



INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Área departamental de Engenharia Civil



AVALIAÇÃO IMOBILIÁRIA
COM
REDES NEURONAIS ARTIFICIAIS

BRUNO GUIA NETO MACHADO

LICENCIADO

Dissertação para obtenção do grau de Mestre em Engenharia Civil – Área de
especialização em Edificações

Orientador:

Doutor Filipe Manuel Vaz Pinto Almeida Vasques

Júri:

Presidente: Doutor Paulo Jorge Henriques Mendes

Vogais:

Licenciado Especialista Manuel Augusto Gambôa

Doutor Filipe Manuel Vaz Pinto Almeida Vasques

Janeiro de 2022

AGRADECIMENTOS

Ao ISEL por ser, mais uma vez, a instituição que me proporcionou esta nova etapa académica.

A todos os professores que contribuíram para o meu incremento de conhecimentos na área de Engenharia, sempre importantes, independentemente da fase de vida em que nos encontramos.

Ao Professor Vasques pelo entusiasmo e força que incutiu nesta iniciativa, desde que a propôs até à sua conclusão.

Ao meu Pai que apesar da área de Finanças, sempre foi um poço de cultura e um excelente provedor na arte de bem escrever. À minha Mãe pelo seu apoio emocional e incondicional positivismo.

Aos amigos e familiares que leram e criticaram este trabalho, colaborando assim para o seu aprimoramento.

Sem qualquer um dos acima, não teria sido possível a execução deste trabalho.

Bem hajam.

RESUMO

Esta dissertação desenvolve, avalia, compara e critica, de forma inovadora, um caso prático de avaliação imobiliária na área metropolitana de Lisboa, com recurso a redes neuronais artificiais (RNA).

Para contextualizar o trabalho, é introduzido um enquadramento histórico das RNA, expondo o seu crescimento conceptual e científico (matemático), dando por fim a conhecer a sua aplicabilidade nas diversas áreas científicas com enfoque na avaliação imobiliária, demonstrando o potencial da sua utilização nesta área.

Seguidamente, é desenvolvido um caso prático, expondo todos os passos necessários à realização do método, desde a recolha de dados essenciais para amostra de referência por intermédio de programa específico, e com critérios pré-determinados para a avaliação pretendida, até à utilização de programas também específicos que sustentam a utilização do método.

São apresentados em detalhe todos os aspetos que envolvem a programação da RNA, desde o tratamento necessário dar à amostra, a metodologia e os modelos testados apontando o modelo de RNA mais adequado para este tipo de avaliação, por forma a obter os melhores resultados.

Para validação do método de estimação por RNA, é aplicado à mesma amostra, com os mesmos critérios e variáveis, um modelo hedónico de regressão, servindo de método comparativo.

Por fim, são apresentados resultados que demonstram a fiabilidade do método de RNA na estimação de valores imobiliários, bem como a sua superioridade relativamente a métodos hedónicos de regressão linear (MHRL), apontando as principais vantagens e desvantagens na sua utilização e evidenciando deste modo a necessidade de novos estudos e aprofundamento deste método.

PALAVRAS-CHAVE:

Avaliação imobiliária; RNA; Redes Neuronais Artificiais.

ABSTRACT

This dissertation develops, evaluates, compares and criticizes, in an innovative way, a practical case of real estate valuation in the metropolitan area of Lisbon, using artificial neural networks (ANN).

To contextualize the work, a historical framework of ANN is introduced, exposing its conceptual and scientific (mathematical) growth, finally revealing its applicability in various scientific areas with a focus on real estate valuation, demonstrating the potential of its use in this area.

Following, a practical case is developed, exposing all the steps necessary to carry out the method, from the collection of essential data for a reference sample through a specific program, and with predetermined criteria for the intended evaluation, to the use of programs that support the use of the method.

All aspects involving ANN programming are presented in detail, from the treatment necessary to give the sample, the methodology and models tested, pointing out the most appropriate ANN model for this type of assessment, in order to obtain the best results.

To validate the ANN estimation method, a hedonic regression model is applied to the same sample, with the same criteria and variables, serving as a comparative method.

Finally, results are presented that demonstrate the reliability of the RNA method in estimating real estate values, as well as its superiority over hedonic linear regression methods (MHRL), pointing out the main advantages and disadvantages in its use and thus evidencing the need for further studies and deepening of this method.

KEYWORDS:

Real estate evaluation, ANN, Artificial Neural Network

Índice

1. INTRODUÇÃO	1
1.1 ENQUADRAMENTO.....	1
1.2 OBJETIVOS.....	1
1.3 ABORDAGEM.....	2
1.4 METODOLOGIA.....	2
1.5 ORGANIZAÇÃO.....	3
2. INTRODUÇÃO ÀS REDES NEURONAS ARTIFICIAIS	5
2.1 COMO E PORQUE SURTIRAM AS RNA	5
2.2 FUNCIONAMENTO DE UMA REDE NEURONAL ARTIFICIAL	9
2.2.1 <i>O que é uma Rede Neuronal Artificial</i>	10
2.2.2 <i>A retro propagação do erro e as RNA na atualidade</i>	14
2.3 AS RNA E A AVALIAÇÃO IMOBILIÁRIA	17
2.3.1 <i>As PropTech e as questões da atualidade</i>	18
2.3.2 <i>Utilização e fiabilidade das RNA na avaliação imobiliária</i>	20
2.3.3 <i>Potencial futuro das RNA no setor imobiliário</i>	22
3. RECOLHA DE DADOS E CRIAÇÃO DE BASES DE INFORMAÇÃO PREDIAL	25
3.1 O MÉTODO DE RECOLHA DE DADOS	25
3.2 RECOLHA, TRATAMENTO E FINALIZAÇÃO DA AMOSTRA.....	28
3.3 ANÁLISE DOS DADOS PARA APRENDIZAGEM E TESTE	31
4. A ESTIMAÇÃO IMOBILIÁRIA.....	35
4.1 O MODELO HEDÓNICO DE REGRESSÃO LINEAR (MHRL).....	35
4.1.1 <i>Introdução ao programa Regressit™</i>	35
4.1.2 <i>O método de regressão linear</i>	35
4.1.3 <i>Análise de resultados</i>	37
4.2 O MODELO DE REDE NEURONAL ARTIFICIAL (RNA).....	40
4.2.1 <i>Introdução ao programa de Estimação Neural Designer®</i>	40
4.2.2 <i>Desenvolvimento do método de RNA</i>	41
4.2.3 <i>Resultados e observações</i>	49
4.3 COMPARAÇÃO DE RESULTADOS	50
5. CONCLUSÕES.....	53
5.1 A RECOLHA E TRATAMENTO DOS DADOS PREDIAIS	53
5.2 A UTILIZAÇÃO DOS PROGRAMAS.....	54
5.3 CONSIDERAÇÕES FINAIS	55
REFERÊNCIAS BIBLIOGRÁFICAS.....	57
WEBGRAFIA	63

Anexo I – Tabela de Valores médios de frações por Estados de Freguesias para deteção de representatividade de valores por freguesia.....	65
Anexo II – Tabela de dados estatísticos da amostra global.....	67
Anexo III – Tabela de análise de resultados e comparativa de dados resultantes das diferentes técnicas de estimação por RNA e MHRL.....	69
Anexo IV - Relatório do Modelo 5 do Neural Designer®(29 Páginas)	71

Índice de figuras

Figura 1.1 – (Imagem de Capa) Fonte : James E. Klein, SIOR (03/10/2020).....	Error!
Bookmark not defined.	
Figura 2.1 – Representações do ‘Perceptron’ (Rosenblatt, 1958).....	6
Figura 2.2 – Modelo de Werbos (Paul Werbos,1975) e marcos cronológicos de desenvolvimentos verificados posteriormente	7
Figura 2.3 – Kasparov vs IBM Deep Blue / Fonte: Stanford.edu	8
Figura 2.4 – Composição e ligação neuronal (sinapses) / Fonte:Saymedia	10
Figura 2.5 – Gráfico da Função Logística / Fonte: ResearchGate - Allysson Alex Araújo	12
Figura 2.6 – Rede Neuronal Simples (Fausett, 1994)	12
Figura 2.7 – Afectação do peso (w) e enviesamento (b) na curva sigmoide / Fonte: towardsdatascience – Shane De Silva.....	13
Figura 2.8 – ‘Sub-ajustado’ (linha laranja), Ajustado (linha preta), Sobre-ajustado (linha verde) / Fonte: Medium – Jaron Collis	14
Figura 2.9 – PropTech na área Residencial e Comercial / Fonte: Thomvest Ventures Real Estate Tech Review (2019).....	19
Figura 3.1 – Exemplo de dados de imóvel detalhados	25
Figura 3.2 – Exemplo de dados de imóvel pouco detalhados	25
Figura 3.3 – Programa WebHarvy [®] a coletar dados.....	27
Figura 3.4 – Recolha de dados com WebHarvy [®] de páginas de Imobiliárias Oficiais ..	27
Figura 3.5 - Recolha de dados com WebHarvy [®] de páginas de promoção imobiliária livre	28
Figura 3.6 – Compilação de dados coletados.....	29
Figura 3.7 – Tabela de dados rawdata tratados	30
Figura 3.8 – Gráficos comparativos nas relações VO (test/train), N°Im.(test/train), Δ RNA/VO.....	33
Figura 3.9 – Gráficos comparativos nas relações VO(test/train), N°Im.(test/train), Δ RL/VO	34
Figura 4.1 – Tabela de dados para Regressão Linear.....	36
Figura 4.2 – Menu do ‘Add-on’ em Excel® – Regressit TM	36
Figura 4.3 – Janela de definição dos dados para Regressão e resultados pretendidos ...	36
Figura 4.4 – Tabela de Resultados da Regressão Linear.....	38

Figura 4.5 – Gráfico da relação de valores estimados da RL com valores reais de oferta	40
Figura 4.6 – ‘ <i>Data set</i> ’ aplicado no Neural Designer®	42
Figura 4.7 – ‘ <i>Neural Network</i> ’ aplicado no Neural Designer®	43
Figura 4.8 – Rede Neuronal Artificial Fonte: o autor	44
Figura 4.9 – Híper superfície definida pelos dados objetivo (apenas com 2 parâmetros) Fonte: página “Neural Designer® training strategy tutorial”	46
Figura 4.10 – Estratégia de treino ou aprendizagem aplicado no Neural Designer®	47
Figura 4.11 – Sequência de operações numa iteração de treino de uma RNA Fonte: página “Neural Designer® training strategy tutorial”	48
Figura 4.12 – Grafico da relação de valores estimados com valores reais de oferta	50
Figura 4.13 – Gráficos comparativo entre resultados da RNA e Regressão Linear, recorrendo aos desvios de distribuição por Freguesia e por QC	51

Índice de tabelas

Tabela 3.1 – Tabela de Máximos e Mínimos por Estado, Qualidade e Conservação, por Freguesia	31
Tabela 4.1 – Distribuição das incidências percentuais pelas percentagens de erro.....	39
Tabela 4.2 – Comparativo de resultados entre métodos de regularização L1 e L2, PR e IR variáveis	47
Tabela 4.3 – Distribuição das incidências percentuais pelas percentagens de erro.....	50
Tabela 4.4 – Diferenciais entre % de incidências por percentis de erro absoluto	52

Índice de equações

Equação 2.1 – Função Logística ou Sigmoides	12
Equação 2.2 – Produto escalar entre entradas e pesos correspondentes (Winston, 2015)	12
Equação 2.3 – Função logística com peso e enviesamento (Winston, 2015).....	13
Equação 4.1 – Escalonamento de saída por máximo-mínimo.....	43
Equação 4.2 – Erro Quadrático Normalizado	45

Simbologia

RNA	Rede Neuronal Artificial (<i>ANN – Artificial Neural Network</i>)
MHRL	Método Hedónico de Regressão Linear
RN	Rede Neuronal (<i>NN – Neural Network</i>)
CP	Computador Pessoal (<i>PC – Personal Computer</i>)
IA	Inteligência Artificial (<i>AI – Artificial Intelligence</i>)
RNP	Rede Neuronal Profunda (<i>DNN – Deep Neural Net</i>)
EQM	Erro quadrático médio
EQN	Erro quadrático normalizado
EQP	Erro quadrático ponderado
EEC	Erro de entropia cruzada
EM	Erro de Minkowski (<i>ME – Minkowski Error</i>)
MAA	Modelo Automatizado de Avaliação
RNC	Rede Neuronal Convolutacional
AC	Aprendizagem Computacional (<i>ML – Machine Learning</i>)
PLN	Processamento de Linguagem Natural (<i>NLP – Natural Language Processing</i>)
VC	Visão Computacional (<i>CV – Computer Vision</i>)
RL	Regressão Linear
Q&C / QC	Qualidade e Conservação
FREG	Freguesia
AB	Área Bruta
VAL	Valor de Oferta
IR	Índice de Redução
TE	Termo de Erro
TR	Termo de Regularização
PR	Peso de Regularização
L1/L2	Métodos de regularização
MQN	Método Quasi-Newton
MGC	Método do Gradiente Conjugado
VO	Valor de Oferta

1. INTRODUÇÃO

1.1 Enquadramento

A avaliação de imóveis é um trabalho muito requerido por instituições bancárias, agências imobiliárias e diversas outras entidades. É normalmente concretizada por avaliadores qualificados e com experiência no mercado imobiliário, que estimam o valor de mercado do imóvel com recurso a diversos métodos, mas que, por forma a se enquadrarem nas exigências de tempo e custo dos requerentes, são normalmente aplicados com base numa grande componente empírica. Torna-se assim relevante a utilização de métodos que possam ser mais sustentados do ponto de vista metodológico de informação de base, sem prejuízo da sua rapidez e facilidade de utilização, quando se pretende uma avaliação massiva fidedigna e num curto intervalo de tempo.

No caso de Portugal, assim como em inúmeros outros países, a avaliação imobiliária tem-se transformado numa tarefa premente e de elevada complexidade, dado o papel que o setor imobiliário tem tido no modelo de crescimento da economia portuguesa, bem como nas alterações profundas recentes na valorização e procura de imóveis.

Face a essas questões, têm surgido novas técnicas associadas aos principais métodos de avaliação permitindo dar uma resposta sustentada, justificada cientificamente e gradualmente mais eficaz. Neste enquadramento, surgem cada vez mais frequentemente novos trabalhos e publicações sobre a aplicação de redes neuronais, que se pretendem que sejam exploradas e comprovadas como técnica para potenciar a aplicação de métodos de previsão suportados em dados diretos de mercado.

1.2 Objetivos

Este trabalho tem por objetivo aplicar a metodologia de previsão não paramétrica (não linear) que simula uma rede neuronal (artificial), no contexto da avaliação de imóveis para habitação, circunscrita a uma área metropolitana.

Pretende-se que este trabalho contribua para o aprofundamento da aplicação das Redes Neurais (RN) no contexto da atividade de estimação de valores imobiliários, aumentando o conhecimento, e o interesse na aplicação desta metodologia preditiva.

Pretende-se também, verificar a fiabilidade e facilidade da utilização desta técnica na avaliação imobiliária feita com base no método de mercado, permitindo que um utilizador desta área a possa empregar, estimulando a utilização de redes neuronais artificiais nesta área, fomentando deste modo a criatividade e desenvolvimento de novas abordagens no futuro.

1.3 Abordagem

Em termos conceptuais, a abordagem a este trabalho foi dividida em 3 etapas.

Na primeira etapa é apresentado o histórico e desenvolvimento científico das RNA em geral. Seguidamente, é analisado o desenvolvimento atual deste método com particular incidência na avaliação imobiliária. Deste modo, é contextualizado de forma sucinta o que são as RNA, como surgiram, quais as utilizações atualmente atribuídas e conhecidas, mas também perspetivando quais as vantagens e inconvenientes na aplicação deste tipo de modelos na vertente de avaliação imobiliária.

A análise bibliográfica serviu para a definição detalhada da metodologia de aplicação, tendo por base as previsíveis vantagens e desvantagens dessa escolha.

Numa segunda etapa é exposta e explicada a metodologia implementada, justificando e desenvolvendo o trabalho proposto que, resumidamente, consiste na recolha de uma amostra e aplicação comparada de dois métodos preditivos de mercado, um baseado nas RNA, cuja aplicabilidade é testada neste trabalho, e outro conhecido e comprovado baseado num modelo hedónico de previsão.

A terceira e última etapa faz a comparação dos dados obtidos, análise e avaliação de resultados, retirando assim e por fim, conclusões sobre o potencial da utilização de RNA na determinação de valores imobiliários.

1.4 Metodologia

A metodologia para a aplicação e validação de RNA na determinação de valores imobiliários é dividida em duas fases.

Na primeira fase são utilizados métodos expeditos e atuais, informatizados, para a recolha sistemática de dados do mercado imobiliário, existente na *Internet*, no sentido de construir uma amostra completa e atualizada. Esta amostra, reúne dados dispersos de vários locais na *Internet*, nem sempre relacionados, mediante critérios de busca específicos e pré-definidos e por conseguinte, que necessitam de um tratamento específico, *a posteriori*, para ser utilizado no modelo preditivo recorrendo a RNA.

Na segunda fase, recorre-se a um programa informático que permite a construção e desenvolvimento do modelo preditivo de RNA, aplicando neste a amostra e colhendo os resultados. Do mesmo modo, se procede à aplicação de um modelo hedónico de regressão à mesma amostra para posterior comparação. Os critérios comumente utilizados na avaliação imobiliária, são aplicados de igual modo, a ambos os métodos para comparação.

1.5 Organização

A dissertação, para além deste capítulo introdutório, desenvolve-se nos capítulos seguidamente indicados (2-5).

Capítulo 2 – Introdução às Redes Neurais Artificiais

Este capítulo foi dividido em três partes e é suportado em revisão bibliográfica. Na primeira parte, é introduzida uma nota histórica sobre a origem das RNA até ao seu desenvolvimento atual. Na segunda parte, é descrito e explicado o funcionamento das RNA, introduzindo o conceito científico que sustenta os modelos atuais e forma de os desenvolver. A terceira parte faz uma análise de como as RNA podem ser utilizadas na avaliação imobiliária, quais os principais critérios na sua utilização atualmente, e qual o seu potencial e possíveis utilizações futuras.

Capítulo 3 – Recolha de dados e criação de bases de informação predial

Este capítulo é dividido em três partes:

- a) A primeira parte consiste na apresentação do método utilizado na pesquisa, seleção e recolha de dados específicos, na *Internet*, para a criação de bases de dados, e que constituíram a amostra. São também observados os critérios de recolha de dados.
- b) A segunda parte consiste na apresentação da recolha e do tratamento efetuado aos dados da amostra, para que esta possa ser finalizada e introduzida no método de avaliação por RNA e no método de Regressão Linear.

- c) A terceira e última parte consiste na análise e avaliação da amostra, apontando questões pertinentes que afetam os modelos de avaliação.

Capítulo 4 – A estimação imobiliária

Este capítulo introduz os programas que executam os modelos de avaliação por RNA e por Regressão Linear. Faz uma descrição do funcionamento dos mesmos na execução de ambos os métodos de avaliação, de forma detalhada, apresentando os seus resultados individualmente, e por fim comparando os mesmos realçando as suas vantagens e desvantagens.

São expostos os critérios que definem ambas as técnicas de avaliação desenvolvidas individualmente e, no final, em comparação. No caso do método por RNA, é escrutinado todo o processo, explicando os passos que são dados e opções tomadas por forma a escolher um modelo para comparação, definir o algoritmo e a estratégia de aprendizagem. É detalhado todo o processo de operação dos programas escolhidos para desenvolvimento dos métodos de mercado e na apresentação de resultados. Após executados ambos os modelos, são extraídos e apresentados os resultados para observação. Por fim, comparam-se os resultados de ambas as técnicas de avaliação, apontando as situações pertinentes que justificam os dados obtidos em cada situação.

Capítulo 5 – Conclusões (Análise crítica ao método)

Este capítulo faz uma análise dos objetivos propostos atingir na utilização de RNA na avaliação imobiliária, destacando o interesse em desenvolvimentos futuros, possibilidades de progresso e desenvolvimento nesta área, determinando assim o potencial inerente ao método de avaliação por RNA. São particularizados os principais constrangimentos, deficiências e dificuldades verificadas na estimação de valores por RNA, comparativamente com o modelo hedónico de regressão, tendo por base os resultados obtidos e toda a metodologia desenvolvida.

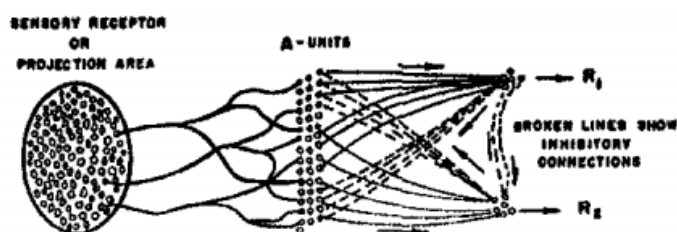
2. INTRODUÇÃO ÀS REDES NEURONAS ARTIFICIAIS

2.1 Como e porque surgiram as RNA

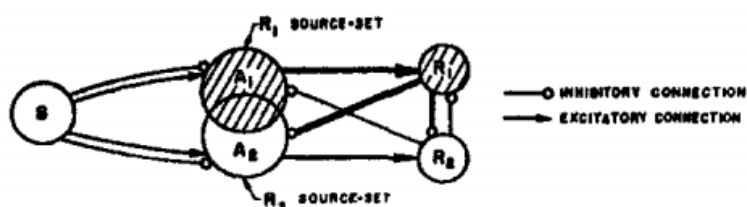
McCulloch e Pitts, efetuaram em 1943 uma analogia entre as células nervosas e um processo elétrico, no contexto de uma publicação sobre o tema, num artigo do *'Bulletin of Mathematical Biophysics'* (McCulloch & Pitts, 1943). Em 1949 Hebb aprofunda o estudo sobre sistemas nervosos complexos (Morris, 1999) e formula uma lei que descreve o funcionamento quantitativo da sinapse e do processo de aprendizagem humano. Em 1951, Minsky como co-fundador do Laboratório de Inteligência Artificial do MIT, construiu o SNARC (*'Stochastic Neural Analog Reinforcement Calculator'*) - o primeiro simulador de cadeia neuronal (Forbus et al., 2016). Estavam assim lançadas as bases para que muitos países e investigadores se debruçassem sobre o desenvolvimento destas temáticas que abraçam o estudo científico do pensamento humano.

Em 1956 McCarthy cunhou o termo 'inteligência artificial' (IA) num tópico da proposta de um projeto de investigação efetuado na 'Conferência de Dartmouth' (Buchanan, 2006; McCorduck, 2004). Nesta conferência foi proposto que "um estudo de inteligência artificial de 2 meses e 10 homens" fosse realizado durante o verão de 1956 no Dartmouth College em Hanover, New Hampshire. O estudo deveria "prosseguir com base na conjectura de que cada aspeto da aprendizagem ou qualquer outra característica da inteligência pode, em princípio, ser descrito com tanta precisão que será possível conceber uma máquina que consiga simulá-lo.". Seria efetuada "uma tentativa de descobrir como fazer para que as máquinas usassem linguagem, conseguissem formar abstrações e conceitos, resolvessem problemas tipo reservados aos humanos, e que conseguissem aperfeiçoar-se.". Acreditava-se que um avanço científico significativo poderia ser atingido, numa ou mais dessas questões, se um grupo cuidadosamente selecionado de cientistas trabalhasse em conjunto durante esse verão (*A proposal for the Dartmouth summer research project on artificial Intelligence*, J. McCarthy, Dartmouth College; M. L. Minsky, Harvard University; N. Rochester, I.B.M. Corporation; C.E. Shannon, Bell Telephone Laboratories - August 31, 1956).

Em 1958, Rosenblatt, criou uma rede de múltiplos neurónios do tipo discriminadores lineares que batizou de rede ‘perceptron’ representando-a como na Figura 2.1. abaixo. Rosenblatt baseou-se nas linhas de pensamento de McCulloch para desenvolver o seu modelo matemático de sinapse humana. Devido às suas complexas pesquisas e inúmeras contribuições técnicas, Rosenblatt é considerado o fundador da neurocomputação. Mas logo de seguida, Minsky e Papert lançaram uma obra também chamada *Perceptron*, na qual demonstravam que o modelo apresentado por Rosenblatt não era muito promissor, devido ao uso de técnicas empíricas, das grandes dificuldades da matemática envolvida, e dos poucos recursos computacionais disponíveis à época. Esta publicação de Minsky e Papert provocou o corte no investimento financeiro nesta área e mergulhou a investigação num período redutor até meados da década de 80 (Fausett, 1994; Teixeira et al., 2010).



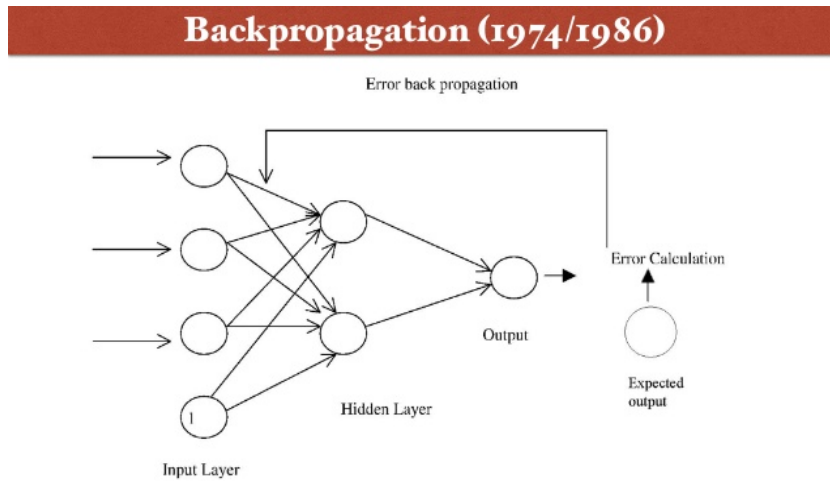
Schematic representation of connections in a simple perceptron.



