-.484	.416	-.072	.192	.040	-.204	.495	.039	.143	-.306	.039		
	[CLI=1.6]	.398	.309	.529	.317	.294	.166	.379	.088	.228	-.030	.576	.265		
	[CLI=1.7]	-.213	-.025	.251	.290	.267	.032	-.016	-.350	-.378	.435	.124	-.362		
	[CLI=1.8]	-.577	.235	-.498	-.082	.264	.146	-.092	-.281	-.540	-.286	.257	-.025		
	[CLI=2.0]	-.056	.245	.464	-.028	.063	.105	.377	.364	-.302	-.158	-.106	-.265		
	[CLI=2.2]	.043	-.479	-.059	.524	-.220	-.238	.391	-.079	.334	-.497	-.233	-.153		
	[CLI=2.4]	.388	-.293	.283	-.501	.502	-.083	.005	.117	-.058	-.528	-.474	.067		
	[CLS=2.0]	.161	-.058	.641	.128	.074	-.378	.232	.676	-.103	.126	.272	-.129		
	[CLS=2.1]	.386	-.376	.497	-.219	.385	.377	.202	.399	.286	-.239	-.213	-.015		
	[CLS=2.3]	.144	.134	-.121	.017	-.103	-.029	.012	.016	.322	-.025	-.344	-.429		
	[CLS=2.4]	-.400	.001	.281	.129	.192	.376	.178	.272	.413	-.183	-.103	.386		
	[CLS=2.5]	.352	-.383	-.170	-.250	.426	.431	-.462	.095	.162	-.257	-.253	-.558		
	[CLS=2.6]	.306	-.146	.106	.421	.312	-.321	.225	-.302	.172	-.301	.425	-.113		
	[CLS=2.8]	-.360	-.031	.116	-.347	-.330	.371	-.060	.148	.405	.255	-.357	-.197		
	[CLS=2.9]	-.458	-.301	-.140	-.306	.065	.057	-.459	.040	-.383	.276	-.154	-.194		
	[CLS=3.0]	-.047	.313	.246	.063	.278	.021	.173	.056	.214	.033	.183	-.014		
	[ZONAIID=A]	-.061	-.388	-.086	.102	.020	.337	-.455	-.212	.448	.503	.412	.260		
	[ZONAIID=B]	-.399	.398	-.202	-.354	-.008	.367	.436	-.212	-.268	.502	-.213	.516		
	[ZONAIID=C]	.021	-.509	-.156	.419	-.033	.196	.208	-.073	-.110	.235	-.497	-.025		
	[ZONAIID=D]	-.046	.275	.011	-.336	-.319	-.024	.435	.048	-.407	.521	-.067	.112		
	[ZONAIID=E]	.094	-.331	.288	-.172	-.256	-.025	-.164	.375	-.484	-.075	.194	.232		
	[ZONAIID=F]	.103	.089	.299	.107	.205	.062	.370	.117	.046	.474	-.606	-.289		
	[ZONAIID=G]	-.548	.260	.349	.165	.495	.041	.575	.391	.309	.404	.003	.134		
	[ZONAIID=H]	.687	.082	-.222	-.083	.028	.155	-.328	.324	-.325	.164	-.255	-.362		
	[ANO=2007]	-.104	-.013	-.011	-.277	-.382	.077	-.396	.415	-.245	-.100	.230	.421		
	[ANO=2008]	-.230	-.368	-.098	.364	.433	-.394	.134	-.295	-.445	-.266	-.528	.190		
	[ANO=2009]	.131	.405	-.349	-.251	-.392	.451								