Venn diagram of the same perceptron (shading shows active sets for R_1 response).

Figura 2.1 – Representações do ‘Perceptron’ (Rosenblatt, 1958)

Em 1974, Werbos formulou o algoritmo de retro propagação do erro (*backpropagation*), demonstrando que as redes neuronais de múltiplas camadas detinham capacidade de aprendizagem (Fausett, 1994). Já na década de 80, muitos investigadores dedicaram-se e publicaram inúmeros estudos sobre redes neuronais, como Hopfield (1982) na neurocomputação, seguindo-se Hopfield na mesma área criticando as teorias apresentadas por Minsky e Papert (1969) (Fausett, 1994; Martins et al., 2016; Teixeira et al., 2010). Distinguem-se nestas três décadas de 70 a 90, três marcos históricos conforme representados na Figura 2.2 seguinte.



1974 Paul Werbos' invents Backpropagation algorithm for NN
1986 Backdrop popularized by Rumelhart, Hinton, Williams
1990: Renewed Interest in NN's

Figura 2.2 – Modelo de Werbos (Paul Werbos, 1975) e marcos cronológicos de desenvolvimentos verificados posteriormente

Alimentada pelas fantasias de Julio Verne do século XIX até à ficção de Isaac Asimov e Kubrick do século XX, o surgimento da IA está inundada de sonhos, possibilidades, demonstrações e promessas, tendo inúmeros autores destas épocas inspirado muitos investigadores (Buchanan, 2006). A título de exemplo, o Xadrez era uma realização que requeria pensamento profundo, pelo que durante muito tempo foi comumente utilizado para estudar mecanismos de inferência e de representação nos estudos de IA. Um marco importante foi alcançado quando o programa *Deep Blue* derrotou o campeão mundial de xadrez, Gary Kasparov, em 1997 – Figura 2.3 (McCorduck, 2004).



Figura 2.3 – Kasparov vs IBM Deep Blue / Fonte: Stanford.edu

Com as invenções eletrônicas do início do século XX e o surgimento de computadores modernos no laboratório de Alan Turing em Manchester, na Escola Moore em Penn, no laboratório de Howard Aiken em Harvard, nos Laboratórios IBM e Bell e outros após a Segunda Guerra Mundial, um mundo de possibilidades sobrepôs-se ao das demonstrações.

No entanto, somente no último meio século, surgiram dispositivos computacionais e linguagens de programação suficientemente poderosos, para construir testes experimentais de ideias sobre o que é inteligência.

Os anos 1980 e 1990 são caracterizados pelo desenvolvimento revolucionário no setor de gestão de informação com a introdução do computador pessoal (PC) e recurso a redes de comunicação por computador. Durante esse período, múltiplos programas foram desenvolvidos e tornaram-se acessíveis.

O crescimento tecnológico exponencial dos últimos 30 anos, deve-se também ao surgimento e desenvolvimento da *Internet*. Esta transformou completamente a comunicação, a forma como se veicula e partilha informação, como se vende e compram produtos, como se recrutam e prestam serviços, entre múltiplas outras facetas que moldam atualmente o funcionamento profissional e social de toda a população do mundo desenvolvido, provando que a *Internet* é pedra basilar para a conectividade mundial e a portadora de muita informação científica que impulsiona presentemente a humanidade.

A maioria das ferramentas dominantes de programas dedicados a qualquer operação, disponíveis hoje, provém muito provavelmente de ideias, formulações empíricas e desenvolvimentos académicos das décadas de 60 e seguintes.

Muito embora o desenvolvimento da IA ainda esteja numa evolucionar, é já real e aplicada de forma diversa. Nos últimos anos, tendo em conta o aumento da capacidade computacional e

de ligação em rede, assistiu-se a um crescimento exponencial do número de aplicações de IA Restrita que estão a mudar de forma decisiva as nossas vidas (Vasques, 2017).

O desenvolvimento tecnológico dos computadores, nomeadamente no que respeita a capacidade de processamento, possibilitou a criação de programas que operam algoritmos complexos e que simulam o comportamento do cérebro humano, tendo por base o seu método biológico e altamente complexo de aprendizagem. Este conceito desenvolveu-se considerando que o ser humano (ou o seu comportamento) pode ser traduzido num algoritmo. Substituem-se os impulsos eletroquímicos que permitem a um ser humano ter emoções e consciência, por processamentos de impulsos elétricos de código binário. O fato de ser um cálculo matemático em vez de processo eletroquímico, torna-o muito menos complexo e menos falível, dado que a análise probabilística que leva à tomada de decisões é apenas influenciada pela experiência da aprendizagem, ao invés de influenciada pelas emoções, criando situações contraditórias e ilógicas (Camacho, 2020).

O desenvolvimento deste pensamento levou à criação de algoritmos com instruções para processar informação de forma específica permitindo, com o auxílio de computadores, criar redes neuronais artificiais suficientemente potentes para ‘aprender’, de forma automática.

A Inteligência Artificial (IA) na última década, foi introduzida em todo o tipo de atividades e funcionalidades, como reconhecimento de cores, números, objetos ou animais em imagens, traduções linguísticas, desenvolvimento da fala/leitura computacional, avaliar padrões de comportamento e definir respostas automáticas, efetuar análises financeiras, procedimentos de indexação, respostas de emergência, diagnósticos de saúde e avaliações imobiliárias, entre muitas outras, sendo atualmente um universo vasto e transversal a todas as áreas científicas. (Choromanska et al., 2014)

Na maioria destas aplicações está a utilização das chamadas *deep neural nets* (DNN) ou RNA profundas (Zhao et al., 2017), onde são empregues programas com algoritmos de aprendizagem que desenvolvem milhares de iterações por segundo com o intuito de fornecer os resultados pretendidos pelo utilizador.

2.2 Funcionamento de uma rede neuronal artificial

No coração de muitos sistemas de programação de inteligência artificial estão as redes neurais artificiais. Estas redes foram inspiradas, como já foi referido, no arranjo e funcionamento biológico do cérebro humano.

Utilizando uma estrutura de ‘neurónios’ interligados, estas redes podem distinguir arranjos numéricos, ‘aprender’ a processar certos estímulos e produzir resultados sem a intervenção humana.

Importa perceber, nas RNA, como ocorre o processo de transformação dos dados de uma amostra, que se pretenda avaliar quando é aplicado um algoritmo, no sentido de obter um resultado específico. E também, em termos leigos, como funciona esse algoritmo.

Por conseguinte, para explicar o funcionamento de uma RNA é necessário introduzir o raciocínio matemático que traduz o seu funcionamento, mesmo que de uma forma simplificada.

2.2.1 O QUE É UMA REDE NEURONAL ARTIFICIAL

Não basta dizer que as RNA reproduzem as estruturas biológicas de neurónios – Figura 2.4, entenda-se, primeiro que tudo, que estas têm parcialmente um comportamento reprodutível matematicamente e que essa componente teve a sua descoberta e evolução ao longo da nossa mais recente história. No entanto, outros tipos de comportamentos mantêm-se no vasto desconhecido do campo neurológico dos nossos cérebros.

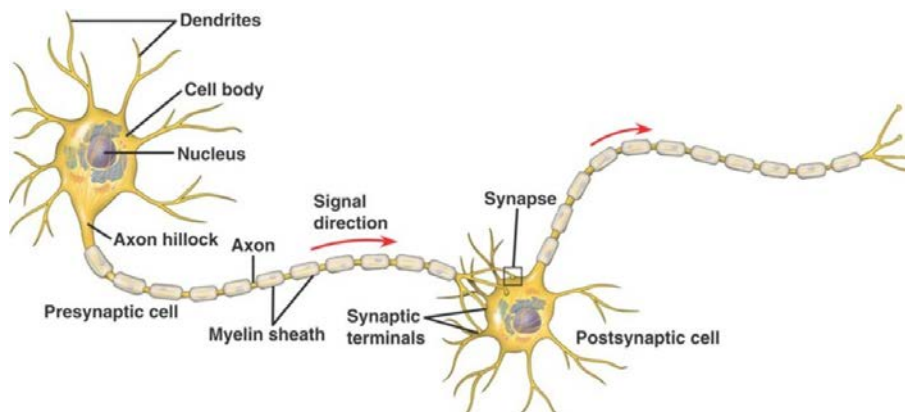


Figura 2.4 – Composição e ligação neuronal (sinapses) / Fonte: Saymedia

As redes neuronais artificiais foram desenvolvidas como generalizações de modelos matemáticos de biologia humana, com base nos pressupostos de que:

- O processamento de informação ocorre em muitos elementos simples chamados neurónios;
- Os sinais são transmitidos entre neurónios por intermédio das sinapses;
- Cada ligação tem um peso associado que, numa rede típica, multiplica ao sinal transmitido;
- Cada neurónio aplica uma função de ativação (geralmente não linear) à sua entrada (soma de sinais de entrada ponderados) para determinar o seu sinal de saída.

Uma rede neuronal caracteriza-se:

- Por um padrão de conexões entre os seus neurónios, comumente designada de *arquitetura* do modelo;

- Pelo seu método de determinar os pesos nas ligações que corresponde ao treino ou aprendizagem ou algoritmo;
- Pela sua função de ativação.

Os pesos representam a informação utilizada pela rede para resolver um problema. As redes neuronais podem ser aplicadas a uma grande variedade de problemas, tais como armazenar e recordar dados ou padrões, classificar padrões, realizar mapeamentos gerais como padrões de entrada e de saída, agrupar padrões semelhantes ou encontrar soluções para problemas constrictos de otimização (Fausett, 1994).

Na Universidade do MIT explicam que “(...), considerando uma rede com várias entradas $x_1 \dots x_n$, às quais são aplicados os pesos $w_1 \dots w_n$, o resultado desse conjunto será a soma $x_1 \cdot w_1 + \dots + x_n \cdot w_n = z_1 \dots z_n$. Esse polinómio, atingindo um determinado limite t , o axónio será estimulado resultando no valor de 1, caso contrário no valor 0. Deste modo diz-se que o resultado ou funcionamento de uma rede neuronal binária, neste modelo, que pretende simular um neurónio biológico, está correlacionado com:

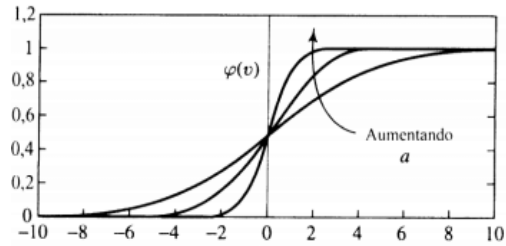
- Tudo ou nada (1 ou 0);
- Influência cumulativa;
- Peso sináptico., sendo que estas serão as componentes possíveis de modelar.”
(Winston, 2015).

De uma forma simplificada, os modelos neuronais fornecem vetores de saída, em função dos vetores de entradas x e dos pesos w e limites t que os influenciam, ou seja, para treinar uma rede neuronal, são ajustados os pesos e limites de modo a obter-se o resultado ótimo ou pretendido. Uma rede neuronal é, por conseguinte, definida por uma função aproximativa ou comumente designada de algoritmo *descendente* ou *decrecente*.

Mas era também necessário fornecer uma função que ajustasse os resultados ao pretendido, simultaneamente descontínuo ou não linear, o que demorou cerca de 25 anos até que Werbos (1974) encontrasse uma resposta.

Para resolver o problema da descontinuidade, Werbos (1974) utilizou a função logística também denominada ‘sigmoide’ pelos países Anglo-saxónicos, como função de ativação. Uma função muito conveniente por ter uma adaptação ideal para os resultados pretendidos circunscritos aos valores binários (0,1) para qualquer entrada α racional (Equação 2.1 e Figura 2.5).

$$f(x) = \frac{1}{1 + e^{(-\alpha)}}$$



Equação 2.1 – Função Logística ou Sigmoid

Figura 2.5 – Gráfico da Função Logística / Fonte: ResearchGate - Allysson Alex Araújo

Tomando como referência a Figura 2.6 abaixo, supondo que o neurónio Y está por sua vez ligado aos neurónios Z1 e Z2, com pesos v_1 e v_2 , respetivamente, o neurónio Y envia o seu sinal para cada uma destas unidades. A presença desta unidade oculta Y, juntamente com uma função de ativação não linear, confere a capacidade de resolver muitos mais problemas do que por uma rede com apenas unidades de entrada e saída. Por outro lado, é mais difícil treinar (ou seja, encontrar valores ideais para os pesos) uma rede com unidades ocultas (Fausett, 1994).

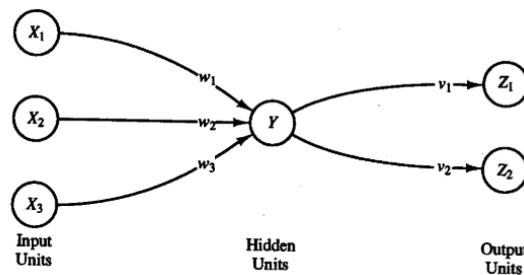


Figura 2.6 – Rede Neuronal Simples (Fausett, 1994)

Regressando ao desenvolvimento dado no MIT, numa rede neuronal com múltiplas entradas que são afetadas pelos pesos correspondentes, obtém-se como resultado do conjunto algo que não é mais do que um produto escalar ou comumente conhecido como *dot product*, antes de servir de entrada na função logística – Equação 2.2 (Winston, 2015).

$$\sum_n^i x_i w_i$$

Equação 2.2 – Produto escalar entre entradas e pesos correspondentes (Winston, 2015)

Este produto escalar determina a representação matemática, numa iteração, das entradas, segundo a distribuição de pesos aleatórios a cada uma delas. Mas para reduzir os valores positivos e negativos deste produto escalar à análise binária que ativa ou não o neurónio, aplica-se a função

logística que reduz as saídas a valores entre 0 (para produtos escalares negativos) e 1 (para produtos escalares positivos). Suplementarmente, adiciona-se a este produto escalar um valor nominal positivo ou negativo chamado enviesamento, também comumente conhecido por ‘bias’, que refletirá situações em que se pretende que o neurónio seja ativado com um nível de significância maior ou menor respetivamente.

Numa representação gráfica que explica sucintamente o acima, podem observar-se comportamentos distintos na função logística com a variação do peso e do enviesamento (Equação 2.3). O aumento do valor do peso w refletirá um acentuar da curvatura ou declive dessa função, e a variação do enviesamento refletirá um deslocamento da função segundo o eixo das abcissas (Figura 2.7). Um deslocamento negativo deste último significará uma maior dificuldade ou probabilidade em o resultado ser próximo de 1 e, por conseguinte, de o neurónio ser ativado.

$$f(x) = \frac{1}{1 + e^{(-x \cdot w - b)}}$$

Equação 2.3 – Função logística com peso e enviesamento (Winston, 2015)

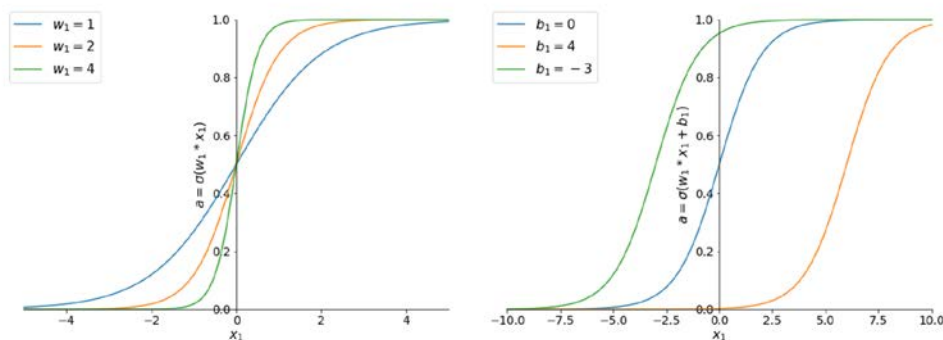


Figura 2.7 – Afecção do peso (w) e enviesamento (b) na curva sigmoide / Fonte: towardsdatascience – Shane De Silva

O termo *enviesamento* é também referido como o erro sistemático do modelo, a tendência de continuar a obter os mesmos resultados incorretamente. Isso tende a surgir quando um modelo é muito simples e, portanto, não consegue representar a complexidade dos dados subjacentes, fazendo com que um algoritmo perca relações relevantes entre as características salientes e os resultados esperados, independentemente de ter dados de aprendizagem mais do que suficientes - um problema conhecido como sub-ajustamento (*Under fitting*). Por outro lado, se o algoritmo for adaptado para memorizar até casos especiais dar-se-ão casos de sobre-ajustamento (*Over fitting*) (Figura 2.8) (Oladipupo, 2010).

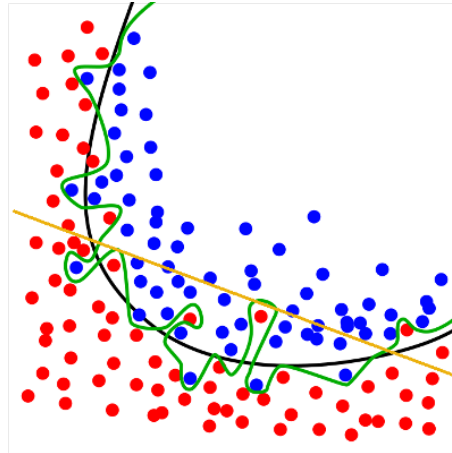


Figura 2.8 – ‘Sub-ajustado’ (linha laranja), Ajustado (linha preta), Sobre-ajustado (linha verde) / Fonte: Medium – Jaron Collis

Como se referiu anteriormente, quando se trata da aprendizagem do modelo, é o *enviesamento* e os pesos que o computador irá calibrar, iterando e convergindo vectorialmente num gradiente descendente para valores mínimos da função, para que os valores de saída se ajustem aos valores reais. A este processo denomina-se ‘retro-propagação’ do erro, mais conhecido pelo termo anglófono *back-propagation*. A retro-propagação não é mais do que um modelo de rede de propagação contínua ou frontal ‘*feedforward*’, com pelo menos uma camada de neurónios ‘escondida’, em que no resultado gerado é avaliado o erro que este tem em relação ao valor real, e em função dessa diferença são reprogramados os pesos e *enviesamento* da rede.

Quando acima se fala em ‘probabilidade’ pretende-se de fato dar a perceber que, sendo quer os pesos quer o enviesamento aleatórios, os valores resultantes na curva sigmoide serão, cada um deles, probabilísticos. No entanto, será por intermédio da ‘aprendizagem’ que a função logística se deformará vectorialmente de modo a reproduzir tendencialmente os valores resultantes reais para cada entrada.

2.2.2 A RETRO PROPAGAÇÃO DO ERRO E AS RNA NA ATUALIDADE

Nas últimas décadas, diversos investigadores têm aprofundado a temática de aprendizagem em modelos de retro propagação, apontando diversas abordagens no que respeita ao tratamento dos dados de entrada, à tipologia de construção da rede, às funções de ativação, verificação do erro, algoritmos de otimização e por fim o tratamento dos dados de saída.

No que respeita aos dados de entrada, há todo um tratamento indispensável a dar inicialmente aos dados de amostra, e que resulta essencialmente de duas questões:

1. Da existência de diversos tipos de erros nos modelos de RNA (i.e., dados espúrios ou ‘*outliers*’ – valores *desenquadrados dos demais a que tem proximidade e por esse motivo mal reconhecidos pelas redes*), o que se torna um processo moroso e

complexo para grandes amostras, mas cujo tratamento é possível com técnicas contemporâneas para detecção de valores discrepantes (Oladipupo, 2010). Este tratamento é imprescindível para se evitar as situações de sobre-ajustamento (*over fitting*);

2. Da conveniência em estruturar os valores de entrada na rede, por forma a terem um intervalo adequado (escala) para uma progressão mais célere na aprendizagem da rede. Torna-se um passo indispensável no caso de os dados de entrada serem abstratos, como cores, objetos ou outros, que tem de ter forçosamente uma conversão numérica podendo por exemplo ser binária. Guo e Berkhan apresentam um estudo onde aprofundam este tema (Guo & Berkahn, 2016). Para entradas numéricas, o dimensionamento não é mais do que um método que atribui uma escala à amostra (quaisquer métodos lineares, e que em geral produzem o mesmo tipo de resultados), sendo os mais usados:
 - Método de escalonamento entre mínimo e máximo ($=(\text{entrada-min})/(\text{max-min})$);
 - Método de escalonamento recorrendo a média e desvio padrão ($=(\text{entrada-média})/(\text{desvio padrão})$);
 - Método de escalonamento por desvio padrão ($=\text{entrada}/\text{desvio padrão}$).

Por outro lado, os dados disponíveis da amostra devem ser divididos em duas partes, ou seja, para a aprendizagem do modelo e para a avaliação do modelo. Não há uma proporção de divisão comumente aceite, no entanto, uma divisão dos dados entre 80% (para aprendizagem do modelo) e 20% (para teste do modelo) é comum na literatura. Isso justifica o fato de a avaliação de propriedades (métodos tradicionais ou avançados) ser uma arte que requer habilidades e a experiência de avaliadores (Chan & Abidoeye, 2019).

Numa rede neuronal artificial de múltiplas entradas, é também comum a utilização de múltiplas camadas de neurónios nas camadas intermédias (entre a camada de entrada e a camada de saída), denominadas camadas ‘escondidas’ não havendo, no entanto, uma fórmula ou um consenso da comunidade quanto ao número de neurónios ou mesmo o número de camadas para casos específicos. Estas camadas são, no entanto, necessárias para a aprendizagem do modelo de retro propagação, sendo que é necessário pelo menos uma para desenvolver um modelo de RNA preditivo (*multilayer Perceptron*), como se pode visualizar na Figura 2.2 anterior. Na verdade, alguns autores testaram redes com mais de uma camada escondida, concluindo que a performance com apenas uma camada escondida produzia melhores resultados que qualquer das restantes (Abidoeye & Chan, 2018; Khalafallah, 2008), resultando mais que uma camada apenas em maior complexidade e mais tempo de processamento.

No que respeita a funções de ativação, muitos desenvolvimentos se tem produzido na literatura ao longo destes últimos anos, para além da função logística (*sigmoide*), entre as mais

conhecidas: a função *tangente hiperbólica*; a função de *ativação linear*; e umas das mais recentes e utilizadas: a função *linear retificada* (*Rectified Linear Unit – ReLU*) (Fausett, 1994; Sattelberg et al., 2020), também nas variantes ‘Leaky ReLU’, ‘Parameterised ReLU’ e outras (Liang & Xu, 2021), a função Swish (Mercioni & Holban, 2020) utilizada pela Google, embora outras soluções possam ser aplicadas (Dudek, 2016; Sütfeld et al., 2020) tendo cada uma um grau de aproximação diferente.

A aprendizagem do algoritmo, depende, no entanto, de uma função essencial que identifica e quantifica o erro. Atualmente utilizam-se pelo menos 5 tipos de equações para o efeito:

- erro quadrático médio (EQM) (mais comumente utilizada);
- erro quadrático normalizado (EQN);
- erro quadrático ponderado (EQP);
- erro de entropia cruzada (EEC);
- erro de Minkowski (EM).

(Bishop, 1986; Jafarian et al., 2018; Núñez-Tabales et al., 2016)

Depois da quantificação e regularização do erro de um ciclo iterativo, prossegue a verificação e validação do critério de paragem. Essa validação é efetuada pelo *índice de redução (IR)* que é vital pois, não só define que função será utilizada seguidamente, como fornece uma quantificação sobre a ‘qualidade’ da aprendizagem do modelo em cada iteração. Este *índice* não é mais do que um termo de regularização somado ao erro obtido. O termo de regularização pode ser obtido segundo dois métodos:

- pela multiplicação de um *peso de regularização* ao somatório de todos os parâmetros da rede (pesos e *enviesamento*);
- pela multiplicação de um *peso de regularização* ao somatório quadrático de todos os parâmetros da rede (pesos e *enviesamento*).

Não chegando ao critério de paragem, o modelo iterativo prossegue mediante o algoritmo de otimização escolhido, sendo que atualmente são conhecidos e referenciados por diversos autores como Baptistella, Bishop, Bock, Liu e Oladipupo, diferentes algoritmos, como sejam:

- o *gradiente descendente*, o mais simples que atualiza em cada iteração na direção do gradiente negativo do erro obtido;
- o *gradiente conjugado*, onde a verificação é realizada com direções conjugadas, o que geralmente produz uma convergência mais rápida do que direções do gradiente descendente;
- o método *Quasi-Newton* de derivadas parciais conforme acima descrito, sendo um processo de alta precisão, mas de cálculo e processamento mais longo e moroso;

- o algoritmo de *Levenberg-Marquardt* que só pode ser considerado com a introdução do erro quadrático EQN ou EQP, requerendo o cálculo do gradiente matricial Jacobiano do erro;
- o algoritmo de *gradiente descendente estocástico*, de natureza diferente dos restantes, atualizando os parâmetros em cada iteração utilizando lotes de dados;
- o algoritmo adaptativo de momento linear (tradução do inglês *Adaptive linear momentum* (ADAM)), similar ao gradiente descendente implementando um método mais sofisticado no cálculo da direção de aprendizagem e que usualmente produz uma convergência mais rápida.

(Baptistella et al., 2007; Bishop, 1986; Bock & Weiß, 2021; LIU et al., 2011; Núñez-Tabales et al., 2016; Oladipupo, 2010).

Por fim, uma solução de saída diz-se *normalizada* quando pequenas alterações nos valores de entrada produzem também pequenas variações de saída, o que é avaliado como foi dito pela minimização do *índice de redução*.

O ciclo iterativo suspender-se-á quando uma condição especificada é satisfeita. Alguns critérios de suspensão comumente usados são (Hidalgo-Cajo et al., 2020):

- A norma de incremento de parâmetros é menor que um valor mínimo;
- O índice de redução numa iteração é menor do que um predeterminado valor;
- O índice foi minimizado para um valor objetivo;
- A norma do gradiente do índice de redução fica abaixo de um determinado objetivo;
- Um número máximo de iterações é atingido;
- O tempo limite de processamento foi atingido;
- O erro num subconjunto selecionado aumenta com várias iterações.

Para, finalmente, se obterem os dados preditivos da rede, é necessário tratar os dados de saída, num reescalamento, com o intuito de devolver as unidades de operação originalmente obtidas na amostra. Este reescalamento assume os mesmos métodos utilizados, mas de forma inversa ao escalonamento das entradas, utilizando-se aqui o método logarítmico (*saída reescalada* = $\min + 0,5(\exp(\text{saída}) + 1) * (\max - \min)$).

2.3 As RNA e a Avaliação Imobiliária

É evidente que a IA já faz parte das nossas vidas, e como referiu o *influencer* Bernard Marr (Forbes, in ‘The Top PropTech Trends: 6 Technologies Disrupting The Property And Real Estate Industry’, Feb, 2020), o “*Big Data*” e a “*AI and machine learning*” estão no top das tendências de PropTech (*property technology*). As RNA combinam ambas, pelo que constituirão

certamente a base e a ferramenta para uma nova era tecnológica, o que ocorre e se desenvolve neste momento, também, no setor imobiliário.

A *Internet* acumula um sem fim de possibilidades no que toca a informação, quer histórica quer atual e utilitária, e na atualidade existem meios de extração e compilação dessa informação em quantidades massivas, num curto espaço de tempo – ação denominada *Web Scraping* (Uzun, 2020), o que potencia a utilização do *Big Data* na aprendizagem de modelos de RNA adaptados para responder a diversas utilidades na avaliação imobiliária. Mas quais as necessidades do imobiliário que se pretendem ver resolvidas recorrendo a novas tecnologias? Que PropTech está presentemente em desenvolvimento? Quais as potencialidades e fiabilidade das RNA nestas questões do presente e no futuro?

2.3.1 AS PROPTech E AS QUESTÕES DA ATUALIDADE

Nos últimos anos tem-se verificado uma evolução significativa dos meios informatizados utilizados no setor imobiliário. O *Web Scraping* permitiu, como exemplo, que rapidamente se desenvolvessem *Census* informáticos no setor, com todas as vantagens associadas. Boeing e Waddell (2017) recolheram por este meio, 11 milhões de registos do mercado de arrendamento americano, tratando posteriormente essa amostra, o que lhes permitiu avaliar a acessibilidade da população ao mercado, pelo cálculo de proporções entre o suporte populacional das rendas e o desenvolvimento urbano, salientando a importância de estes dados fornecerem informação em tempo real e à escala local, sobre características do arrendamento e da habitação, que à época faltavam em fontes alternativas, como em dados do censo.

Por outro lado, na comercialização de imóveis, o número de serviços informatizados cresceu exponencialmente, sendo que milhões de utilizadores em todo o mundo recorrem atualmente a páginas da *Internet*, programadas por empresas que apenas publicitam imóveis dos proprietários ou também medeiam a sua comercialização. Estas páginas possuem atualmente um conjunto de facilidades, que tem vindo a ser incrementadas com o crescimento da sua utilização. Sejam para definir a localização pretendida, a área, tipologia entre outras propriedades, estas utilidades pretendem essencialmente agilizar a procura e a promoção de imóveis.

Verifica-se atualmente e cada vez mais, uma transformação nos processos de procura e na mediação imobiliária, com a incorporação de vários processos de IA que personalizam cada vez mais a experiência do utilizador (Vasques, 2017). Um exemplo relativamente recente, que adaptou novas tecnologias ao setor, foram as virtualizações dos espaços para visitas 3D nas páginas que divulgam e promovem os imóveis.

Muitas e novas tecnologias têm agilizado o setor imobiliário, nas designadas por Baum (2017) de ‘Smart real estate’. Existem um conjunto de plataformas e tecnologias para melhoria da eficiência na Gestão de Manutenção (*Facility Management*), na Gestão da Propriedade

(*Property Management*), otimização comercial dos espaços, otimização da relação com o setor bancário e financeiro, de investimento, automação dos espaços, melhoria dos espaços comuns, gestão de impostos, entre muitos outros (Figura 2.9).

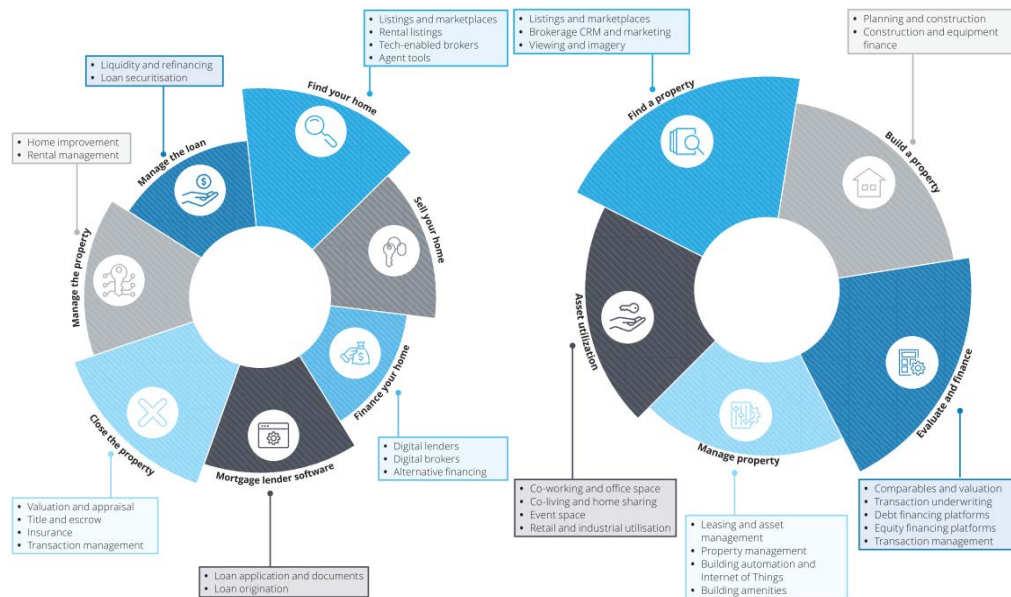


Figura 2.9 – PropTech na área Residencial e Comercial / Fonte: Thomvest Ventures Real Estate Tech Review (2019)

Existe, no entanto, uma preocupação crescente no que respeita à tendência atual do mercado imobiliário, antes tendencialmente lenta, às novas tecnologias, transparecendo uma corrida ao digital numa concorrência global e desafiante espoletada pelas PropTech e que transformam supostamente o mercado imobiliário num mercado movido e dependente do digital (Braesemann & Baum, 2020). A mesma publicação concluiu que “(...) todas essas inovações podem potencialmente levar a ganhos substanciais de eficiência. No entanto, ao mesmo tempo, eles mudam toda a estrutura do mercado imobiliário. Os mercados de dados são geralmente caracterizados por estruturas de mercado de oligopólio (com algumas empresas) ou monopólios que oferecem um único serviço digital disponível. O mercado imobiliário ainda não se tornou um monopólio de dados, mas os gigantes digitais Amazon e Google já deram os primeiros passos no mesmo.

Para evitar a acumulação de poder sobre o mercado, atendendo a que este é dominado pelas PropTech, beneficiando simultaneamente dos ganhos de eficiência associados à introdução de tecnologias digitais no mercado, os utilizadores e proprietários de imóveis necessitam de estar conscientes do valor dos dados que estão a gerar ao incorporar os mesmos nestas tecnologias.” (Braesemann & Baum, 2020).

Todavia, no caso específico da comercialização imobiliária (de maior escala), depende ainda muito do utilizador certas atividades, como encontrar o melhor preço de mercado para o enquadramento pretendido ou imóvel em carteira, estabelecer contatos entre o vendedor e comprador, ou com a empresa de mediação imobiliária, ou a título de exemplo, serem divulgadas

inúmeras questões de detalhe da propriedade. Perceber quais as tendências atuais na valorização das tipologias de imóveis, que o utilizador procura, com dados precisos, é também uma lacuna importante. O mesmo se pode referir a contextos associados à localização, como níveis de criminalidade, escolas na proximidade, transportes, comercio, etc.. Muito importante será, no contexto da contratualização da compra e/ou venda, a antecipação de todos os requisitos necessários para o efeito, dependendo totalmente dos intervenientes a agilização e concretização de todo o processo, independentemente de este ser mais ou menos burocratizado.

Pretende-se, por conseguinte, que estas PropTech agilizem os processos de compra e venda transformando toda a ação numa experiência (quase) caseira. Esse é o desafio, e estas são as necessidades que se pretendem ver respondidas, nas quais diversas empresas investem neste momento.

Há, no entanto, sérias preocupações relativas a segurança de dados e privacidade que se levantam com estas tendências, e que cada vez surgem com maior rapidez, colocando-se atualmente muitas questões e entraves na sua implementação. É para os investidores uma grande preocupação, pois as transações em imobiliário envolvem por vezes muita informação privada que vale muito dinheiro, para além de dados pessoais dos intervenientes. Donati, A. (Forbes, 2020) publicou um artigo sobre esta temática, explicando as preocupações que o promotor imobiliário deverá ter com um conjunto de fatores que são essenciais na segurança informática. Expõe também o que as PropTech podem e devem atualmente fazer nesse âmbito.

No entanto, Brien (Deloitte US, 2020) refere que “As PropTech terão um impacto crescente no setor imobiliário no futuro, com potencial para destabilizar uma das maiores e mais antigas indústrias do mundo. Será expectável o surgimento de muitas novas PropTechs e, eventualmente, é provável que haja alguma consolidação à medida que estas combinam os seus recursos e capacidades para se tornarem mais influentes no setor imobiliário. Como qualquer área de start-ups, no espaço das PropTech existirão vencedores e perdedores, e o envolvimento ativo de investidores e operadores imobiliários ajudará a determinar quem serão esses vencedores.” (Brien, 2020).

2.3.2 UTILIZAÇÃO E FIABILIDADE DAS RNA NA AVALIAÇÃO IMOBILIÁRIA

No desenvolvimento atrás dado, pretendeu-se dar uma noção de como surgiram, como funcionam e como se desenvolveram as redes neuronais artificiais, permitindo aqui resumir que:

- O funcionamento de uma RNA depende da definição de um conjunto de parâmetros como a sua arquitetura, a sua função de ativação, a verificação do erro e o algoritmo de otimização utilizado;

- A aprendizagem e eficácia da rede é tão mais elevada e rápida, quanto melhor afinados forem os parâmetros do modelo em utilização, bem como o tratamento efetuado aos dados de entrada (tratamento da amostra) e aos dados de saída (permitindo a que a rede entenda melhor o erro em cada iteração, e também perceber quando chegou ao limite mínimo do gradiente descendente).

Constata-se atualmente um nível de desenvolvimento e avaliação destes modelos pela comunidade científica já bastante elevado em diversas áreas, embora não se verifique no topo da lista a avaliação imobiliária, consegue-se já um volume considerável de publicações sobre o tema, principalmente de 2010 até à atualidade, de uma forma crescente.

Será, portanto, expectável que os modelos automatizados de avaliação (MAA) comumente utilizados (modelos hedónicos de regressão múltipla) com as limitações amplamente conhecidas, venham a ser rapidamente substituídos pela utilização deste novos modelos de RNA (Vasques, 2017), já em uso mas essencialmente noutras áreas por colossos como, a Amazon (Avrutskiy, 2020; Oladipupo, 2010) nas designadas Redes Neuronalis Profundas (RNP) e Redes Convolucionais (RNC) (*DNN – Deep Neural Nets, CNN – Convolutional Neural Nets*), assim como a Google, que inclusive fornece atualmente cursos gratuitos online sobre aprendizagem computacional (AC) (*ML - Machine Learning*). Acredita-se que serão estes colossos, bem como eventualmente outros como o Facebook e o Youtube, que irão intervir diretamente no setor imobiliário, de forma direta como investimento, ou nas ferramentas essenciais para que as Proptech desenvolvam a sua atividade, pois são estas empresas que possuem todos os dados necessários para o fazer.

Alguns casos de estudo podem ser apontados, onde se concluiu o potencial da aplicação de RNA na avaliação imobiliária com erros abaixo dos 10% e performances que ultrapassavam os modelos hedónicos (Abidoeye & Chan, 2018; Shybaiev et al., 2019). Numa revisão a vários autores, houve quem concluísse que as RNA possibilitavam uma avaliação livre de erro (Peter et al., 2020). Alguns investigadores chineses acreditam que as RNA serão o futuro nesse âmbito (LIU et al., 2011; Shi, 2009; Zhou et al., 2018) tendo Shi (2009) inclusive fundamentando todo o desenvolvimento matemático na sua exposição.

Será nesse âmbito que se desenvolverá o estudo desta dissertação, aplicando as tecnologias mais recentes e cujo desenvolvimento tem sido bastante acentuado nos últimos anos, reforçando e demonstrando a eficácia e facilidade no uso corrente das RNA na avaliação imobiliária.

2.3.3 POTENCIAL FUTURO DAS RNA NO SETOR IMOBILIÁRIO

Pela leitura dos capítulos anteriores, pode constatar-se que o potencial de utilização das RNA no setor imobiliário é já uma realidade embora ainda pouco testado e utilizado, que requer mais investigação e investimento, para além do que já tem vindo a ser realizado nas PropTech. Se estas, por conseguinte, irão utilizar massivamente redes neuronais artificiais, ou outros modelos, é algo que se aguarda para ver. Embora tudo aponte, que serão estas empresas a cimentar e por fim demonstrar massivamente os resultados e fiabilidade destes modelos, bem como alargar criativamente a sua aplicabilidade no setor, alimentadas pelo *'Big Data'*.

Verificam-se já inúmeras empresas cujo *core business* é a avaliação imobiliária, mas será que estas estão a investir em modelos preditivos recorrendo a RNA?

Numa revisão à publicação de “PROPTECH 3.0” de 2017, Baum, Saull e Braesemann lançaram uma mais recente publicação onde reavaliam o futuro do setor imobiliário. Nesse documento é publicado que “(...)a crescente disponibilidade de dados permitiu uma modelação quantitativa mais financeira, e o programa de avaliação imobiliária e os sistemas de gestão de propriedades e portfólio, baseiam-se agora em computadores e tecnologia (informática).”.

Seguidamente afirmam que em termos de investimento “A métrica de sucesso usada pela maioria das empresas é o lucro. No entanto, no mundo da tecnologia, onde o crescimento é o melhor indicador de valor futuro, mais investimento significa avaliações mais altas e maior hipótese de retorno do investidor.”. Ao que investigam e demonstram os investimentos mais recentes efetuados em PropTech e a facturação das empresas *'Unicórnio'* ligadas ao setor, (*Startup's* tecnológicas, termo cunhado em 2013 por Aileen Lee (1970), fundadora da CowboyVentures, referindo-se a empresas que atingem o valor de 1.000 milhões de dólares sem estarem cotadas na bolsa.), nos principais países investidores, constatando a quantidade de informação (*dados*) que estas empresas de PropTech necessitam de *'capturar'* para aumentar a sua eficácia, sendo que esta informação, na sua maioria é privada e está incluída em documentos analógicos, não tão fáceis de aceder.

Acresce que “(...) aqueles que estão a usar sistemas digitalizados provavelmente fazem-no por meio de programa genérico, com 60% dos executivos a informar que nas suas empresas ainda usam tabelas como ferramenta principal para relatórios, 51% para avaliação e análise de fluxos de caixa e 45% para orçamento e previsão. E Isso impede o acesso ao chamado *'Big Data'*.”.

Por fim, dedicam um capítulo inteiro à inteligência artificial e aprendizagem computacional (*'AI and Machine Learning'*), afirmando que as RNA produzem atualmente resultados muito fiáveis e que estes modelos “(...) estão cada vez mais a ser utilizados em análises e previsões imobiliárias, encontrando correlações além do que é possível por meio de modelos de regressão tradicionais.”.

Para além das RNA, apontam o Processamento de Linguagem Natural (PLN) e as técnicas de Aprendizagem Computacional (AC) conjuntamente, como algo com forte aplicabilidade futura no setor. Outra vertente com o mesmo potencial será o desenvolvimento da Visão Computacional (VC) que permite o reconhecimento facial e qualquer tipo de imagem. Neste capítulo os autores concluem que as RNA cresceram 60 vezes só entre 2013 e 2016 e que o erro verificado em VC foi de apenas 3,5% comparado com o erro humano de 5%, idem para ML que desde 2017 baixou dos 5%, e que com a velocidade a que a AI está a ser desenvolvida, ultrapassou já claramente a nossa capacidade de regular e legislar a manipulação de dados privados, criando-se um vazio tecnológico que muitas empresas vieram a explorar:

- *“Muitas das práticas associadas à capitalização dessas oportunidades recentemente perfeccionadas desafiaram as normas sociais (de privacidade) ... e são contestadas como violações de direitos e leis” (Zuboff, 2015: 85). Essas práticas permitiram que o poder fosse usurpado por fornecedores de tecnologia, enquanto o uso dos seus serviços tornou-se um requisito para inclusão social.”*

(Andrew Baum et al., 2020).

3. RECOLHA DE DADOS E CRIAÇÃO DE BASES DE INFORMAÇÃO PREDIAL

3.1 O método de recolha de dados

A recolha de dados de informação predial tem forçosamente de recorrer a um método expedito, dada a quantidade massiva de dados necessários, para aplicar os métodos de mercado. Dependendo do grau de rigor e da tipologia de dados pretendida, a tarefa poderá ser mais ou menos simplificada, pretendendo-se uma escala razoável para um estudo de validação como o da presente dissertação.

Atendendo a que inúmeros serviços imobiliários têm vindo a crescer na *Internet*, nomeadamente na comercialização de imóveis, é pelos mesmos disponibilizada uma informação extensa, quantitativamente e não qualitativamente ou suficientemente detalhada, sobre o mercado imobiliário. Como exemplo, podem apresentar-se duas situações distintas, onde numa se constata dados que sugerem maior precisão, e noutra dados de um imóvel onde não distingue área bruta de útil, bem como isento de inúmeros outros dados úteis para a sua classificação – Figura 3.1 e Figura 3.2 respetivamente.



330 000 € id. 120611692-8

Venda 72 m² 1 3

Área Bruta Privativa m ²	72	Área Útil m ²	62
Total do lote m ²	246	Ano de construção	1961
Quartos	3	WCs	1
Piso	R/C	Estacionamento	Não
Elevador	Não	Eficiência energética	D

Figura 3.1 – Exemplo de dados de imóvel detalhados



465 000 € Apartamento T3

COMPRAR Av. Óscar Monteiro Torres, Areiro, Lisboa

Dados do imóvel

ESTADO Recuperado	ÁREA ÚTIL 102 m ²	ÁREA BRUTA 102 m ²	ANO DE CONSTRUÇÃO 1956
PUBLICADO EM 03/08/2021	CERTIFICAÇÃO ENERGÉTICA	VISUALIZAÇÕES 464	CLIQUE S 139

Figura 3.2 – Exemplo de dados de imóvel pouco detalhados

Embora alguns dados sejam de um rigor não mensurável, pois são introduzidos pelos proprietários, estes podem ser triados e tratados de modo a possibilitar a verificação de tendências de crescimento, efetuar análises comparativas, variações e estimação de valor, entre outras.

A existência de informação útil para várias utilizações, em larga escala na *Internet*, e acumulada ao longo das últimas décadas, fez com que, associado a interesses científicos e de mercado, surgissem meios rápidos de recolha destes dados, como diferentes programas informáticos dedicados a essa função, e que abrange naturalmente a área do imobiliário. Dada a dificuldade em adquirir informação, maioritariamente privilegiada, relativa à promoção e venda de imóveis, e existindo por vezes um diferencial grande entre ambos, esta é não só confidencial como tem um valor de mercado considerável.

Por outro lado, com o desenvolvimento destas ferramentas, designadas por ‘coletores de dados’ ou Web Scrapers, foram posteriormente surgindo mecanismos de defesa por parte dos criadores de páginas da *Internet*, dificultando o acesso aos dados públicos destes serviços imobiliários. Muitos destes mecanismos são conhecidos, como os verificadores de operador humano (verificação de imagens, números, horizontalidade de figuras etc.), ou apenas por análise de incidências de pesquisa e limitação de acesso por bloqueamento de IPs (número de endereço único do dispositivo).

No âmbito deste estudo, foi, por conseguinte, necessário:

- Identificar um fornecedor de dados imobiliários que garantisse um volume considerável de imóveis para constituir a amostra. E que simultaneamente, fornecesse junto dessa informação, o maior número de dados relativos a imóveis, de uma forma compartimentada e organizada;
- identificar um programa de ‘colheita’ que ultrapassasse a maioria dos mecanismos de defesa da página do fornecedor escolhido;
- utilizar uma VPN para não permitir o rastreamento por IP;
- sistematizar e adaptar (programar) o programa de colheita para os diversos campos dos imóveis a coletar.

Refira-se todos os dados extraídos são públicos, e na legislação portuguesa, não há naturalmente qualquer referência a uma proibição da utilização destes mecanismos para recolha de dados públicos e exibidos nas páginas da *Internet*. No entanto, os mesmos foram utilizados de forma confidencial e sem identificar os imóveis unitariamente, com o propósito exclusivo de fazerem parte de uma amostra, para aplicação e comparação de resultados, conforme definido nos objetivos desta dissertação. O Programa WebHarvy[®] escolhido para este fim (©Copyright SysNucleus) dedica uma página à legalidade do procedimento e da utilização do programa, alertando apenas para questões éticas, relacionadas com o fim ou propósito na utilização dos dados coletados (<https://www.webharvy.com/articles/is-scraping-legal.html>).

O WebHarvy[®] foi, por conseguinte, programado para a recolha de dados nos *sites* pretendidos, fazendo sair uma tabela em formato de base de dados, com os campos escolhidos, e que foi aberta e tratada seguidamente no Microsoft Excel[®] – Figura 3.3.

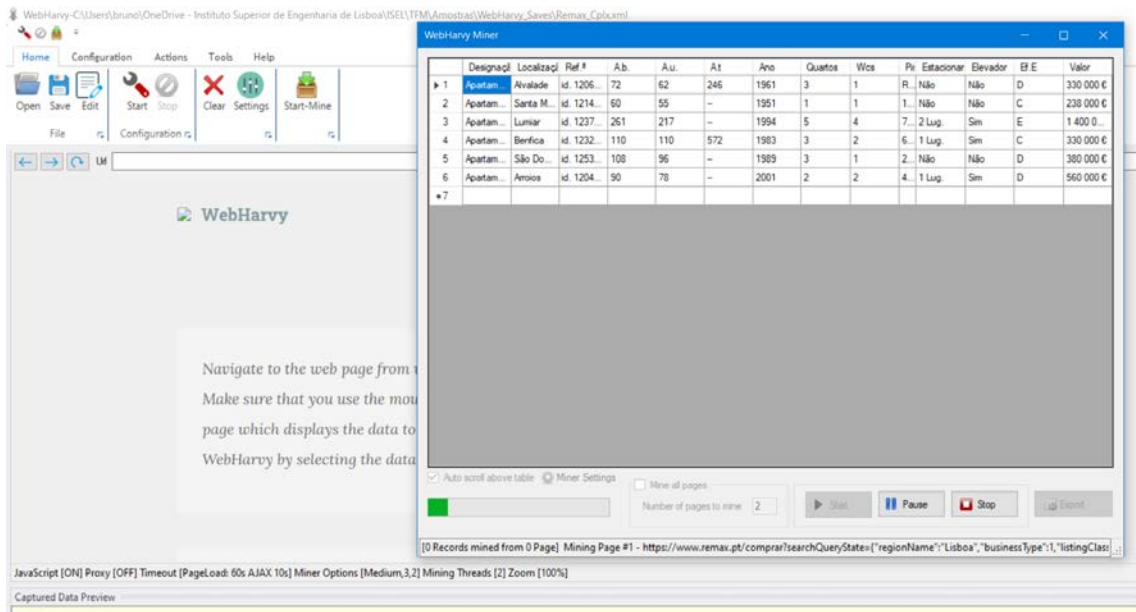


Figura 3.3 – Programa WebHarvy® a coletar dados

Abaixo demonstram-se algumas tabelas agrupadas, como exemplos resultantes de dados extraídos de páginas de imobiliárias oficiais, especializadas na promoção de imóveis – Figura 3.4.

ID	Ref.#	Tipologia	Designação	Freguesia	A.b.	A.u.	A.t.	Ano	Wcs	Piso	Estacion	Elevad	Estad	C.En.	Valor
1	121701274-188	T3	Apartamento - Lisboa	Santa Maria Maior	103	90	--	1951	2	3ª Andar	Não	Não		C	525.000
2	120261696-24	T6	Apartamento - Lisboa	Avenidas Novas	176	176	--	1986	2	2ª Andar	Não	Não		C	750.000
3	120611191-1221	T2	Apartamento - Lisboa	Penha de França	80	68	--	1950	1	R/C	Não	Não		D	270.000
4	121701274-187	T2	Apartamento - Lisboa	Carnide	97	87	--	2000	3	4ª Andar	1 Lug.	Sim		C	385.000

ID	Ref.#	Tipolog.	Designação	Freguesia	A.b.	A.u.	A.t.	Ano	Wc	Pisc	Estacion	Eleva	Estado	C.En.	Valor
1	CO429-00030	T1	T1,Benfica	Benfica	50	50	--	1985	1				Bom	C	325.000,00
2	CO223-03456	T5	T5,Duplex,Manvila	Manvila,Bairro,das,Amendoeira	165	150	--	1985	2				Média	B	315.000,00
3	CO418-GORE-00096	T5	Apartamento	Lisboa	55	55	--	1928			0 Rua			E	211.000,00

ID	Ref.#	Tipolog	Designação	Freguesia	A.b.	A.u.	A.t.	Ano	Wcs	Pisc	Estacion	Elevad	Estado	C.En.	Valor
1	1 / REF: 113210007	T3	Apartamento 3 quartos	Santo António	115	100			2				Usado	E	325.000,00
2	2 / REF: 113210009	T5	Apartamento 5 quartos	Santa Maria Maior	174	151			3				Novo	C	935.000,00
3	3 / REF: 334210007	T3	Apartamento 3 quartos	Lumiar	99,75				2		1		Usado	C	485.000,00

Figura 3.4 – Recolha de dados com WebHarvy® de páginas de Imobiliárias Oficiais

De seguida demonstram-se tabelas de dados agrupados na Figura 3.5, extraídos de páginas dedicadas à promoção imobiliária livre, inserida pelos proprietários ou por qualquer empresa do ramo da promoção imobiliária.

A	B	C	D	E	F	G	H	I
1	Tip	Freg	Val	Estado	Aut	Ref	CEn	Texto
2	Apartamento T1	Alameda (Alto do Pina), Areeiro, Lisboa	260 000 €	ESTADORrecuperado	ÁREA ÚTIL45 m²	PUBLICADO EM		Venha con
3	Apartamento T2	Benfica, Lisboa	289 900 €	ESTADORenovado	ÁREA ÚTIL60 m²	APA_11566		<div class="energyCla Apartamen
4	Apartamento T4	Alcântara, Lisboa	228 000 €					Apartamen
5	Apartamento T3	Morais Soares (Penha de França), Penha d	499 900 €	ESTADOUssado	ÁREA ÚTIL136 m²	ANO DE CONSTRUÇÃO		Apartamen
6	Apartamento T4	Avenidas Novas, Lisboa	1 150 000 €					Apartamen
7	Apartamento T3	Corte Inglês (São Sebastião da Pedreira),	1 460 000 €	ESTADOUssado	ÁREA ÚTIL100 m²	PUBLICADO EM		Lisboa cen

A	B	C	D	E	F	G	H	I	J	
1	Desc	Tip	Areab	AreaUt	Ano	WC	CEn	Texto	Data	Ref
2	Apartamento T2 para Venda	T2	83 m2	---	1969	---	D	Apartamento remodelad	Inserido a: 27-03-2021	Referência
3	Apartamento T1+1 para Venda	T1+1	63 m2	57 m2	---	---	---	Apartamento T1+1 Duple	Inserido a: 05-12-2020	Referência
4	Apartamento T2 para Venda	T2	66 m2	60 m2	1994	1	C	Apartamento T2 em pré	Inserido a: 02-03-2021	Referência
5	Apartamento T1 para Venda	T1	62.00 m2	57.73 m2	1950	1	D	Apartamento T1 Principe	Inserido a: 15-11-2020	Referência
6	Apartamento T1 para Venda	T1	50 m2	30 m2	---	---	D	Conta a lenda, que o edif	Inserido a: 09-01-2021	Referência
7	Apartamento T2 para Venda	T2	97.00 m2	92.00 m2	---	2	B-	Apartamento T2 remodel	Inserido a: 20-06-2020	Referência
8	Apartamento T1 para Venda	T1	82.25 m2	73 m2	---	---	---	Apartamento T1, de Novi	Inserido a: 28-01-2021	Referência
9	Apartamento T3 para Venda	T3	156.1 m2	---	2021	---	A	Lisboa>Belem : Luxuoso	Inserido a: 10-03-2020	Referência
10	Apartamento T1 para Venda	T1	---	86 m2	---	---	B-	Apartamento T1 com qu	Inserido a: 21-11-2020	Referência

A	B	C	D	E	F	G	H
1	Desc	Valor	Tip	T	Area	Piso	Texto
2	Apartamento T2 em Graça, São Vicente	590.000€	Garagem incluída	T2	88 m² constru	3º andar com elevador	Apartamento T2 situado n
3	Apartamento T2 na rua Neves Ferreira, 1	325.000€	Garagem incluída	T2	75 m² constru	4º andar com elevador	Vendo ou permuto por T1
4	Moradia geminada na travessa Olival à G	175.000€		T2	90 m² constru		Moradia na zona de Santa
5	Apartamento T2 em Santa Engrácia, São	525.000€		T2	73 m² constru	1º andar	Localizado entre a Estaçãc
6	Apartamento T2, Av. Casal Ribeiro, Salda	610.000€	Garagem incluída	T2	73 m² constru	9º andar com elevador	Apartamento T2 com 73 m
7	Apartamento T2, Av. Casal Ribeiro, Salda	625.000€	Garagem incluída	T2	81 m² constru	9º andar com elevador	O Oásis 28! Apartamento 1

Figura 3.5 - Recolha de dados com WebHarvy® de páginas de promoção imobiliária livre

Do acima foi possível concluir que destes dois segmentos de mercado, o primeiro é mais estruturado, embora limitado em termos quantitativos, enquanto o segundo menos estruturado, com menos campos de informação e maiores lacunas de informação, qualitativamente pior mas quantitativamente muito superior.

Agregar a recolha de várias páginas em qualquer dos segmentos, implica o risco de se juntar inúmeros imóveis da mesma fonte ou iguais, principalmente com o último segmento por envolver maior número de dados.

Tendo a possibilidade de recolher estes dados com significativa rapidez, permitiu esta análise, bem como uma tomada de decisão, quanto à escolha das páginas, organização dos campos, e dados elegíveis.

3.2 Recolha, tratamento e finalização da amostra

A especificidade e dimensão do mercado português, com imóveis de natureza muito diversificada, abrangendo outras questões, como a de atravessar atualmente uma vaga de reabilitação urbana significativa, obrigou a optar por reduzir o número de campos da amostra. Deste modo foi possível que a mesma tivesse uma dimensão apreciável e fosse simultaneamente bem estruturada e simples de tratar.

Obteve-se assim, no Concelho de Lisboa, um total de 18.401 imóveis coletados entre Março e Abril de 2021.

O tratamento dos dados foi efetuado numa base de dados em *Microsoft Excel*®.

Por tratamento refere-se a um conjunto de passos essenciais, para que a amostra represente uma base sólida para a sua utilização, quer no modelo hedónico de regressão linear, quer no modelo de RNA.

Esse tratamento teve duas fases:

1ª Fase – A organização e filtragem dos dados relevantes e fidedignos;

2ª Fase – A análise estatística dos valores para remoção de dados não padronizados designadamente ‘outliers’ da amostra, e reestruturação dos campos de distribuição dos dados da amostra;

A primeira fase consistiu na:

a) Compilação dos dados das várias fases de colheita – Figura 3.6;

Desig	Estado	Area	Val	Texto	Freguesia
Apartamento T0 em Prazeres, Estrela	Usado	25	129.500	Estúdio junto à beira-	Estrela
Apartamento T1 na travessa Rebelo da	Usado	40	185.000	T1 com cozinha Equi	Arroios
Apartamento T0 em Pena, Arroios	Usado	29	200.000	Apartamento T0 com	Arroios
Apartamento T2 em Olivais, Lisboa	Usado	54	165.000	Apartamento T2 trans	Olivais
Apartamento T3 em Olivais, Lisboa	Usado	80	189.500	T3 Olivais Sul a 13 m	Olivais
Apartamento T1 na rua do Vale de St.º	Usado	59	190.000	Apartamento renovad	São Vicente
Apartamento T2 em Alfama - Sé, Santa	Usado	35	175.000	Apartamento T2 local	Santa Maria Maior
Apartamento T0 em São Vicente de For	Usado	35	175.000	Excelente Apartamen	São Vicente
Apartamento T2 na rua Pedro de Queir	Usado	50	155.000	Excelente apartament	Lumiar
Apartamento T3 em Olivais. Lisboa	Usado	85	194.500	Apartamento de 4 ass	Olivais

Figura 3.6 – Compilação de dados coletados

A necessidade de segmentar a coleta de dados adveio das proteções implementadas pela página à robotização na sua operação.

b) Transformação dos dados ‘rawdata’ alfanuméricos saídos do programa de colheita, para dados de texto e numéricos qualificáveis e quantificáveis, respetivamente, na tabela de Excel®, separando em colunas os campos de Freguesia, Estado, Área e Valor (Campos escolhidos inicialmente para o programa WebHarvy® coletar) – Figura 3.7.

Freguesia	Estado	Área Bruta	Valor (€)	Valor /m2	Texto
Ajuda	Usado	127	379.000 €	2.984 €	A Inicial J convida-o a conh
Avenidas Novas	Usado	94	390.000 €	4.149 €	A Inicial J convida-o a conh
Olivais	Usado	68	200.000 €	2.941 €	A Inicial J convida-o a conh
Olivais	Usado	68	200.000 €	2.941 €	A Inicial J convida-o a conh
Areiro	Usado	171	620.000 €	3.626 €	A Inicial J convida-o a conh
Benfica	Usado	42	239.000 €	5.690 €	A intervenção arquitetónica
Estrela	Usado	100	470.000 €	4.700 €	A Lapa é considerado um d
Estrela	Usado	228	1.850.000 €	8.114 €	A Lapa é um dos bairros de
Misericórdia	Usado	28	215.000 €	7.679 €	A localização deste apartan
Santo António	Usado	54	647.000 €	11.981 €	A localização do Glória é fa
Santo António	Usado	50	615.000 €	12.300 €	A localização do Glória Stuc
Santo António	Usado	28	535.000 €	19.107 €	A localização do Glória Stuc
Santo António	Usado	54	647.000 €	11.981 €	A localização do Glória Stuc
Santa Maria Maior	Usado	69	515.000 €	7.464 €	A localização não podia ser
Santa Maria Maior	Usado	55	450.000 €	8.182 €	A localização não podia ser
Avenidas Novas	Usado	58	395.000 €	6.810 €	A localização que sempre d
Marvila	Usado	70	225.000 €	3.214 €	A localização: O imóvel enc
Estrela	Novo	67	440.000 €	6.567 €	A Localização: Junta de

Figura 3.7 – Tabela de dados rawdata tratados

A **segunda fase** determina a fiabilidade dos dados para utilização nos modelos de previsão, recorrendo à criação de tabelas estatísticas dos dados coletados, e que permitem o apuramento das situações que carecem de análise e tratamento.

A existência de situações que desfavorecem os modelos de previsão, como sejam situações de contrassenso, obriga à análise da representatividade dos dados.

Como exemplo do acima referido, tem-se Freguesias com imóveis ‘usados’ mais caros que imóveis ‘novos’, por questões essencialmente relacionadas com a localização, mas também por outros fatores que interferem de forma diferenciada sobrevalorizando imóveis mais antigos. São estes exemplos de não representatividade que foram detetadas com recurso a tabelas elaboradas em Excel como as do **Anexo I**.

Dada a falta de informação que respeita às características intrínsecas do produto, e atendendo à amplitude de valores por metro quadrado e por freguesia, foi criada uma segmentação em 5 percentis de valor, classificados como Qualidade e Conservação (Q&C) do imóvel. Esta segmentação permite uma maior representatividade do produto e por conseguinte uma classificação mais fidedigna no modelo.

Da Tabela 3.1 seguinte, pode observar-se que os imóveis novos não só desvirtuam em termos de valor a amostra, dada também a dispersão de valores dos imóveis usados, como quantitativamente não são suficientemente representativos (11%), o que levou à decisão de os **remover da amostra**.

Mínimos / Máximos	Usado										Novo		NºImóveis	
	MED	MAU	RAZ	BOM	MBOM	EXCLUIDOS	Usado	Novo						
Ajuda	776 €	3.621 €	3.621 €	6.466 €	6.466 €	9.310 €	9.310 €	12.155 €	12.155 €	15.000 €	3.413 €	5.066 €	368	21
Alcântara	1.228 €	3.642 €	3.642 €	6.056 €	6.056 €	8.470 €	8.470 €	10.884 €	10.884 €	13.298 €	4.231 €	14.124 €	421	143
Alvalade	1.835 €	3.598 €	3.598 €	5.362 €	5.362 €	7.125 €	7.125 €	8.889 €	8.889 €	10.652 €	4.828 €	10.500 €	428	108
Areiro	2.324 €	3.530 €	3.530 €	4.737 €	4.737 €	5.943 €	5.943 €	7.150 €	7.150 €	8.356 €	3.942 €	6.918 €	369	25
Arroios	939 €	3.055 €	3.055 €	5.172 €	5.172 €	7.289 €	7.289 €	9.405 €	9.405 €	11.522 €	3.465 €	9.851 €	1782	176
Avenidas Novas	2.063 €	4.188 €	4.188 €	6.314 €	6.314 €	8.439 €	8.439 €	10.565 €	10.565 €	12.690 €	2.245 €	10.000 €	1328	229
Beato	1.233 €	3.387 €	3.387 €	5.540 €	5.540 €	7.693 €	7.693 €	9.847 €	9.847 €	12.000 €	3.133 €	5.194 €	232	4
Belém	2.155 €	4.093 €	4.093 €	6.031 €	6.031 €	7.970 €	7.970 €	9.908 €	9.908 €	11.846 €	4.227 €	14.839 €	487	104
Benfica	1.200 €	2.713 €	2.713 €	4.225 €	4.225 €	5.738 €	5.738 €	7.250 €	7.250 €	8.763 €	3.684 €	5.313 €	542	9
Campo de Ourique	1.977 €	4.462 €	4.462 €	6.947 €	6.947 €	9.433 €	9.433 €	11.918 €	11.918 €	14.403 €	4.361 €	9.973 €	664	46
Campolide	1.303 €	3.042 €	3.042 €	4.781 €	4.781 €	6.520 €	6.520 €	8.259 €	8.259 €	9.998 €	5.167 €	14.312 €	381	63
Carnide	1.877 €	2.730 €	2.730 €	3.583 €	3.583 €	4.436 €	4.436 €	5.288 €	5.288 €	6.141 €	3.952 €	5.529 €	88	15
Estrela	880 €	3.599 €	3.599 €	6.317 €	6.317 €	9.036 €	9.036 €	11.755 €	11.755 €	14.474 €	4.433 €	11.692 €	1142	134
Lumiar	1.800 €	2.823 €	2.823 €	3.847 €	3.847 €	4.870 €	4.870 €	5.894 €	5.894 €	6.917 €	2.131 €	3.780 €	525	21
Marvila	1.500 €	2.589 €	2.589 €	3.679 €	3.679 €	4.768 €	4.768 €	5.858 €	5.858 €	6.947 €	3.308 €	7.273 €	258	111
Misericórdia	1.786 €	5.929 €	5.929 €	10.071 €	10.071 €	14.214 €	14.214 €	18.357 €	18.357 €	22.500 €	5.031 €	21.408 €	1376	158
Olivais	1.674 €	2.768 €	2.768 €	3.861 €	3.861 €	4.955 €	4.955 €	6.049 €	6.049 €	7.143 €	2.811 €	3.727 €	374	29
Parque das Nações	2.045 €	5.031 €	5.031 €	8.017 €	8.017 €	11.002 €	11.002 €	13.988 €	13.988 €	16.974 €	5.235 €	19.588 €	600	120
Penha de França	1.525 €	2.808 €	2.808 €	4.090 €	4.090 €	5.372 €	5.372 €	6.654 €	6.654 €	7.937 €	4.348 €	5.928 €	950	13
Santa Clara	1.364 €	2.309 €	2.309 €	3.255 €	3.255 €	4.200 €	4.200 €	5.146 €	5.146 €	6.091 €	€	€	126	0
Santa Maria Maior	1.870 €	4.889 €	4.889 €	7.907 €	7.907 €	10.926 €	10.926 €	13.945 €	13.945 €	16.964 €	5.000 €	11.036 €	1595	73
Santo António	1.468 €	6.106 €	6.106 €	10.744 €	10.744 €	15.382 €	15.382 €	20.020 €	20.020 €	24.658 €	3.570 €	12.324 €	1030	237
São Domingos de Benf.	792 €	2.180 €	2.180 €	3.567 €	3.567 €	4.955 €	4.955 €	6.342 €	6.342 €	7.729 €	5.441 €	8.209 €	482	2
São Vicente	372 €	2.742 €	2.742 €	5.112 €	5.112 €	7.482 €	7.482 €	9.852 €	9.852 €	12.222 €	3.826 €	8.200 €	904	108
													16452	1949
													89%	11%

Tabela 3.1 – Tabela de Máximos e Mínimos por Estado, Qualidade e Conservação, por Freguesia

A criação de tabelas com dispersão de valores médios por metro quadrado por segmentos de áreas, bem como por freguesia e qualidade e conservação (explicado no capítulo seguinte) – ver **Anexo II**, permitiu o apuramento de situações singulares como imóveis com valores desprezáveis segundo critérios de Chauvenet, considerando-se os mesmos espúrios ou ‘outliers’ e, por conseguinte, também retirados da amostra.

Na finalização da amostra, esta ficou reduzida a 10.465 imóveis distribuídos por 24 Freguesias, correspondendo a uma redução de quase 43% de todos os dados coletados.

3.3 Análise dos dados para aprendizagem e teste

Das 10.465 incidências, que constituem 100% da amostra, 80% (8.372) foram destinadas para utilização nos métodos de estimação por RNA e RL (regressão linear), e 20% (2.093) para teste de ambas as técnicas.

Importa referir e recordar, que os valores ‘objetivo’ da amostra correspondem a valores de oferta (VO) e não valores de venda.

Em Portugal a única entidade que trata e publica indicadores relativos a Preços de Venda Transacionados (PVT) é a *Confidencial Imobiliário*, dados esses que a servir de amostra para este modelo de estimação, potenciaria consideravelmente o interesse.

Por conseguinte, e sumarizando, de modo a aumentar a coerência da amostra e consequentemente a eficácia dos modelos de previsão, foram tomadas duas medidas fundamentais:

- a) Retirada de todos os dados espúrios, incluindo os imóveis em mau estado ou devolutos, e todos os imóveis com áreas menores que 35m² e maiores que 400m². Para suprir as situações acima identificadas no capítulo 3.2, foram também retirados todos os imóveis classificados com o estado 'Novo';
- b) No que respeita à falta de informação quanto à **qualidade** dos imóveis e à **amplitude de valores por metro quadrado por freguesia**, foi criado um novo campo de classificação dos imóveis, de modo a segmentar cada Freguesia em 5 partes iguais em termos de valorização por metro quadrado. Denominou-se este campo de 'Qualidade e Conservação', classificados os imóveis como '*Mediocres*', '*Mau*', '*Razoável*', '*Bom*' e '*Muito Bom*'. Com a criação desta nova classificação e atendendo ao exposto no ponto anterior, foi eliminado o campo 'Estado' dos imóveis da tabela que compunha a amostra.

Do acima, resultou a tabela do **Anexo II**.

A amostra total foi seguidamente sumarizada e analisada atendendo essencialmente às distribuições quantitativas e de valor médio por freguesia e qualidade e conservação. Com recurso novamente ao Excel[®], tendo sido possível constatar o seguinte (ver **Anexo III**):

- a) A aleatoriedade da oferta em termos quantitativos variou entre freguesias com um desvio médio de **2,502%** à divisão 80-20% e naturalmente entre a sua qualidade e conservação (Q&C) (segmento de preço) com um desvio médio de **1,045%**. As situações mais desviantes são o Areeiro cujos imóveis desviaram-se em média 13% na distribuição quantitativa 80-20%. Estes desvios deveriam provocar deficiências na aprendizagem e no teste. Poderia especular-se que, nos casos em que o valor médio de imóveis numa dada freguesia for inferior na parcela de dados destinado ao treino, comparativamente com a média da parcela destinada a teste, os modelos de mercado deveriam ter tendencialmente valores preditivos mais afastados dos reais para valores inferiores. Na situação inversa, também mais afastados para valores superiores. Mas na realidade isto não acontece, não existindo, por conseguinte, uma linearidade neste critério de pensamento.
- b) A aleatoriedade na divisão entre os imóveis selecionados para aprendizagem dos modelos, introduziu também diferenciais entre os valores médios da amostra, da parcela para aprendizagem e da parcela para teste, com um desvio médio absoluto de **4,205%** entre freguesias, e **1,203%** entre diferentes estados de Q&C. Verificaram-se freguesias em que esse diferencial atingiu os **-10%** e **14,5%**, nas freguesias de Belém

e Campolide respetivamente. Nesses casos específicos, Belém deveria ser prejudicado na aprendizagem das ferramentas de avaliação, e Campolide beneficiado, o que também não acontece.

A análise destas tabelas não demonstra uma coerência na correlação entre a quantidade de dados na aprendizagem e no teste, bem como nos desvios entre valorizações médias das freguesias na mesma proporção de divisão da amostra.

Verifica-se, porém, que a tendência no comportamento dos desvios é idêntica por freguesias, mas não por grau de Q&C, como se pode verificar pelos gráficos da Figura 3.8 e Figura 3.9 seguintes.

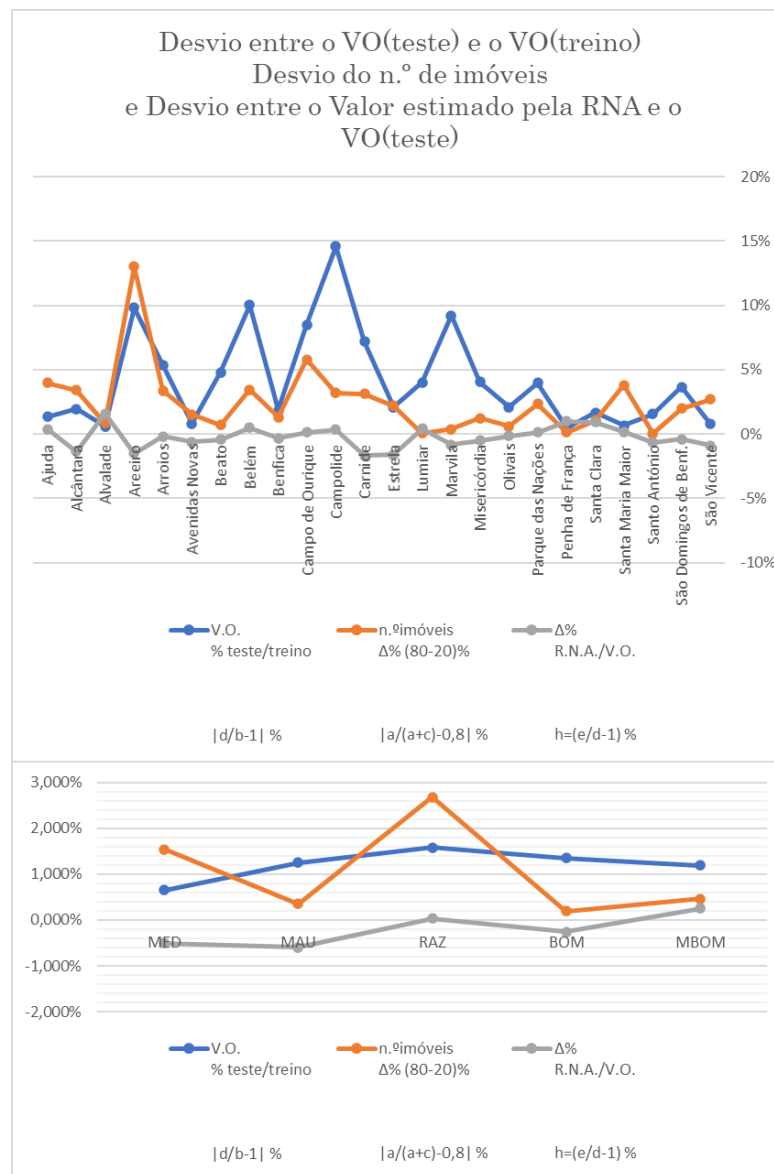


Figura 3.8 – Gráficos comparativos nas relações VO (test/train), N°Im.(test/train), ΔRNA/VO

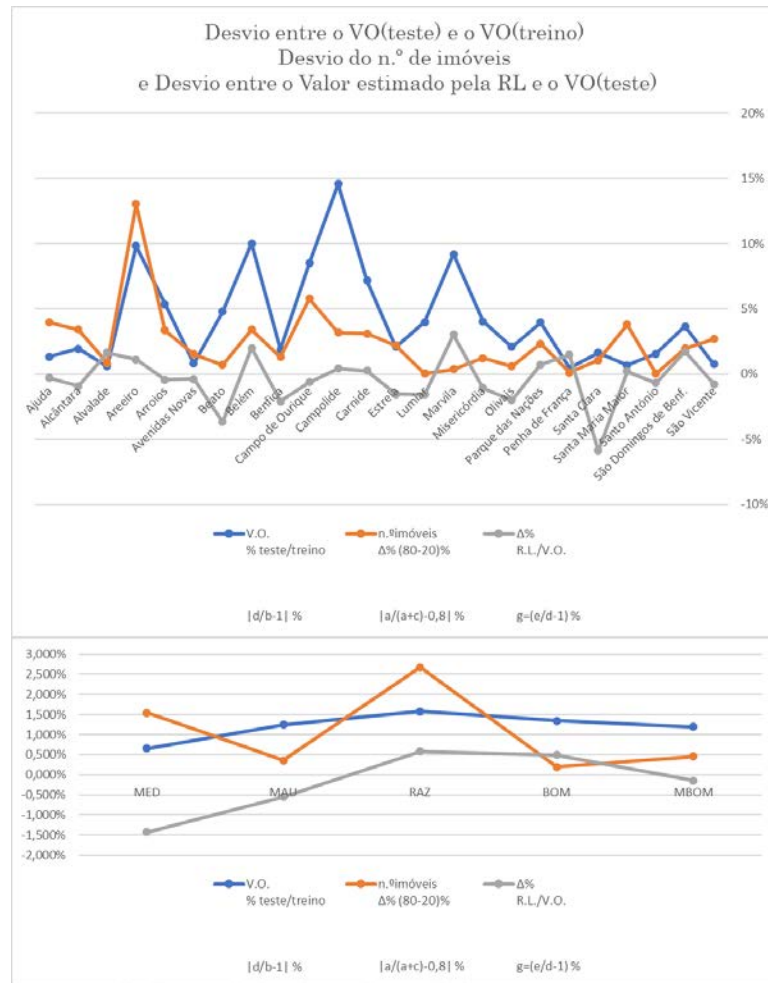


Figura 3.9 – Gráficos comparativos nas relações VO(test/train), N°Im.(test/train), ΔRL/VO

4. A ESTIMAÇÃO IMOBILIÁRIA

No que refere ao estudo de avaliação imobiliária aqui em dissertação, foram ensaiados alguns modelos distintos de RNA, aplicados à mesma amostra, procurando obter os melhores resultados e tentando abranger o que a literatura aponta como mais adequado a este tipo de estimação. São expostos neste capítulo os critérios adotados, bem como comentados os resultados obtidos para diferentes modelos.

De forma idêntica, mas sem aprofundar um método de mercado já consolidado (Rosen, 1974; Sirmans et al., 2005), são também definidos os critérios para o modelo hedónico de regressão, comentando os resultados obtidos.

4.1 O Modelo Hedónico de Regressão Linear (MHRL)

4.1.1 INTRODUÇÃO AO PROGRAMA *REGRESSIT*TM

Para modelos de regressão linear, o Microsoft Excel[®] possui um módulo que permite realizar a regressão linear múltipla com variáveis binárias e reais. No entanto esse módulo limita a regressão linear a 16 variáveis dependentes, e no caso em estudo era necessário incluir 28 variáveis, pelo que foi necessário procurar um aplicativo que não tivesse esta limitação.

Para o caso em estudo foi utilizado um ‘*add-on*’ criado para funcionar com o Excel[®] denominado *Regressit*TM (*RegressIt Excel add-in developers and history - RegressIt*).

4.1.2 O MÉTODO DE REGRESSÃO LINEAR

Para desenvolver o modelo, foi criada uma tabela com os dados da amostra tratados para o efeito, o que se fez recorrendo ao Excel[®]. Assim, foi criada uma tabela com a decomposição binária de cada variável independente, excluindo uma de cada ‘campo’ (uma freguesia e uma classificação de QC). Por conseguinte, cada incidência terá um total de 23 variáveis de freguesias, mais 4 variáveis de QC e uma correspondente à AB, no total de 28 variáveis independentes e uma dependente (Valor de Oferta – VAL) - Figura 4.1.

AJU	ALC	ALV	ARE	ARR	AVE	BEA	BEL	BEM	CAM	CMP	CRN	EST	LUM	MAR	MIS	OLI	PAN	PEF	SAC	SAM	SAN	SDM	MED	MAU	RAZ	BOM	AR	VAL	
Aljuda	Alcântara	Alvalade	Areeiro	Arroios	Avenidas Novas	Beato	Belém	Benfica	Campo de Ourique	Campolide	Carnide	Estrela	Lumiar	Marvila	Misericórdia	Olhais	Parque das Nações	Penha de França	Santa Clara	Santa Maria Maior	Santo António	São Domingos de Benf.	MED	MAU	RAZ	BOM	AR	VAL	
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	58,00	4.483 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	45,00	4.111 €
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	145,00	6.138 €
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	182,00	3.819 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	200,00	3.250 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	92,00	3.424 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	176,00	3.352 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	75,00	2.600 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	54,00	3.426 €	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	55,00	3.900 €	

Figura 4.1 – Tabela de dados para Regressão Linear

Com a utilização do ‘Add-on’ em Excel® – Regressit™, obtém-se um menu criado especificamente para o efeito onde se opera o programa. Assim, são seleccionados os dados da amostra, definindo as variáveis dependentes e independente, bem como os gráficos e análises pretendidas, conforme se pode visualizar na Figura 4.2 e Figura 4.3.

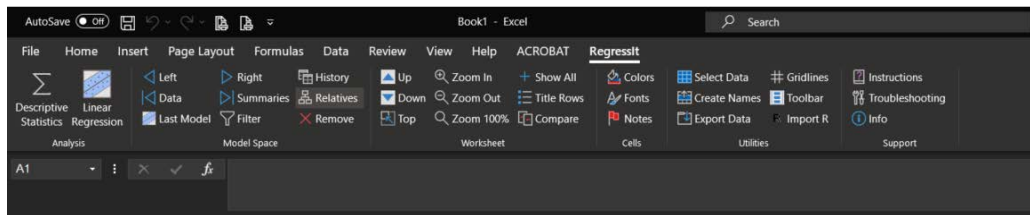


Figura 4.2 – Menu do ‘Add-on’ em Excel® – Regressit™

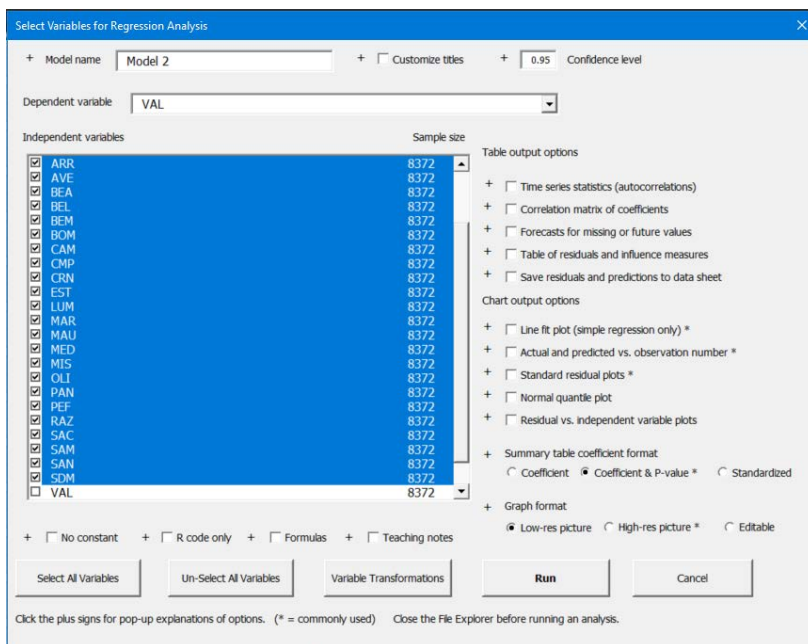


Figura 4.3 – Janela de definição dos dados para Regressão e resultados pretendidos

4.1.3 ANÁLISE DE RESULTADOS

Depois de executado o programa, obtém-se os dados da Figura 4.4 seguinte, onde se pode observar entre outros parâmetros, a função que, aplicada à listagem de teste, devolve os valores estimados pelo método em avaliação:

$$\begin{aligned} \text{VAL} = & 8\,902 - 981.012 \cdot \text{AJU} + 1025 \cdot \text{ALC} + 632.373 \cdot \text{ALV} - 0.015 \cdot \text{AR} - 263.523 \cdot \text{ARE} \\ & + 285.929 \cdot \text{ARR} + 1261 \cdot \text{AVE} - 1184 \cdot \text{BEA} + 1127 \cdot \text{BEL} - 1910 \cdot \text{BEM} - 1542 \cdot \text{BOM} + \\ & 1011 \cdot \text{CAM} + 397.985 \cdot \text{CMP} - 1605 \cdot \text{CRN} + 705.018 \cdot \text{EST} - 1260 \cdot \text{LUM} - 908.693 \cdot \text{MAR} - \\ & 4491 \cdot \text{MAU} - 5842 \cdot \text{MED} + 1364 \cdot \text{MIS} - 1353 \cdot \text{OLI} + 1187 \cdot \text{PAN} - 760.264 \cdot \text{PEF} - 3042 \cdot \text{RAZ} - \\ & 2321 \cdot \text{SAC} + 1089 \cdot \text{SAM} + 1587 \cdot \text{SAN} - 771.864 \cdot \text{SDM} \end{aligned}$$

Este critério é aplicado apenas a 80% da amostra para treino, reservando-se os 20% remanescentes para teste.

Na mesma tabela, podem verificar-se outros parâmetros da regressão, como sejam o coeficiente de determinação \mathbf{R}^2 (*R-squared*), que afere a qualidade do modelo em relação à sua habilidade na estimação correta dos valores da variável dependente. Pelo valor obtido, verifica-se que o modelo linear ajusta-se em 91,8% da variância da variável dependente, a partir das variáveis independentes incluídas no modelo.

Para além do \mathbf{R}^2 , outros parâmetros atestam a qualidade da regressão, como seja o Valor-P (*P-value*), muito próximo de zero, significando uma reduzida probabilidade de a hipótese nula ser falsa (que a estimativa do modelo em teste seja falsa). Idem para os parâmetros T e F (*t-Statistic* e *F-Statistic*) que devolvem a significância dos resultados, apresentando valores elevados, reforçando conjuntamente com o Valor-P a credibilidade do modelo. O Valor T define se cada variável é estatisticamente significativa, e um valor F define se o conjunto de variáveis utilizadas no modelo, são conjuntamente significativas.

Model: Model 1

Dependent Variable: VAL

Independent Variables:

AJU, ALC, ALV, AR, ARE, ARR, AVE, BEA, BEL, BEM, BOM, CAM, CMP, CRN, EST, LUM, MAR, MAU, MED, MIS, OLI, PAN, PEF, RAZ, SAC, SAM, SAN, SDM

Equation:

Predicted VAL = 8.902 - 981.012*AJU + 1025*ALC + 632.373*ALV - 0.015*AR - 263.523*ARE + 285.929*ARR + 1261*AVE - 1184*BEA + 1127*BEL - 1910*BEM - 1542*BOM + 1011*CAM + 397.985*CMP - 1605*CRN + 705.018*EST - 1260*LUM - 908.693*MAR - 4491*MAU - 5842*MED + 1364*MIS - 1353*OLI + 1187*PAN - 760.264*PEF - 3042*RAZ - 2321*SAC + 1089*SAM + 1587*SAN - 771.864*SDM

Regression Statistics: Model 1 for VAL (28 variables, n=8372)

R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std.Dep.Var.	# Fitted	# Missing	Critical t	Confidence
0,918	0,917	519,813	1.808	8372	0	1,960	95,0%

Coefficient Estimates: Model 1 for VAL (28 variables, n=8372)

Variable	Coefficient	Std.Err.	t-Statistic	P-value	Lower95%	Upper95%	VIF	Std. Coeff.
Constant	8.902	44,686	199,222	0,000	8.815	8.990	0,000	0,000
AJU	-981,012	47,317	-20,733	0,000	-1.074	-888,259	1,340	-0,075
ALC	1.025	44,537	23,016	0,000	937,765	1.112	1,405	0,086
ALV	632,373	39,967	15,822	0,000	554,027	710,718	1,567	0,062
AR	-0,015	0,093	-0,155	0,877	-0,198	0,169	1,128	-0,000518
ARE	-263,523	43,789	-6,018	0,000	-349,360	-177,686	1,446	-0,023
ARR	285,929	30,325	9,429	0,000	226,485	345,374	2,591	0,048
AVE	1.261	32,140	39,234	0,000	1.198	1.324	2,344	0,189
BEA	-1.184	59,515	-19,895	0,000	-1.301	-1.067	1,193	-0,068
BEL	1.127	39,095	28,838	0,000	1.051	1.204	1,699	0,118
BEM	-1.910	40,551	-47,112	0,000	-1.990	-1.831	1,539	-0,184
BOM	-1.542	41,319	-37,328	0,000	-1.623	-1.461	3,841	-0,230
CAM	1.011	37,579	26,893	0,000	936,969	1.084	1,676	0,109
CMP	397,985	44,910	8,862	0,000	309,950	486,020	1,400	0,033
CRN	-1.605	69,576	-23,073	0,000	-1.742	-1.469	1,138	-0,077
EST	705,018	32,493	21,698	0,000	641,323	768,712	2,193	0,101
LUM	-1.260	38,013	-33,134	0,000	-1.334	-1.185	1,700	-0,136
MAR	-908,693	52,002	-17,474	0,000	-1.011	-806,756	1,271	-0,062
MAU	-4.491	37,167	-120,835	0,000	-4.564	-4.418	10,326	-1,220
MED	-5.842	37,648	-155,173	0,000	-5.916	-5.768	8,928	-1,457
MIS	1.364	31,562	43,213	0,000	1.302	1.426	2,313	0,207
OLI	-1.353	44,680	-30,290	0,000	-1.441	-1.266	1,407	-0,113
PAN	1.187	37,810	31,386	0,000	1.113	1.261	1,697	0,128
PEF	-760,264	35,135	-21,639	0,000	-829,137	-691,391	1,843	-0,092
RAZ	-3.042	38,196	-79,634	0,000	-3.117	-2.967	7,385	-0,680
SAC	-2.321	64,146	-36,178	0,000	-2.446	-2.195	1,162	-0,123
SAM	1.089	31,359	34,718	0,000	1.027	1.150	2,347	0,167
SAN	1.587	33,301	47,663	0,000	1.522	1.653	2,098	0,217
SDM	-771,864	39,899	-19,345	0,000	-850,076	-693,652	1,578	-0,076

Analysis of Variance: Model 1 for VAL (28 variables, n=8372)

Source	Deg. Freedom	Sum Squares	Mean Square	F-Statistic	P-value
Regression	28	25.107.073.513	896.681.197	3318.509	0,000
Residual	8343	2.254.329.185	270.206		
Total	8371	27.361.402.698			

Error Distribution Statistics: Model 1 for VAL (28 variables, n=8372)

	Mean Error	RMSE	MAE	Minimum	Maximum	MAPE	J-B stat
Fitted (n=837)	-1,073E-11	518,912	421,452	-1.996	1.920	9,3%	9.40 (P=0.009)

Figura 4.4 – Tabela de Resultados da Regressão Linear

Recorrendo à função resultante obtida – ‘Equation’, e aplicando a mesma aos 20% da Amostra de teste, obtém-se os dados que sumarizados na tabela do **Anexo III**, apresentados na coluna a cinza à qual é atribuída a variável *e*. Imediatamente a seguir são apresentados os diferenciais percentuais dos resultados da estimação por RL aos valores de oferta reais da amostra de teste – representados pela variável $g = (e/d-1) \%$.

Verifica-se na mesma tabela, que o resultado global médio para todas as freguesias, obtém um diferencial médio de -0,38%, com módulo máximo de 5,88% e módulo mínimo de 0,27%, em relação ao Valor de Oferta (VO).

A dispersão dos percentis de erro entre os valores estimados e valores de oferta, na amostra de teste, demonstrou uma distribuição conforme a Tabela 4.1 seguinte:

Regressão Linear	
%ERRO ABS	%INCID
>900%	0,000%
600-900%	0,000%
300-600%	0,000%
100-300%	0,000%
90-100%	0,000%
80-90%	0,000%
70-80%	0,000%
60-70%	0,956%
50-60%	0,430%
40-50%	0,239%
30-40%	2,150%
20-30%	5,303%
10-20%	28,906%
5-10%	28,906%
0-5%	33,923%
Média	6,721%

Tabela 4.1 – Distribuição das incidências percentuais pelas percentagens de erro

A média percentual de incidências com diferenciais entre os 0 e 100% foi de **6,721%** para este modelo.

Na Figura 4.5 apresenta-se gráfico da relação entre os valores estimados pelo modelo (RL) e os valores reais de oferta (VO).

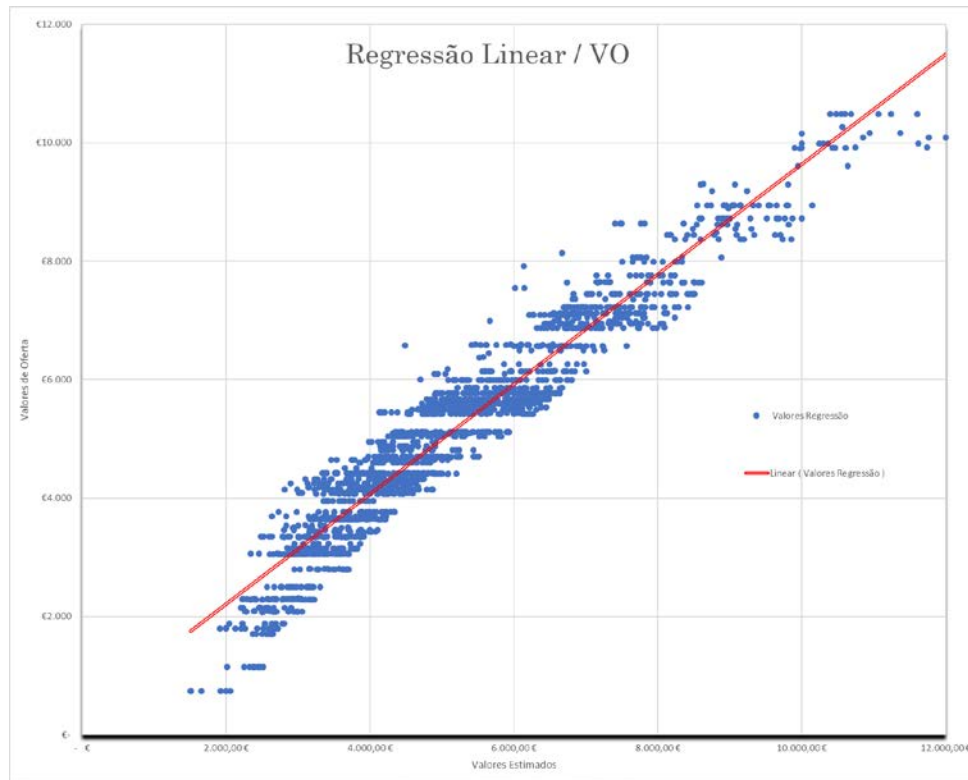


Figura 4.5 – Gráfico da relação de valores estimados da RL com valores reais de oferta

4.2 O Modelo de Rede Neuronal Artificial (RNA)

4.2.1 INTRODUÇÃO AO PROGRAMA DE ESTIMAÇÃO NEURAL DESIGNER®

A metodologia de estimação com recurso a redes neuronais artificiais, possui infindáveis modelos possíveis, constituídos pela conjugação de diversos parâmetros, já referidos anteriormente, como a sua arquitetura, a sua função de ativação, a verificação do erro e o algoritmo de otimização utilizado.

Atualmente existem programas específicos, para quem não pretende desenvolver todo o processo matemático por detrás do funcionamento de uma RNA, mas sim verificar os resultados da sua implementação recorrendo a diversos modelos.

No caso do presente estudo, foi utilizado o programa Neural Designer® (ND), cuja utilização pode ser comprovada em diversos outros trabalhos de investigação científica, nas mais variadas áreas, sejam em Engenharia, Energia, Mineralogia, Informática, Tráfego, etc., (Abrukov et al., 2020; Alnahwi & Loucks, 2019; Arhin et al., 2020; Gautam et al., 2020; Jamala & Abu-Naser, 2018; Obe & Alese, 2017).

A plataforma de aprendizagem do Neural Designer® foi criada pela Artelnics, uma empresa especializada no desenvolvimento e aplicação de técnicas de inteligência artificial para resolver problemas da atualidade.

O programa permite a definição da arquitetura de rede, escolher dentro de várias funções de ativação, definir os escalonamentos para as entradas e saídas e definir a estratégia de aprendizagem (índices de redução e algoritmos de otimização). Efetua ainda o cálculo dos valores de saída para ficheiro de base de dados, possível de importar para Excel®, e formula a expressão matemática do modelo.

4.2.2 DESENVOLVIMENTO DO MÉTODO DE RNA

O método foi implementado conforme explicado em 2.2.2, recorrendo às funções e parâmetros disponibilizados pelo programa ND.

4.2.2.1 A introdução da amostra ou ‘Data set’

A amostra utilizada contém 2 campos de **variáveis independentes alfanuméricas** cujo universo é compreendido por 24 freguesias e 5 estados de qualidade e conservação, 1 campo com uma **variável independente numérica** correspondente à área bruta do imóvel, e um campo com uma **variável dependente** correspondente ao valor de oferta de cada imóvel.

Foi seguidamente introduzida a amostra – ‘Data set’, com as 8.372 incidências de imóveis (80% da amostra total), no programa Neural Designer® em formato de tabela com 4 colunas correspondentes aos 4 campos de variáveis: Freguesia (FREG), Qualidade & Conservação (QC), Área bruta (AR) e Valor de Oferta (VAL) – Figura 4.6

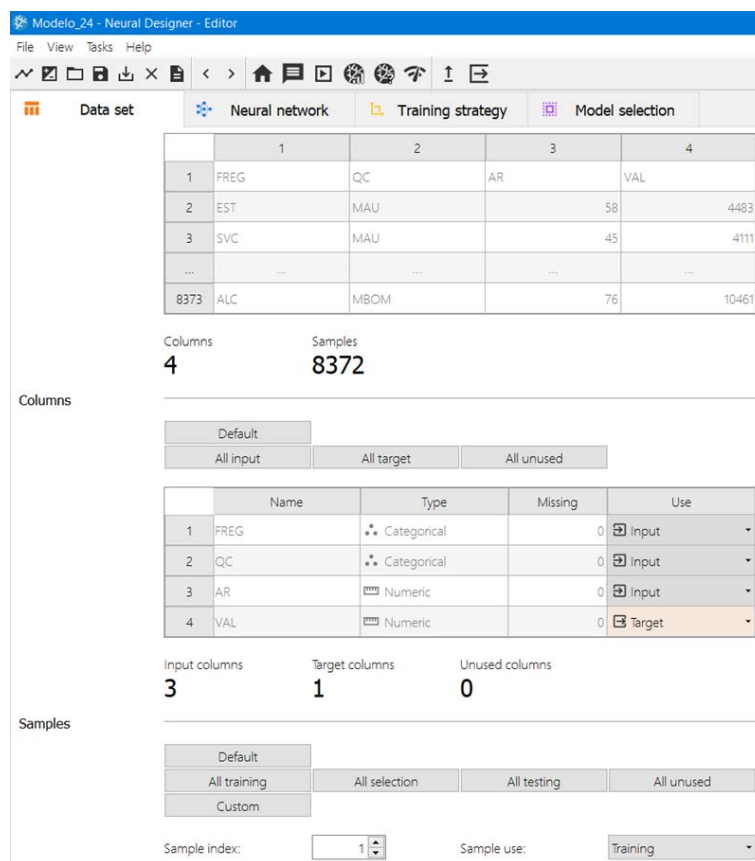


Figura 4.6 – ‘Data set’ aplicado no Neural Designer®

4.2.2.2 A Arquitetura de Rede

Para se definir a arquitetura de rede, começa-se por atribuir um número específico de camadas escondidas (as camadas *Perceptron*), e de seguida as funções de ativação que se pretendem utilizar, e por fim o escalonamento necessário aos dados de entrada e saída da rede.

Estes parâmetros constituíram a base do método, escolhidos em função do estudo efetuado nesta dissertação bem como toda a literatura que a suportou, mantidos para os vários modelos testados de RNA.

Deste modo, para a primeira camada de *Perceptron* foi utilizada uma função Tangente Hiperbólica (TH), sendo uma das mais utilizadas, também sigmoide e similar à função logística, mas com variação de -1 a 1.

Na segunda camada foi utilizada uma função linear, que não é mais do que a simples combinação da função com os dados de saída, de cada neurónio específico da camada anterior, dado que não se pretende mais do que uma redução do número de neurónios dessa camada *Perceptron* a um único neurónio com o valor estimado de cada imóvel.

O método de escalonamento à entrada pode ser automatizado pelo programa, o que no caso de existirem entradas binárias (categóricas) e numéricas, como é o caso para as FREG, QC e AR respetivamente, é ideal pois vão ser utilizados métodos diferentes. Para os dados de saída,

optou-se por um escalonamento ‘máximo-mínimo’ que é usualmente aplicado a valores de saída entre -1 e 1, como é o caso, e devolver aos mesmos a sua amplitude inicial – Equação 4.1.

$$\text{Escalação de saída} = \frac{\text{valor de entrada} - \text{média}}{\text{desvio padrão}}$$

Equação 4.1 – Escalonamento de saída por máximo-mínimo

Os dados acima descritos são parâmetros introduzidos no ND conforme Figura 4.7 seguinte.

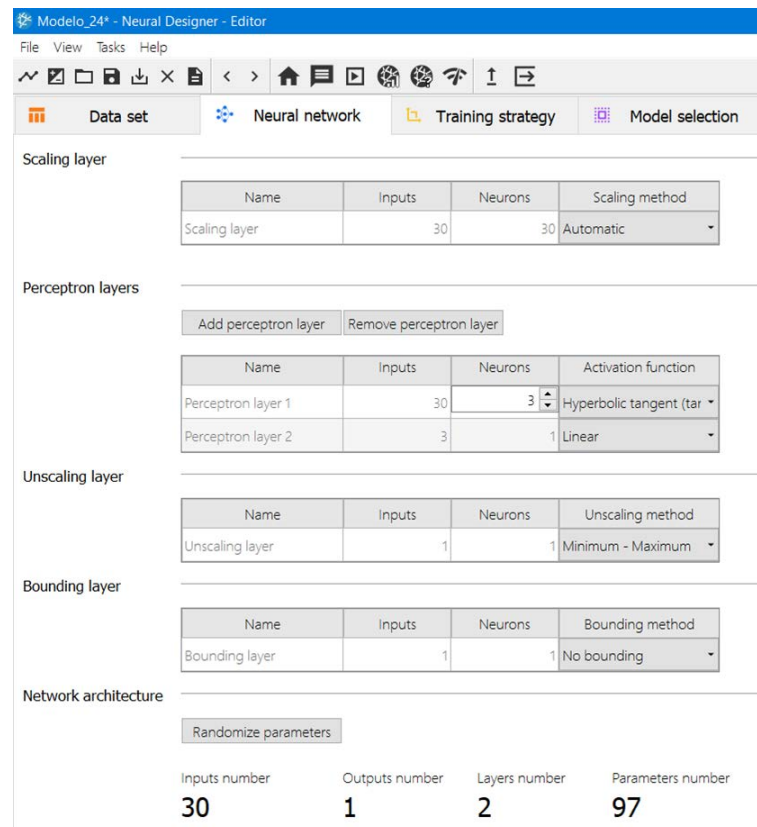


Figura 4.7 – ‘Neural Network’ aplicado no Neural Designer®

Sumarizando, a arquitetura de base, será composta por uma camada inicial para escalonamento com 30 entradas (24 freguesias + 5 estados de Q&C + Área bruta), apenas duas camadas escondidas de ‘Perceptron’, e uma camada de escalonamento de saída com um neurónio, que produzirá o resultado da estimação do valor de cada imóvel. Esta é a arquitetura usualmente aceite pela comunidade para redes neuronais simplificadas, como é o caso, e cuja representação se pode visualizar na Figura 4.8 seguinte.

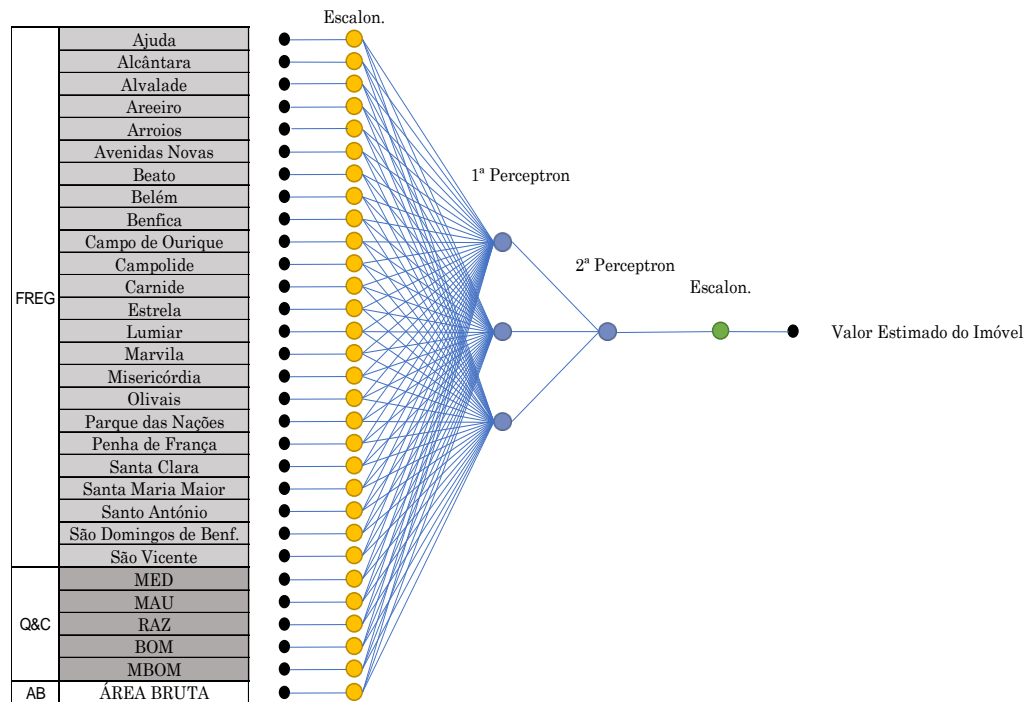


Figura 4.8 – Rede Neuronal Artificial | Fonte: o autor

4.2.2.3 A estratégia de treino ou aprendizagem da rede

Ao procedimento utilizado para executar o processo de aprendizagem chama-se estratégia de treino da RNA.

A estratégia é aplicada à rede de modo a obter os diferenciais mínimos possíveis aos dados *objetivo* (Variável dependente – VAL). Isto é feito procurando um conjunto de parâmetros, que moldam a rede neuronal (composição das suas variáveis dependentes) aos dados objetivo de cada incidência da listagem de treino (variáveis dependentes).

Esta estratégia depende de dois fatores independentes:

1. O índice de redução (IR);
2. O algoritmo de otimização.

Como foi explicado no capítulo 2.2.2, o IR tem um papel vital na execução da rede neuronal, pois define que tipo de ação a tomar e a qualidade da representação necessária, para esta aprender. A escolha do IR mais correto depende da estrutura do modelo que está a ser utilizada na Rede.

Uma solução é considerada regularizada ou atingida, quando pequenas mudanças nas variáveis de entrada levam a pequenas mudanças nas saídas. Uma abordagem para problemas não lineares, é controlar a complexidade efetiva da rede neuronal. Isso pode ser alcançado usando um termo ou coeficiente de regularização no índice de redução.

Para definir o IR, dois termos devem ser escolhidos: - o termo de erro (TE); - o termo de regularização (TR). O IR será obtido pela soma destes dois termos.

O termo de erro, é o mais importante, pois define o quão próximo a rede deve aproximar-se do valor objetivo.

Não obstante existirem 3 opções distintas no programa para os índices de redução, no que respeita ao termo de erro, a mais comum na literatura para resolver problemas de aproximação é o Erro Quadrático Normalizado (EQN) – Equação 4.2.

O erro quadrático normalizado divide o erro quadrático do diferencial entre os valores de saída da rede neuronal e os valores objetivo da amostra (preço de oferta) por um coeficiente de normalização. Se o erro quadrático normalizado tem um valor da unidade, então a rede neuronal está a prever os dados 'na média', enquanto um valor de zero significará uma previsão perfeita dos dados.

$$EQN = \frac{\sum(\text{val. saída} - \text{objetivo})^2}{\text{coeficiente de normalização}}$$

Equação 4.2 – Erro Quadrático Normalizado

Os coeficientes ou termos de regularização geralmente medem os valores dos parâmetros numa rede neuronal. Adicionar esse coeficiente ao erro fará com que a rede neuronal tenha pesos e *enviesamentos* menores, o que forçará uma resposta mais *ligeira*.

Os tipos de regularização mais usados, e disponibilizados pelo programa, são os seguintes:

- a) Regularização L1;
- b) Regularização L2.

O método de regularização L1 consiste na soma dos valores absolutos de todos os parâmetros da rede neuronal. Entenda-se por **parâmetros** todos os pesos e *enviesamentos* atribuídos aleatoriamente em cada iteração.

$$L1 = \text{peso de regularização (PR)} * \sum|\text{parâmetros}|$$

O método de regularização L2 consiste na soma quadrática de todos os parâmetros da rede neuronal.

$$L2 = \text{peso de regularização (PR)} * \sum[\text{parâmetros}]^2$$

Verifica-se então que o peso de regularização é ponderado por um parâmetro. Se a solução for muito *ligeira*, o peso deve ser diminuído. Por outro lado, se a solução oscilar muito, o peso deve ser aumentado.

Este método disponibilizado pelo Neural Designer® tem sido abordado por outros autores, desde o reconhecimento de padrões (Bishop, 1986), à previsão de chegada de autocarros em Washington (Arhin et al., 2020), entre outros ((Li et al., 2017).

Segundo o Neural Designer®, e como referido por Sattelberg (2020), o índice de redução depende da função apresentada pela rede neuronal e é medido no conjunto de dados resultantes.

Estes dados podem ser visualizados como uma híper superfície com os seus parâmetros como coordenadas, conforme na Figura 4.9 seguinte.

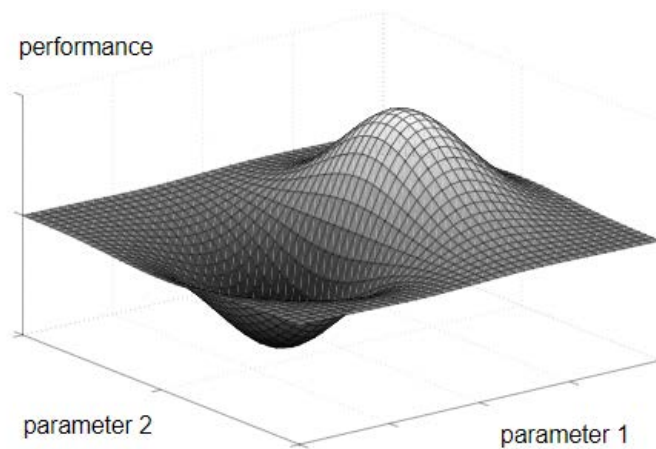


Figura 4.9 – Híper superfície definida pelos dados objetivo (apenas com 2 parâmetros) | Fonte: página “Neural Designer® training strategy tutorial”

A aprendizagem nas redes neurais, pode então ser definida, pela procura de uma função para a qual o índice de redução assume o valor mínimo. Ou seja, encontrar o conjunto de parâmetros que minimizam a função acima (L1 ou L2).

Dos testes realizados, alterando os parâmetros do programa Neural Designer® conforme se pode ver na Figura 4.10, verificou-se que os modelos quer com L1 quer com L2 apresentam resultados globais muito similares, apesar de com L1 se obter uma maior abrangência no percentil de erro (diferença da estimativa ao valor real de oferta do imóvel), e uma progressão mais linear no aumento de incidências na direção do menor percentil de erro. Com L2 reduzem as incidências de erro nos percentis de erro maiores, e aumentam as incidências nos percentis menores, de uma forma mais condensada – ver Tabela 4.2 (modelos 1 e 2).

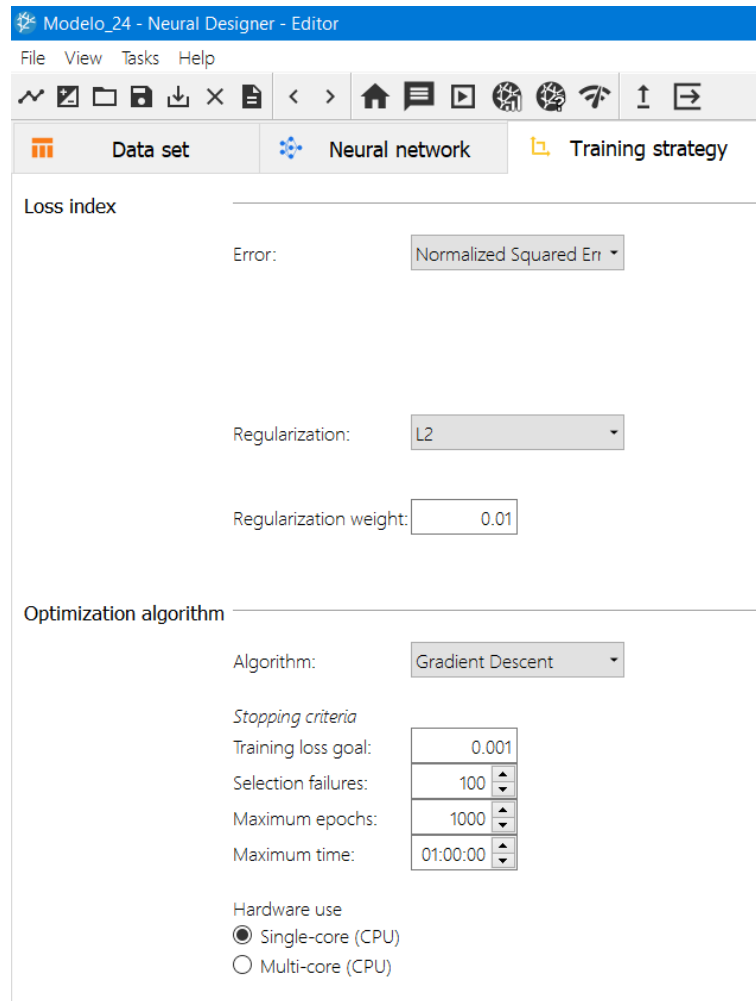


Figura 4.10 – Estratégia de treino ou aprendizagem aplicado no Neural Designer®

Relativamente à redução drástica do índice de redução (IR) – Tabela 4.2 (Modelos 3 a 5), pode constatar-se uma estabilização com tendência para 0 (zero), perdendo significância a partir de 0,000001 (limite apresentado pelo programa).

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7	Modelo 8
	L1 PR=0,01 IR=0,001	L2 PR=0,01 IR=0,001	L2 PR=0,01 IR=1,0	L2 PR=0,01 IR=0,0001	L2 PR=0,01 IR=0,000001	L2 PR=0,001 IR=0,000001	L2 PR=0,0001 IR=0,000001	L2 PR=0,1 IR=0,000001
%ERR ABS	%INCID.	%INCID.	%INCID.	%INCID.	%INCID.	%INCID.	%INCID.	%INCID.
>900%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%
600-900%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%	0,000%
300-600%	0,000%	0,000%	0,000%	0,000%	0,000%	0,048%	0,000%	0,000%
100-300%	0,000%	0,000%	2,292%	0,000%	0,000%	0,191%	0,000%	0,000%
90-100%	0,000%	0,000%	1,051%	0,000%	0,000%	0,000%	0,000%	0,000%
80-90%	0,000%	0,000%	1,910%	0,000%	0,000%	0,096%	0,000%	0,000%
70-80%	0,000%	0,000%	1,958%	0,000%	0,000%	0,000%	0,000%	0,000%
60-70%	0,000%	0,000%	2,292%	0,000%	0,000%	0,000%	0,000%	0,000%
50-60%	0,000%	0,000%	3,391%	0,000%	0,000%	0,048%	0,000%	0,000%
40-50%	0,096%	0,000%	4,585%	0,000%	0,000%	0,048%	0,048%	0,143%
30-40%	0,478%	0,573%	7,689%	0,573%	0,573%	0,430%	0,525%	0,478%
20-30%	2,006%	1,528%	13,419%	1,528%	1,480%	1,910%	1,719%	2,292%
10-20%	29,226%	26,600%	24,212%	26,313%	26,457%	25,836%	25,788%	29,704%
5-10%	32,187%	33,524%	15,664%	33,859%	33,811%	33,763%	34,288%	32,665%
0-5%	35,960%	37,727%	21,490%	37,679%	37,631%	37,584%	37,584%	34,670%
Média	6,663%	6,663%	6,663%	6,663%	6,663%	6,663%	6,663%	6,663%

Tabela 4.2 – Comparativo de resultados entre métodos de regularização L1 e L2, PR e IR variáveis

A alteração do valor do PR revelou-se instável e ineficaz para valores acima e abaixo de 0,01.

Atendendo aos ensaios dos modelos acima, considerou-se como mais eficaz o Modelo 5, por apresentar resultados mais próximos dos valores reais de oferta, com incidências crescentes e significativas (26,3%) a partir de erros absolutos menores que 20%.

4.2.2.4 O Algoritmo de otimização

Conforme foi dito, o problema da aprendizagem da rede neuronal consiste na procura de parâmetros cujo índice de redução assume um valor mínimo. Quando esta condição necessária se verifica, o gradiente é zero.

De acordo com o Neural Designer®, o IR é, de forma genérica, uma função não-linear dos parâmetros da RNA. Como consequência, não é possível encontrar algoritmos que atinjam a situação de mínimos de uma forma direta. Em vez disso, é considerada a sua procura pelos parâmetros da RNA numa sucessão de passos, ou iterações. Em cada iteração ('epoch'), o IR será reduzido ajustando os parâmetros da rede. À operação que intermeia cada iteração denomina-se *incremento de parâmetro*.

Deste modo, para treinar uma RNA, começa-se por definir um vetor aleatoriamente com esses parâmetros. É gerada uma sequência de vetores de modo que o IR reduza em cada iteração do algoritmo, conforme Figura 4.11.

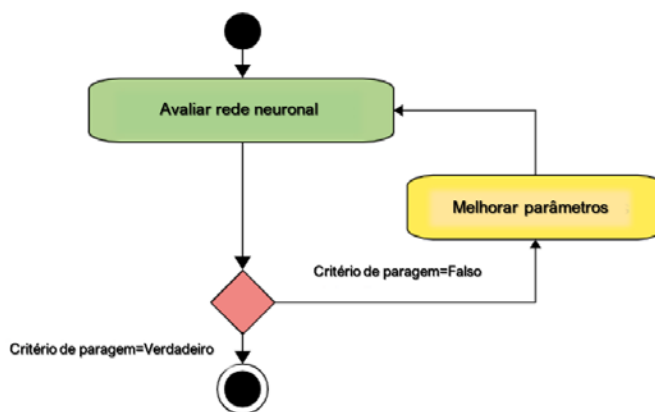


Figura 4.11 – Sequência de operações numa iteração de treino de uma RNA | Fonte: página “Neural Designer® training strategy tutorial”

Assim, o algoritmo de otimização determina como o ajustamento dos parâmetros na rede neuronal é efetuado, em cada iteração.

Conforme foi explicado no sub-capítulo 2.2.2, diferentes algoritmos são utilizados na atualidade, sendo aconselhado pelo Neural Designer®, para a dimensão da nossa amostra, o método Quasi-Newton (MQN) ou o de Gradiente Conjugado (MGC), adaptado a problemas

intermédios. No entanto, realizaram-se testes para o modelo 5 com diferentes algoritmos e foi com o **gradiente descendente** que se obtiveram melhores resultados – Figura 4.10 anterior.

4.2.3 RESULTADOS E OBSERVAÇÕES

Executado o programa Neural Designer® com a arquitetura e estratégia definidas nos subcapítulos anteriores, o mesmo fornece a função matemática que traduz os resultados do modelo de previsão depois de aplicada à amostra reservada para testagem do modelo – ver **Anexo IV**. Introduzindo a tabela de teste, o Neural Designer® devolve automaticamente os valores resultantes da aplicação dessa função.

Deste modo, obtém-se os dados que sumarizados na tabela **Anexo III**, apresentados na coluna a azul à qual é atribuída a variável f . Imediatamente a seguir são apresentados os diferenciais percentuais dos resultados da estimação por RNA aos valores de oferta reais da amostra de teste – representados pela variável $h=(e/d-1)$ %.

Verifica-se na mesma tabela, que o resultado global médio dos valores estimados de todas as freguesias, têm um diferencial médio de -0,24%, com módulo máximo de 1,68% e módulo mínimo de 0,12%, em relação ao Valor de Oferta (VO).

Como se verificou anteriormente nos gráficos da Figura 3.8 e Figura 3.9, existe uma relação entre os desvios dos resultados de avaliação da RNA aos valores de oferta, e os desvios das médias dos valores de oferta dos imóveis selecionados para teste em relação aos selecionados para treino, bem como dos desvios da distribuição de imóveis cabimentados nos percentis 80-20% também para treino-teste da rede.

Resumindo, na relação por Freguesias, quanto maior o afastamento ou desvio dos valores médios de oferta dos imóveis utilizados para testar a rede aos valores médios de oferta dos imóveis utilizados para aprendizagem da rede (linha azul), maior também o desvio da estimativa da RNA aos valores de oferta de testagem (linha cinza). O mesmo ocorre e é agravado, quanto maiores forem os desvios das médias do número de imóveis, entre as listagens para treino e teste, na repartição 80-20% (linha laranja).

Os resultados, são, por conseguinte, coerentes com os dados e distribuição da amostra aleatória introduzida no programa de cálculo, uma vez que o pressuposto de base da aprendizagem destes modelos é a quantidade e homogeneidade/coerência dos dados introduzidos.

A dispersão dos percentis de erro entre os valores estimados e valores de oferta, na amostra de teste, demonstrou uma distribuição conforme a Tabela 4.3 seguinte:

RNA	
%ERRO ABS	%INCID.
>900%	0,000%
600-900%	0,000%
300-600%	0,000%
100-300%	0,000%
90-100%	0,000%
80-90%	0,000%
70-80%	0,000%
60-70%	0,000%
50-60%	0,000%
40-50%	0,000%
30-40%	0,573%
20-30%	1,480%
10-20%	26,457%
5-10%	33,811%
0-5%	37,631%
Média	6,663%

Tabela 4.3 – Distribuição das incidências percentuais pelas percentagens de erro

A média percentual de incidências com diferenciais entre os 0 e 100% foi de **6,663%** para este modelo. Na Figura 4.12 apresenta-se gráfico da relação entre os valores estimados pelo modelo (RNA) e os valores reais de oferta (VO).

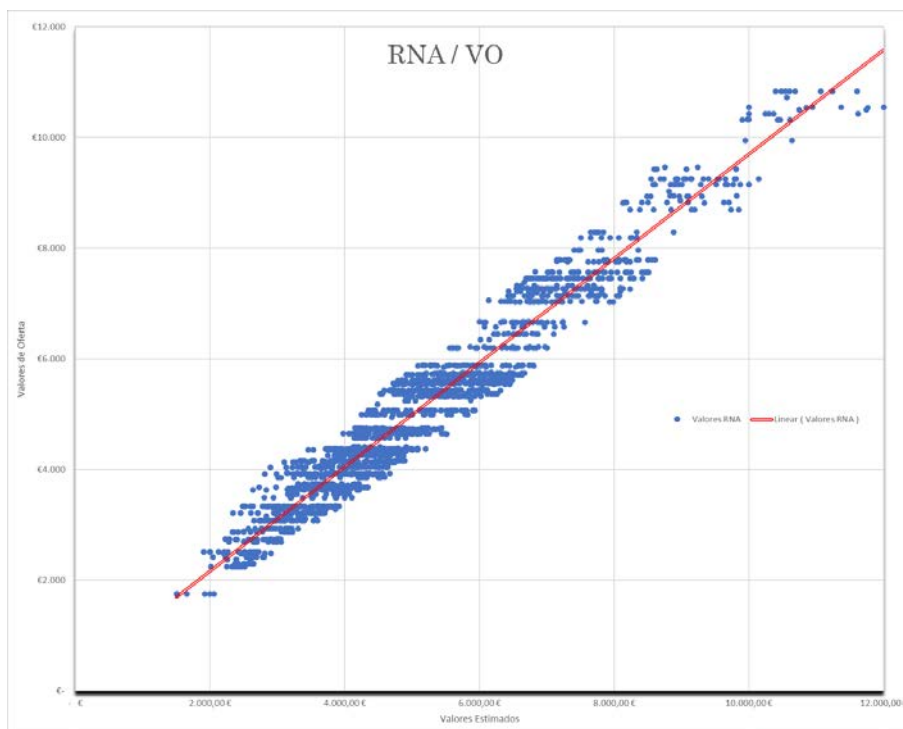


Figura 4.12 – Gráfico da relação de valores estimados com valores reais de oferta

4.3 Comparação de resultados

Relacionando os resultados obtidos e demonstrados nos subcapítulos 4.1.2 e 4.2.3, rapidamente se verificam os benefícios da avaliação imobiliária com recurso a RNA.

Com apenas 8.372 incidências foi possível obter uma otimização cujo diferencial se demonstra nos gráficos da Figura 4.13 seguinte, onde se observa uma menor amplitude na percentagem de desvio aos valores de oferta, ou seja, uma maior regularidade na proximidade de valores estimados aos valores reais de oferta:

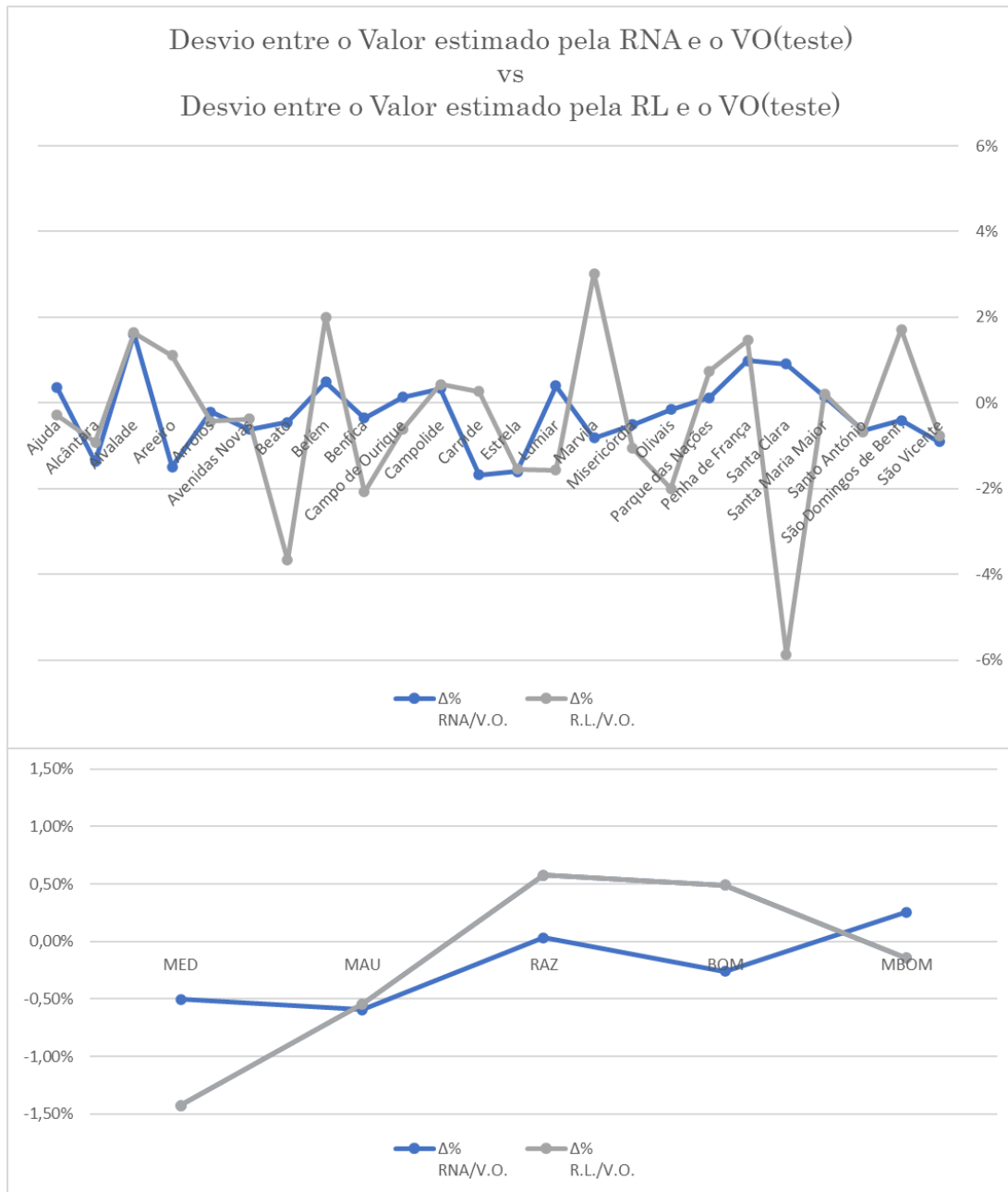


Figura 4.13 – Gráficos comparativo entre resultados da RNA e Regressão Linear, recorrendo aos desvios de distribuição por Freguesia e por QC

Comparando os dados da Tabela 4.1 e Tabela 4.3, apresentadas nos capítulos anteriores, o diferencial entre a média percentual de incidências (com diferenciais entre os 0 e 100% de erro) entre os modelos de regressão linear e rede neuronal artificial, foi de 0,058% (**6,721% - 6,663%**),

destacando a RNA em eficácia nos resultados comparativos, e que figuram seguidamente na Tabela 4.4.

Regressão Linear		RNA		Δ %
%ERRO ABS	%INCID.	%ERRO ABS	%INCID.	
>900%	0,000%	>900%	0,000%	0,000%
600-900%	0,000%	600-900%	0,000%	0,000%
300-600%	0,000%	300-600%	0,000%	0,000%
100-300%	0,000%	100-300%	0,000%	0,000%
90-100%	0,000%	90-100%	0,000%	0,000%
80-90%	0,000%	80-90%	0,000%	0,000%
70-80%	0,000%	70-80%	0,000%	0,000%
60-70%	0,956%	60-70%	0,000%	-0,956%
50-60%	0,430%	50-60%	0,000%	-0,430%
40-50%	0,239%	40-50%	0,000%	-0,239%
30-40%	2,150%	30-40%	0,573%	-1,577%
20-30%	5,303%	20-30%	1,480%	-3,823%
10-20%	28,906%	10-20%	26,457%	-2,449%
5-10%	28,906%	5-10%	33,811%	4,905%
0-5%	33,923%	0-5%	37,631%	3,709%
Média	6,721%		6,663%	-0,057%

Tabela 4.4 – Diferenciais entre % de incidências por percentis de erro absoluto

Conforme referido no capítulo 4.2.3, a RNA apresenta valores médios de erro ligeiramente superiores aos obtidos pelo MHRL, mas máximos e mínimos percentuais significativamente inferiores (diferencial de máximos -7,99%; diferencial de mínimos -48,24%, em valores absolutos), corroborando a constatação de maior regularidade e menor amplitude gráfica na dimensão do erro.

Conclui-se assim, que o método de avaliação com recurso a RNA, é eficaz, embora com um diferencial reduzido comparativamente com o modelo hedónico, o que se deve primordialmente à dimensão da amostra.

No que respeita à facilidade no uso e aplicabilidade do modelo de RNA, verifica-se ser este mais complexo que o MHRL tão somente pela necessidade de se entender a aplicabilidade dos diversos parâmetros que uma RNA abrange e que são disponibilizados pelo programa que a desenvolve, como sejam nos modelos aqui testados e apresentados. No entanto, a maior morosidade neste processo de análise e comparação foi na recolha e tratamento da amostra, pelas razões já enunciadas anteriormente.

5. CONCLUSÕES

Conforme objetivado, este trabalho desenvolveu uma análise detalhada ao método de avaliação imobiliária com recurso a rede neuronal artificial. Recorrendo a outro modelo de mercado (modelo hedónico de regressão linear), foi possível uma credibilização dos dados obtidos e por conseguinte uma comparação de resultados.

Com o trabalho desenvolvido foi também possível refletir e evidenciar nestas conclusões, quais as principais dificuldades encontradas na utilização de RNA na avaliação imobiliária, as possibilidades práticas da sua aplicação na atualidade, o que pode e deve ser melhorado, e quais as potencialidades deste método atualmente.

As RNA são já exploradas em muitas e diversas áreas científicas bem como em quase todas as plataformas eletrónicas que se utilizam no dia-a-dia. Continuarão a ser alvo de novos estudos e terão certamente muito para progredir dentro dos conceitos de IA e Aprendizagem Computacional (*‘Machine Learning’*).

As RNA, sendo modelos que ‘aprendem’ padrões numéricos por tentativa e erro, com a diversidade e quantidade de informação massiva que a *Internet* providencia, tem um vasto potencial no que respeita a inúmeras aplicações. Quanto mais vasta a informação, mais fidedignos serão os resultados de uma RNA.

5.1 A recolha e tratamento dos dados prediais

Conforme evidenciado, há atualmente uma forte acessibilidade a programas que facilitam a recolha massiva de dados na *Internet*. Ressalva-se, no entanto, a necessidade de aprender a operar com os mesmos e a necessidade de contornar alguns obstáculos, como sejam a proteção dos dados que se pretendem recolher, bem como o tratamento dos mesmos para uniformização de toda a informação recolhida.

Existe, no entanto, para um mercado como o Português, uma diversidade de informação predial que levanta a bastantes dificuldades na extração de informação que se pretende padronizável e útil como amostra, para ser aplicada num modelo de estimação. Dados como, o elevado número de imóveis em mau estado, devolutos ou por recuperar e o elevado número de

imóveis recuperados sobrevalorizados, Freguesias com uma dispersão de valores elevada por terem zonas muito diferenciadas, imóveis com e sem elevador, com e sem garagem, entre outros, dificultam a recolha de uma amostra com um número significativo de imóveis. Esta diversidade não abona naturalmente ao rastreio de uma amostra com a dimensão requerida para testar de forma fidedigna modelos de avaliação. Foi, por conseguinte, neste tema que se perdeu mais tempo de análise e preparação.

Por outro lado, existem também variáveis que ainda são pouco mensuráveis em termos de valor, como seja a certificação energética que está muito associado não só à eficiência energética, mas também ao ciclo de vida do imóvel.

Estará facilitada esta questão, se se caminhar para uma uniformização e organização dos dados públicos disponíveis, e atendendo a que cada vez mais existem programas que recolhem e corrigem a informação, crê-se que num futuro próximo esta informação será cada vez mais acessível e credível.

Conforme dito acima, é fundamental para os bons resultados dos modelos por RNA, utilizar uma amostra de dimensão proporcional à necessidade de aprendizagem, pelo que a solução para combater as dificuldades encontradas, foi de reduzir o número de campos de classificação ao essencial, escolhendo os que detinham maior capacidade de caracterização do imóvel. Desta forma, foi possível máximizara dimensão da amostra, garantindo simultaneamente a melhor caracterização do mercado ou universo de onde esta foi colhida. Recorde-se que, com a retirada de todos os subconjuntos enviesados da amostra, reduziu-se a mesma em cerca de 43%.

O programa Neural Designer[®] já investe em outras vertentes de apoio ao tratamento de dados na remoção de ‘*outliers*’, na divisão da amostra por % definidas pelo utilizador para aprendizagem e teste, bem como na comparação dos dados resultantes com uma regressão linear. Isto demonstra inequivocamente uma segurança e necessidade de afirmação, por parte do produtor deste programa, em sobrepor-se aos demais modelos de aplicação no mercado. Há, no entanto, situações que ainda não são observadas por este programa, como sejam p.ex. os desvios evidenciados no capítulo 3.3, que afetam a qualidade dos resultados. Optou-se, por conseguinte, desconsiderar estas facilidades oferecidas pelo aplicativo, e utilizar outro para testar o modelo hedónico de regressão linear, e assim se evidenciasse inequivocamente as diferenças entre o modelo de RNA e de RL.

5.2 A utilização dos programas

O recurso aos programas alvo desta dissertação, requereu o estudo com profundidade das RNA ao nível da compreensão dos seus mecanismos e seu funcionamento, bem como dos

parâmetros que permitem afinar o modelo, sem ter, no entanto, a necessidade de se aprofundar a matemática que sustenta o programa.

O estudo de vários modelos, permitiu um aprofundamento útil, sobre como as RNA podem ser otimizadas nos casos específicos de avaliação imobiliária, definindo a melhor arquitetura e estratégia de aprendizagem.

5.3 Considerações finais

Avaliações efetuadas por métodos pouco credibilizados podem ter um efeito nefasto no investimento imobiliário. De forma inversa, muito se ganhará se o exercício e validação de novos métodos como o da aplicação de RNA, for melhorado de modo a consolidar a sua eficácia.

Com este estudo, efetuado sobre uma amostra significativa, crê-se ter atingido o objetivo de demonstrar não só a eficácia do modelo de avaliação por RNA, como as vantagens e desvantagens na sua utilização, comparativamente com o método hedónico clássico. Consta-se a sua eficácia e superioridade do ponto de vista dos resultados comparativos, conforme apresentado no capítulo 4.3. Perante os resultados detalhados no capítulo 4.2.3, pode também afirmar-se que este método é utilizável de forma fidedigna e adaptável a uma qualquer avaliação imobiliária pretendida empreender com recurso a este modelo.

Foi também examinado e apresentado um resumo histórico das RNA, e uma antevisão do que poderá ser no futuro ambicionado, com a aplicação deste tipo de modelo no setor imobiliário. Como é aliás evidente, pelo desenvolvimento dado atualmente por cientistas e investigadores de todo o mundo nas mais diversas áreas de aplicação, bem como da aceitação generalizada da utilização de novas técnicas associadas a inteligência artificial e aprendizagem computacional, o que de alguma forma nos últimos anos, não tem sido tão evidente no setor imobiliário.

Desde a década de 90 que investigadores se debruçam sobre a avaliação imobiliária com recurso a RNA. Desde o estudo pioneiro de Borst (McCluskey & Borst, 1996), surgindo seguidamente muitos outros estudos apoiantes e contraditórios, essencialmente por nesta década os computadores serem ainda pouco capacitados em rapidez de processamento para suportar o funcionamento de uma RNA. Mas outros mais recentes como Chan e Abidoye, que desde 2016 se debruçaram e aprofundaram a investigação nesta área, indicando que mais de 82% dos artigos que comparavam a RNA com outros métodos apontavam maior eficácia da RNA, e cerca de 11% uma eficácia equivalente (Abidoye & Chan, 2018), inclusive corroborando as conclusões desta dissertação, testando de forma similar, na metrópole de Lagos (Chan & Abidoye, 2019).

Conclui-se, por conseguinte neste estudo, que uma rede neuronal artificial pode modelar comportamentos não lineares de variáveis de entrada, estimando preços dos imóveis para cada entrada aleatória, desde que os dados de entrada estejam na amplitude dos dados de aprendizagem

da rede. Denota-se principal robustez neste método por conseguir detetar e estabelecer uma correlação não linear entre varáveis. O método apresenta sobretudo um grau muito satisfatório de precisão, e garante a possibilidade da sua aplicação.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abidoeye, R. B., & Chan, A. P. C. C. (2018). Improving property valuation accuracy: a comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, 24(1), 71–83. <https://doi.org/10.1080/14445921.2018.1436306>
- Abrukov, V. S., Lukin, A. N., Chandrasekaran, N., Oommen, C., U.K, T., Mariappan, A., Sanal Kumar, V. R., & Anufrieva, D. A. (2020). Genome Approach and Data Science Methods for Accelerated Discovery of New Solid Propellants with Desired Properties. In *AIAA Propulsion and Energy 2020 Forum*. American Institute of Aeronautics and Astronautics. <https://doi.org/doi:10.2514/6.2020-3929>
- Alnahwi, A., & Loucks, R. G. (2019). Mineralogical composition and total organic carbon quantification using x-ray fluorescence data from the Upper Cretaceous Eagle Ford Group in southern Texas. *AAPG Bulletin*, 103(12), 2891–2907. <https://doi.org/10.1306/04151918090>
- Arhin, S., Manandhar, B., & Baba-Adam, H. (2020). Predicting travel times of bus transit in Washington, D.C. using artificial neural networks. In *Civil Engineering Journal (Iran)* (Vol. 6, Issue 11). <https://doi.org/10.28991/cej-2020-03091615>
- Avrutskiy, V. I. (2020). Enhancing Function Approximation Abilities of Neural Networks by Training Derivatives. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2), 916–924. <https://doi.org/10.1109/tnnls.2020.2979706>
- Baptistella, M., Pós-graduação, P. De, Numéricos, M., Cunico, L., & Steiner, M. (2007). *O Uso de Redes Neurais na Engenharia de Avaliações : Determinação dos Valores Venais de Imóveis Urbanos Neural Network Use in Evaluation Engineering : Figuring Out the Value of Urban Real Estate Tax*. 9, 1–229.
- Baum, A. (2017). the Future of Real Estate. *SAID Business School*, 1–95. www.sbs.oxford.edu
- Baum, Andrew, Saull, S., & Braesemann, F. (2020). *Proptech 2020: the Future of Real Estate*. 107.
- Bishop, C. M. (1986). Pattern Recognition and Machine Learning. In *Compstat*. https://doi.org/10.1007/978-3-642-46890-2_1

- Bock, S., & Weiß, M. G. (2021). *Local Convergence of Adaptive Gradient Descent Optimizers*. 1–15. <http://arxiv.org/abs/2102.09804>
- Boeing, G., & Waddell, P. (2017). New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings. *Journal of Planning Education & Research*, 37(4), 457. <https://doi.org/10.2139/ssrn.2781297>
- Braesemann, F., & Baum, A. (2020). PropTech: Turning Real Estate Into a Data-Driven Market? *SSRN Electronic Journal*, January 2020. <https://doi.org/10.2139/ssrn.3607238>
- Brien, B. O. (Deloitte U. (2020). PROPTECH: A disruptive force in Real Estate? *REflexions Magazine*, 11, 20–25.
- Buchanan, B. G. (AAAI). (2006). A (very) brief history of artificial intelligence. *AI Magazine*, 26–4(4), 53–60. <https://doi.org/10.1051/medsci/2020189>
- Camacho, J. (2020). *Inteligência artificial no imobiliário: ameaça ou oportunidade?* <https://doi.org/rcaap.com.ipl.10400.21.12466>
- Chan, A. P. C., & Abidoye, R. B. (2019). Advanced property valuation techniques and valuation accuracy: Deciphering the artificial neural network technique. *RELAND: International Journal of Real Estate & Land Planning; Vol 2 (2019), 2*, 2623–4807. <https://doi.org/10.26262/reland.v2i0.6748>
- Choromanska, A., Henaff, M., Mathieu, M., Ben Arous, G., Lecun, Y., Arous, G. Ben, Lecun, Y., Ben Arous, G., & Lecun, Y. (2014). The Loss Surface of Multilayer Networks. *Journal of Machine Learning Research*, 38(February 2015), 192–204.
- Dudek, G. (2016). Extreme learning machine as a function approximator: Initialization of input weights and biases. *Advances in Intelligent Systems and Computing*, 403(November), 59–69. https://doi.org/10.1007/978-3-319-26227-7_6
- Fausett, L. (1994). *Fundamentals of Neural Networks Architectures, Algorithms, and Applications*.
- Forbus, K. D., Kuipers, B., & Lieberman, H. (2016). Remembering Marvin Minsky. *AI Magazine*, 37(3), 94–98. <https://doi.org/10.1609/aimag.v37i3.2677>
- Gautam, A., Gopakumar, R., & Deepa, G. (2020). *Artificial Neural Network and Partial Pattern Recognition to Detect Malware BT - Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies* (V. K. Gunjan, S. Senatore, A. Kumar, X.-Z. Gao, & S. Merugu (eds.); pp. 1–8). Springer Singapore. https://doi.org/10.1007/978-981-15-3125-5_1
- Guo, C., & Berkahn, F. (2016). *Entity Embeddings of Categorical Variables. 1*, 1–9. <http://arxiv.org/abs/1604.06737>
- Hidalgo-Cajo, I. M., Yasaca-Pucuna, S., Hidalgo-Cajo, B. G., Hidalgo-Cajo, D. P., & Latorre-Benalcázar, N. B. (2020). Algoritmos Backpropagation (Bp) Y Múltiple Linear Regression

- (Mlr) a Través Del Análisis Estadístico De Datos Aplicado a Redes Neuronales Artificiales Backpropagation (Bp) and Multiple Linear Regression (Mlr) Algorithms Compar- Ative Study of the. *Revista Boletín Redipe*, 144–152. <https://doi.org/https://doi.org/10.36260/rbr.v9i3.939>
- Jafarian, A., Measoomy Nia, S., Khalili Golmankhaneh, A., & Baleanu, D. (2018). On artificial neural networks approach with new cost functions. *Applied Mathematics and Computation*, 339, 546–555. <https://doi.org/10.1016/j.amc.2018.07.053>
- Jamala, M., & Abu-Naser, S. (2018). Predicting MPG for Automobile Using Artificial Neural Network Analysis. *Information Systems Research*, 2, 5–21.
- Li, H., Wang, J., Tang, M., & Li, X. (2017). Polarization-dependent effects of an Airy beam due to the spin-orbit coupling. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 34(7), 1114–1118. <https://doi.org/10.1002/ecs2.1832>
- Liang, X. L., & Xu, J. (2021). Biased ReLU neural networks. *Neurocomputing*, 423, 71–79. <https://doi.org/10.1016/j.neucom.2020.09.050>
- LIU, X. S., DENG, Z., & WANG, T. L. (2011). Real estate appraisal system based on GIS and BP neural network. *Transactions of Nonferrous Metals Society of China*, 21(SUPPL. 3), s626–s630. [https://doi.org/https://doi.org/10.1016/S1003-6326\(12\)61652-5](https://doi.org/https://doi.org/10.1016/S1003-6326(12)61652-5)
- Martins, E. E. dos S. (Escola S. de T. e G., Jorge, M. J. da S., Ester, E., & Martins, S. (2016). *O uso de redes neuronais artificiais na estimação do preço das habitações na ilha do sal, em Cabo Verde*. <http://hdl.handle.net/10400.8/2092>
- McCluskey, W. J., & Borst, R. A. (1996). *Artificial Neural Networks* (pp. 659-679 BT- Quantitative Methods in Property). RMIT University.
- McCorduck, P. (2004). Machines Who Think. In *Jill Fineberg* (Vol. 254, Issue 5036). <https://doi.org/10.1126/science.254.5036.1291-a>
- McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- Mercioni, M. A., & Holban, S. (2020). P-Swish: Activation Function with Learnable Parameters Based on Swish Activation Function in Deep Learning. *2020 14th International Symposium on Electronics and Telecommunications, ISETC 2020 - Conference Proceedings*, 48–51. <https://doi.org/10.1109/ISETC50328.2020.9301059>
- Morris, R. G. M. M. (1999). D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949. *Brain Research Bulletin*, 50(5–6), 437. [https://doi.org/10.1016/S0361-9230\(99\)00182-3](https://doi.org/10.1016/S0361-9230(99)00182-3)
- Núñez-Tabales, J. M., Rey-Carmona, F. J., & Caridad Y Ocerin, J. M. (2016). Commercial properties prices appraisal: Alternative approach based on neural networks. *International Journal of Artificial Intelligence*, 14(1), 53–70.
- Obe, O. O., & Alese, B. K. (2017). *Remote Operating System Identification Using Artificial*

- Neural Networks*, 4(10), 8310–8313.
- Oladipupo, T. (2010). Types of Machine Learning Algorithms. *New Advances in Machine Learning, August*. <https://doi.org/10.5772/9385>
- Peter, N. J., Okagbue, H. I., Obasi, E. C. M., Akinola, A. O., Emmanuela C.M, O., & Akinola, A. O. (2020). Review on the Application of Artificial Neural Networks in Real Estate Valuation. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 2918 – 2925. <https://doi.org/10.30534/ijatcse/2020/66932020>
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55. <http://www.jstor.org/stable/1830899>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
- Sattelberg, B., Cavalieri, R., Kirby, M., Peterson, C., & Beveridge, R. (2020). Locally Linear Attributes of ReLU Neural Networks. *ArXiv*, 1–18. <http://arxiv.org/abs/2012.01940>
- Shi, H. (2009). Determination of Real Estate Price Based on Principal Component Analysis and Artificial Neural Networks. *2009 Second International Conference on Intelligent Computation Technology and Automation, Intelligent Computation Technology and Automation, 2009. ICICTA '09. Second International Conference On, 1*, 314–317. <https://doi.org/10.1109/ICICTA.2009.83>
- Shybaiev, S., Otradskaia, V., Shybaieva, O., Rudnichenko, D., Stepanchuk, V., Shybaiev, D. S., Otradskaia, T. V., Stepanchuk, M. V., Shybaieva, N. O., Rudnichenko, N. D., D.S., S., T.V., O., N.O., S., N.D., R., & M.V., S. (2019). Predicting system for the estimated cost of real estate objects development using neural networks Automation. *Odessa College of Computer Technologies*, 1(83), 154–160. [https://doi.org/10.26642/tn-2019-1\(83\)-154-160](https://doi.org/10.26642/tn-2019-1(83)-154-160)
- Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 13(1), 3–43. <http://www.jstor.org/stable/44103506>
- Sütfeld, L. R., Brieger, F., Finger, H., Füllhase, S., & Pipa, G. (2020). Adaptive Blending Units: Trainable Activation Functions for Deep Neural Networks. *Advances in Intelligent Systems and Computing*, 1230 AISC(February 2019), 37–50. https://doi.org/10.1007/978-3-030-52243-8_4
- Teixeira, M. C. C., Caridad y Ocerin, J. M., & Villamandos, N. C. (2010). Redes Neuronais Artificiais para estimar o preço da habitação em Portugal. *International Meeting on Regional Science - The Future of the Cohesion Policy, November*, 1–20. <http://repositorio.ipcb.pt/handle/10400.11/1127>

- Uzun, E. (2020). A Novel Web Scraping Approach Using the Additional Information Obtained From Web Pages. *IEEE Access, Access, IEEE*, 8, 61726–61740. <https://doi.org/10.1109/ACCESS.2020.2984503>
- Vasques, F. (Instituto S. de E. de L. (ISEL)). (2017). A inteligência artificial no imobiliário - Evolução e disrupção. *Ci: Research - Artigos Científicos, Outubro*, 27–29.
- Winston, P. (2015). Artificial Intelligence - Massachusetts Institute of Technology: MIT OpenCourseWare. In *Creative Commons BY-NC-SA*.
- Zhao, H. Z., Liu, F. X., Li, L. Y., & Luo, C. (2017). Improving deep convolutional neural networks with mixed maxout units. *Tongxin Xuebao/Journal on Communications*, 38(7), 105–114. <https://doi.org/10.11959/j.issn.1000-436x.2017145>
- Zhou, G., Ji, Y., Chen, X., & Zhang, F. (2018). Artificial Neural Networks and the Mass Appraisal of Real Estate. *International Journal of Online Engineering (IJOE)*, 14(03), 180–187. <https://doi.org/10.3991/ijoe.v14i03.8420>

WEBGRAFIA

Bibliografia oficial (Janeiro a Agosto de 2021):

<https://www.b-on.pt/>

<https://www.sciencedirect.com/>

<https://www.researchgate.net/>

<https://www.mendeley.com/>

<https://deepai.org/>

Recolha de dados para amostragem (Abril de 2021):

<https://www.idealista.pt/>

<https://www.imovirtual.pt/>

<https://www.sapo.pt/>

<https://www.remax.pt/>

<https://www.century21.pt/>

<https://www.era.pt/>

Programa de Web Scraping e modelação (Janeiro 2021):

<https://www.webharvy.com/>

<https://www.octoparse.com/>

<https://www.mathworks.com/products/matlab.html>

<https://www.neuraldesigner.com/>

Páginas de apoio informativo:

Universidades

(Jan 2021) <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

(Jan 2021) <http://stanford.edu/~cpiech/cs221/apps/deepBlue.html>

(Fev 2021) <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/#>

Forums, Revistas, Influencers

(Mar 2021) <https://www.forbes.com/sites/bernardmarr/2020/02/03/the-top-propotech-trends-6-technologies-disrupting-the-property-and-real-estate-industry/?sh=36bba123dc16>

(Mar 2021) <https://www.forbes.com/sites/angelicakrystledonati/2020/03/28/this-is-what-real-estate-and-propotech-should-be-doing-to-protect-your-data/?sh=7ad431cb6797>

(Jan 2021) <https://towardsdatascience.com/a-beginners-guide-to-neural-nets-5cf4050117cb>

(Jan 2021) <https://towardsdatascience.com/why-data-should-be-normalized-before-training-a-neural-network-c626b7f66c7d>

(Fev 2021) <https://towardsdatascience.com/neural-network-optimization-algorithms-1a44c282f61d>

(Mar 2021) <http://www.cse.unsw.edu.au/~billw/aidict.html>

(Fev 2021) <https://medium.com/deeper-learning/glossary-of-deep-learning-bias-cf49d9c895e2>

(Fev 2021) <https://medium.com/nerd-for-tech/types-of-optimization-algorithms-used-in-neural-networks-and-ways-to-optimize-gradient-descent-1e32cdbc6f6c>

(Fev 2021) <https://medium.com/@jaroncollis/how-i-learned-to-stop-worrying-and-love-the-error-6c4c488c97da>

(Mar 2021) <https://tjmachinelearning.com/lectures/1819/nn2/>

(Mar 2021) <https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/>

(Fev 2021) <https://ai-summary.com/>

(Jan 2021) <https://towardsdatascience.com/6-web-scraping-tools-that-make-collecting-data-a-breeze-457c44e4411d>

(Jan 2021) <https://towardsdatascience.com/optimizers-for-training-neural-network>

(Mar 2021) <https://developers.google.com/machine-learning/crash-course?hl=pt>

(Fev 2021) <https://bdaily.co.uk/articles/2019/08/02/google-backed-ai-focused-propotech-startup-secures-15m-seed-funding>

ANEXO I – TABELA DE VALORES MÉDIOS DE FRAÇÕES POR ESTADOS DE FREGUESIAS PARA DETECÇÃO DE REPRESENTATIVIDADE DE VALORES POR FREGUESIA

Fração a avaliar	V.Médios	Áreas																	
		<50m2		50<=x<75		75<=x<100		100<=x<150		150<=x<200		200<=x<300		300<=x<400		x=>400			
		n.Imóv	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	Qnt	Méd(€/m2)	
	5.328 €	12.354	1.079	5.901 €	2.581	5.222	2.405	5.120	3.213	5.229	1.642	5.449	1.118	5.477	220	5.488	96	6.094	
Estrela	Novo	6.779 €	120	7	6.875 €	31	6.772 €	31	6.413 €	38	6.997 €	9	6.899 €	4	7.150 €	0		0	
	Usado	5.339 €	779	96	5.650 €	157	5.077 €	128	5.391 €	194	5.553 €	103	4.941 €	77	5.467 €	17	5.541 €	7	4.011 €
Lumiar	Novo	2.908 €	18	0		1	3.500 €	5	2.762 €	5	2.930 €	4	2.857 €	3	2.987 €	0		0	
	Usado	3.687 €	419	1	3.222 €	33	3.950 €	65	3.376 €	147	3.682 €	108	3.840 €	53	3.626 €	7	3.973 €	5	3.129 €
Marvila	Novo	5.472 €	103	10	6.930 €	19	6.314 €	15	6.002 €	30	4.773 €	19	4.872 €	8	4.881 €	2	4.725 €	0	
	Usado	3.183 €	163	8	4.152 €	43	2.743 €	44	2.949 €	40	2.843 €	14	3.949 €	13	4.839 €	1	6.037 €	0	
Misericórdia	Novo	7.808 €	131	16	7.111 €	31	7.518 €	14	6.929 €	21	8.245 €	25	7.789 €	14	7.761 €	2	6.833 €	8	11.102 €
	Usado	6.157 €	897	131	6.341 €	248	6.335 €	179	5.743 €	166	5.822 €	87	6.212 €	64	6.582 €	15	7.156 €	7	8.178 €
Olivais	Novo	3.132 €	29	0		0		0		1	3.687 €	2	3.322 €	22	3.115 €	4	2.992 €	0	
	Usado	3.179 €	243	7	3.203 €	90	3.187 €	67	2.995 €	41	3.381 €	13	2.975 €	22	3.467 €	3	2.970 €	0	
Parque das Nações	Novo	7.502 €	86	8	8.011 €	7	7.776 €	16	7.053 €	27	6.786 €	23	7.676 €	4	10.294 €	0		1	12.895 €
	Usado	5.689 €	440	16	7.337 €	47	5.965 €	61	5.482 €	138	5.808 €	104	5.149 €	54	5.668 €	11	5.695 €	9	7.239 €
Penha de França	Novo	5.174 €	11	1	4.565 €	2	5.444 €	4	5.205 €	3	4.904 €	1	5.928 €	0		0		0	
	Usado	3.784 €	538	37	4.504 €	157	3.905 €	165	3.683 €	128	3.621 €	36	3.512 €	11	3.493 €	4	5.051 €	0	
Santa Clara	Novo		0	0		0		0		0		0		0		0		0	
	Usado	2.640 €	95	0		14	2.748 €	23	2.419 €	46	2.619 €	6	3.000 €	6	3.043 €	0		0	
Santa Maria Maior	Novo	8.050 €	59	9	8.172 €	4	7.143 €	18	8.309 €	23	7.780 €	3	9.030 €	2	8.616 €	0		0	
	Usado	5.858 €	897	203	5.793 €	205	5.308 €	142	5.936 €	218	6.115 €	78	5.898 €	39	6.979 €	9	6.937 €	3	6.728 €

ANEXO II – TABELA DE DADOS ESTATÍSTICOS DA AMOSTRA GLOBAL

Fração a avaliar	Qualid.&Conserv.	Qnt.	Min	Máx	V.Médios 5.062 €	N. Imóveis	Dif. Máx/Mn.	Áreas								Áreas										
								n.Imóv	10m2	15m2	20m2	25m2	30m2	35m2	40m2	45m2	50m2	10m2	15m2	20m2	25m2	30m2	35m2	40m2	45m2	50m2
Ajuda	MED	44	2.037 €	3.058 €	2.548 €	217	5.106 €	10.463	780	2.321	2.101	2.699	1.399	985	199	1	5.668 €	4.984 €	4.845 €	4.961 €	5.157 €	5.308 €	5.431 €	4.500 €		
	MAU	82	3.058 €	4.079 €	3.569 €			44	1	19	15	8	0	1	0	0	0	0	3.000 €	2.644 €	2.777 €	2.883 €	- €	2.500 €	- €	- €
	RAZ	76	4.079 €	5.101 €	4.590 €			82	5	40	19	14	1	2	1	0	0	0	3.731 €	3.598 €	3.546 €	3.461 €	3.693 €	3.501 €	3.906 €	- €
	BOM	11	5.101 €	6.122 €	5.611 €			76	16	31	11	8	8	2	0	0	0	0	4.696 €	4.505 €	4.621 €	4.519 €	4.537 €	4.233 €	- €	- €
	MBOM	4	6.122 €	7.143 €	6.632 €			11	8	2	0	2	0	1	0	0	0	0	5.324 €	5.754 €	- €	5.346 €	- €	5.556 €	- €	- €
Alcântara	MED	107	2.800 €	4.609 €	3.705 €	235	9.046 €	107	9	41	30	20	4	2	1	0	4.279 €	3.922 €	3.578 €	3.866 €	4.231 €	4.483 €	4.592 €	- €		
	MAU	79	4.609 €	6.418 €	5.514 €			107	11	12	17	28	8	8	1	0	0	5.346 €	5.389 €	5.313 €	5.390 €	5.368 €	5.257 €	5.115 €	- €	
	RAZ	23	6.418 €	8.228 €	7.323 €			23	2	5	7	7	0	2	0	0	0	7.121 €	6.935 €	6.806 €	7.720 €	- €	7.465 €	- €	- €	
	BOM	15	8.228 €	10.037 €	9.132 €			15	0	0	10	2	0	3	0	0	0	- €	- €	- €	9.406 €	8.432 €	- €	9.501 €	- €	- €
	MBOM	11	10.037 €	11.846 €	10.942 €			11	0	3	7	0	0	1	0	0	0	- €	- €	- €	11.442 €	10.580 €	- €	10.955 €	- €	- €
Alalade	MED	107	2.640 €	4.242 €	3.441 €	339	8.012 €	107	0	8	18	48	21	13	3	0	- €	3.763 €	3.582 €	3.677 €	3.694 €	3.641 €	3.579 €	- €		
	MAU	140	4.242 €	5.845 €	5.044 €			140	1	25	32	54	20	5	3	0	0	5.475 €	5.023 €	4.959 €	4.804 €	4.916 €	4.665 €	5.352 €	- €	
	RAZ	68	5.845 €	7.447 €	6.646 €			68	1	10	4	30	9	10	4	0	0	6.729 €	6.790 €	6.831 €	6.751 €	6.289 €	7.018 €	6.679 €	- €	
	BOM	22	7.447 €	9.050 €	8.249 €			22	0	8	6	3	3	2	0	0	0	- €	- €	- €	7.940 €	8.011 €	7.596 €	7.857 €	7.985 €	- €
	MBOM	2	9.050 €	10.652 €	9.861 €			2	0	1	1	0	0	0	0	0	0	- €	- €	- €	10.652 €	10.000 €	- €	- €	- €	- €
Areeiro	MED	86	2.555 €	3.715 €	3.135 €	312	5.801 €	86	0	2	10	23	17	14	0	0	- €	3.544 €	3.526 €	3.286 €	3.207 €	3.320 €	- €	- €		
	MAU	104	3.715 €	4.876 €	4.295 €			104	0	13	21	23	28	17	2	0	0	- €	4.032 €	4.255 €	4.315 €	4.392 €	4.165 €	4.355 €	- €	
	RAZ	87	4.876 €	6.036 €	5.456 €			87	0	11	19	34	14	9	0	0	0	- €	5.446 €	5.470 €	5.466 €	5.291 €	5.315 €	- €	- €	
	BOM	43	6.036 €	7.196 €	6.616 €			43	0	3	5	23	8	8	0	0	0	- €	6.343 €	6.477 €	6.459 €	6.746 €	6.336 €	- €	- €	
	MBOM	12	7.196 €	8.356 €	7.776 €			12	0	2	1	3	5	0	1	0	0	- €	- €	- €	7.874 €	7.813 €	7.505 €	7.442 €	- €	7.759 €
Arroios	MED	351	2.338 €	3.939 €	3.138 €	1016	8.007 €	351	11	55	78	120	58	28	1	0	3.552 €	3.488 €	3.346 €	3.298 €	3.262 €	3.413 €	3.600 €	- €		
	MAU	393	3.939 €	5.541 €	4.740 €			393	38	97	87	109	52	28	2	0	0	4.874 €	4.536 €	4.556 €	4.602 €	4.760 €	4.632 €	4.545 €	- €	
	RAZ	192	5.541 €	7.142 €	6.341 €			192	28	38	40	43	32	12	2	0	0	6.241 €	6.143 €	6.328 €	6.137 €	6.235 €	6.111 €	6.708 €	- €	
	BOM	72	7.142 €	8.743 €	7.943 €			72	12	23	17	4	8	5	3	0	0	7.751 €	7.913 €	7.544 €	7.592 €	8.184 €	8.180 €	8.073 €	- €	
	MBOM	8	8.743 €	10.345 €	9.544 €			8	1	4	1	2	0	0	0	0	0	9.000 €	9.271 €	9.048 €	10.094 €	- €	- €	- €		
Avenidas Novas	MED	213	3.053 €	4.775 €	3.914 €	817	8.606 €	213	3	23	25	82	42	43	15	0	4.426 €	4.040 €	3.963 €	4.009 €	4.201 €	4.057 €	4.379 €	- €		
	MAU	330	4.775 €	6.496 €	5.635 €			330	7	41	40	101	56	72	13	0	0	5.701 €	5.863 €	5.734 €	5.682 €	5.602 €	5.696 €	5.336 €	- €	
	RAZ	209	6.496 €	8.217 €	7.356 €			209	15	20	28	78	39	30	3	0	0	7.523 €	6.936 €	7.231 €	7.372 €	7.249 €	7.102 €	7.083 €	- €	
	BOM	52	8.217 €	9.938 €	9.077 €			52	9	12	6	9	10	5	1	0	0	8.865 €	9.011 €	8.707 €	8.847 €	8.817 €	9.113 €	9.207 €	- €	
	MBOM	13	9.938 €	11.659 €	10.799 €			13	2	3	0	0	3	5	0	0	0	11.151 €	10.059 €	- €	- €	10.126 €	10.537 €	- €	- €	
Beato	MED	21	1.777 €	2.841 €	2.309 €	114	5.324 €	21	2	9	6	3	1	0	0	0	2.433 €	2.579 €	2.465 €	2.336 €	2.579 €	- €	- €			
	MAU	66	2.841 €	3.906 €	3.374 €			66	9	34	9	12	2	0	0	0	3.450 €	3.269 €	3.177 €	3.196 €	3.287 €	- €	- €			
	RAZ	21	3.906 €	4.971 €	4.439 €			21	9	10	1	1	0	0	0	0	4.336 €	4.294 €	4.667 €	3.913 €	- €	- €	- €			
	BOM	4	4.971 €	6.036 €	5.503 €			4	0	3	0	0	1	0	0	0	- €	- €	- €	5.128 €	- €	- €	5.568 €	- €	- €	
	MBOM	2	6.036 €	7.101 €	6.568 €			2	0	0	0	1	1	0	0	0	- €	- €	- €	7.101 €	- €	- €	7.101 €	- €	- €	
Belém	MED	154	2.931 €	4.714 €	3.823 €	374	8.915 €	154	3	18	25	35	25	45	14	0	4.245 €	3.911 €	3.930 €	3.772 €	4.133 €	3.949 €	3.988 €	- €		
	MAU	128	4.714 €	6.497 €	5.606 €			128	5	7	15	9	31	42	19	0	0	5.482 €	5.038 €	5.651 €	5.270 €	5.287 €	5.564 €	5.625 €	- €	
	RAZ	45	6.497 €	8.280 €	7.389 €			45	0	2	0	11	11	20	1	0	0	- €	6.950 €	- €	7.301 €	7.281 €	7.166 €	6.571 €	- €	
	BOM	27	8.280 €	10.063 €	9.172 €			27	0	0	18	2	1	6	0	0	0	- €	- €	- €	9.534 €	8.479 €	9.000 €	9.349 €	- €	
	MBOM	10	10.063 €	11.846 €	10.955 €			10	0	2	8	1	0	1	0	0	0	- €	- €	- €	11.618 €	10.762 €	10.204 €	- €	10.955 €	
Benfica	MED	29	1.724 €	2.522 €	2.123 €	321	3.990 €	29	0	7	8	12	1	1	0	0	- €	2.306 €	2.232 €	2.335 €	2.314 €	2.156 €	- €	- €		
	MAU	130	2.522 €	3.320 €	2.921 €			130	1	31	48	34	9	7	0	0	0	3.063 €	2.978 €	3.032 €	2.953 €	2.997 €	2.856 €	- €		
	RAZ	111	3.320 €	4.118 €	3.719 €			111	6	35	37	24	8	1	0	0	0	3.676 €	3.683 €	3.624 €	3.703 €	3.562 €	4.079 €	- €		
	BOM	39	4.118 €	4.916 €	4.517 €			39	6	16	7	8	2	0	0	0	0	4.406 €	4.406 €	4.489 €	4.345 €	4.403 €	- €	- €		
	MBOM	12	4.916 €	5.714 €	5.315 €			12	2	5	1	0	3	1	0	0	0	5.702 €	5.301 €	5.385 €	- €	5.130 €	5.233 €	- €		
Campo de Ourique	MED	99	2.717 €	4.503 €	3.610 €	450	8.926 €	99	4	38	21	23	8	5	0	0	3.948 €	3.890 €	3.836 €	3.979 €	3.945 €	3.628 €	- €			
	MAU	218	4.503 €	6.288 €	5.395 €			218	17	50	43	66	29	9	4	0	0	5.608 €	5.330 €	5.326 €	5.507 €	5.259 €	5.363 €	5.537 €	- €	
	RAZ	97	6.288 €	8.073 €	7.181 €			97	7	14	21	30	13	12	0	0	0	6.793 €	6.730 €	7.015 €	7.053 €	6.955 €	7.034 €	- €		
	BOM	21	8.073 €	9.859 €	8.966 €			21	3	1	6	1	7	3	0	0	0	9.028 €	9.306 €	8.469 €	8.759 €	8.974 €	8.846 €	- €		
	MBOM	15	9.859 €	11.644 €	10.751 €			15	2	4	0	1	3	3	2	0	0	10.535 €	10.601 €	- €	- €	11.288 €	10.650 €	10.562 €	9.987 €	
Campolide	MED	95	2.672 €	4.137 €	3.404 €	250	7.326 €	95	4	26	28	19	8	9	1	0	3.846 €	3.364 €	3.383 €	3.613 €	3.626 €	3.552 €	3.684 €	- €		
	MAU	83	4.137 €	5.602 €	4.889 €			83	5	25	15	17	12	5	4	0	0	4.337 €	4.793 €	4.803 €	4.817 €	4.894 €	4.585 €	4.338 €	- €	
	RAZ	33	5.602 €	7.067 €	6.325 €			33	2	1	8	15	1	5	1	0	0	5.929 €	5.982 €	6.382 €	6.371 €	5.730 €	6.275 €	6.824 €	- €	
	BOM	31																								

ANEXO III – TABELA DE ANÁLISE DE RESULTADOS E COMPARATIVA DE DADOS RESULTANTES DAS DIFERENTES TÉCNICAS DE ESTIMAÇÃO POR RNA E MHRL

Freguesia	Total Imóveis n	Δ (V.Máx./V.Min.) z	$=1/n \cdot z$	n.ºimóveis (treino) a	V.O.Médios (treino) b	n.ºimóveis (teste) c	V.O.Médios (teste) d	V.O. % teste/treino $ d/b-1 $ %	n.ºimóveis $\Delta\%$ (80-20%) $ a/(a+c)-0,8 $ %	RESULTADOS R.L. Val.Méd. (amostra de teste - 20%) e	$\Delta\%$ R.L./V.O. $g=(e/d-1)$ %	RESULTADOS R.N.A. Val.Méd. (amostra de teste - 20%) f	$\Delta\%$ R.N.A./V.O. $h=(f/d-1)$ %	RNA/R.L. h/g
Ajuda	217	3.51	0.02	165	3.910 €	52	3.858 €	1.347%	3.963%	3.847 €	-0.288%	3.872 €	0.361%	0.651%
Alcântara	235	4.23	0.02	196	5.386 €	39	5.281 €	1.937%	3.404%	5.233 €	-0.917%	5.209 €	-1.367%	0.454%
Alvalade	339	4.03	0.01	274	5.114 €	65	5.084 €	0.584%	0.826%	5.167 €	1.643%	5.165 €	1.602%	0.040%
Areiro	312	3.27	0.01	209	4.673 €	103	5.132 €	9.820%	13.013%	5.189 €	1.111%	5.055 €	-1.494%	-2.577%
Arroios	1016	4.43	0.00	847	4.709 €	169	4.962 €	5.371%	3.366%	4.940 €	-0.427%	4.951 €	-0.202%	0.226%
Avenidas Novas	817	3.82	0.00	666	5.961 €	151	5.912 €	0.813%	1.518%	5.890 €	-0.372%	5.876 €	-0.613%	0.242%
Beato	114	4.00	0.04	92	3.482 €	22	3.316 €	4.787%	0.702%	3.194 €	-3.669%	3.301 €	-0.447%	3.345%
Belém	374	4.04	0.01	312	5.526 €	62	4.973 €	10.006%	3.422%	5.073 €	2.004%	4.998 €	0.490%	-1.484%
Benfica	321	3.31	0.01	261	3.430 €	60	3.366 €	1.876%	1.308%	3.296 €	-2.680%	3.354 €	-0.348%	1.769%
Campo de Ourique	450	4.28	0.01	334	5.610 €	116	6.087 €	8.504%	5.778%	6.049 €	-0.616%	6.095 €	0.132%	0.753%
Campolide	250	3.74	0.01	192	4.826 €	58	5.529 €	14.567%	3.200%	5.553 €	0.432%	5.548 €	0.334%	0.098%
Carnide	77	3.27	0.04	64	3.654 €	13	3.917 €	7.175%	3.117%	3.927 €	0.266%	3.851 €	-1.676%	-1.937%
Estrela	736	3.96	0.01	605	5.306 €	131	5.416 €	2.075%	2.201%	5.333 €	-1.543%	5.329 €	-1.601%	0.058%
Lumiar	414	3.84	0.01	331	3.723 €	83	3.574 €	3.990%	0.048%	3.519 €	-1.566%	3.589 €	0.397%	1.994%
Marvila	162	3.66	0.02	129	3.112 €	33	3.398 €	9.195%	0.370%	3.500 €	3.022%	3.370 €	-0.819%	-3.728%
Misericórdia	867	3.84	0.00	683	6.045 €	184	6.291 €	4.054%	1.223%	6.224 €	-1.055%	6.259 €	-0.505%	0.556%
Olivais	242	3.92	0.02	195	3.187 €	47	3.120 €	2.086%	0.579%	3.058 €	-1.994%	3.115 €	-0.155%	1.876%
Parque das Nações	430	4.14	0.01	334	5.691 €	96	5.466 €	3.960%	2.326%	5.506 €	0.741%	5.472 €	0.120%	0.616%
Penha de França	532	3.85	0.01	425	3.766 €	107	3.783 €	0.457%	0.113%	3.839 €	1.466%	3.820 €	0.986%	0.473%
Santa Clara	95	3.54	0.04	77	2.649 €	18	2.605 €	1.654%	1.053%	2.452 €	-5.883%	2.629 €	0.910%	7.217%
Santa Maria Maior	840	4.14	0.00	704	5.821 €	136	5.783 €	0.662%	3.810%	5.795 €	0.216%	5.791 €	0.151%	0.065%
Santo António	684	3.57	0.01	547	6.746 €	137	6.851 €	1.562%	0.029%	6.805 €	-0.673%	6.807 €	-0.651%	0.022%
São Domingos de Benf	355	3.47	0.01	277	4.035 €	78	4.184 €	3.669%	1.972%	4.255 €	1.720%	4.166 €	-0.498%	-2.092%
São Vicente	586	4.33	0.01	453	4.515 €	133	4.550 €	0.767%	2.696%	4.515 €	-0.761%	4.509 €	-0.996%	0.146%
Total /Média				8372		2093		4.205%	2.502%		-0.384%		-0.238%	0.183%
Qualidade e Conservação														
MED	2915	3.72	0.001	2377	3.577 €	538	3.553 €	0.650%	1.544%	3.503 €	-1.427%	3.536 €	-0.502%	0.939%
MAU	4235	3.22	0.001	3403	4.702 €	832	4.760 €	1.248%	0.354%	4.734 €	-0.545%	4.732 €	-0.592%	-0.047%
RAZ	2227	3.09	0.001	1722	6.178 €	505	6.081 €	1.579%	2.676%	6.116 €	0.578%	6.083 €	0.034%	-0.541%
BOM	827	2.91	0.004	660	7.814 €	167	7.708 €	1.346%	0.193%	7.746 €	0.492%	7.689 €	-0.258%	-0.746%
MBOM	261	2.68	0.010	210	9.352 €	51	9.464 €	1.192%	0.460%	9.450 €	-0.146%	9.488 €	0.254%	0.400%
Total /Média				8372		2093		1.203%	1.045%		-0.210%		-0.213%	0.001%

**ANEXO IV - RELATÓRIO DO MODELO 5 DO
NEURAL DESIGNER®(29 PÁGINAS)**

Modelo_5

1 Neural network

1.1 Task description

The neural network represents the predictive model. In Neural Designer neural networks allow deep architectures, which are a class of universal approximator.

1.2 Inputs

The number of inputs is 30. The next table depicts some basic information about them, including the name, the units and the description.

	Name
1	EST
2	SVC
3	CMP
4	PEF
5	OLI
6	ALC
7	CAM
8	SAM
9	MIS
10	MAR
11	SAC
12	ARR
13	BEM
14	SAN
15	SDM
16	PAN
17	AJU
18	AVE
19	ALV
20	LUM
21	BEA
22	BEL

1

23	ARE
24	CRN
25	MAU
26	RAZ
27	MED
28	BOM
29	MBOM
30	AR

1.3 Scaling layer

The size of the scaling layer is 30, the number of inputs. The scaling method for this layer is the Minimum - Maximum. The following table shows the values which are used for scaling the inputs, which include the minimum, maximum, mean and standard deviation.

	Minimum	Maximum	Mean	Deviation
EST	0	1	0.0722647	0.258923
SVC	0	1	0.0541089	0.226236
CMP	0	1	0.0229336	0.149685
PEF	0	1	0.0507645	0.219506
OLI	0	1	0.0232919	0.150835
ALC	0	1	0.0234114	0.151209
CAM	0	1	0.0398949	0.195706
SAM	0	1	0.0840898	0.27753
MIS	0	1	0.0815815	0.273723
MAR	0	1	0.0154085	0.123168
SAC	0	1	0.00919732	0.095458
ARR	0	1	0.101171	0.301563
BEM	0	1	0.0311753	0.173795
SAN	0	1	0.0653368	0.247125
SDM	0	1	0.0330865	0.178865
PAN	0	1	0.0398949	0.195708
AJU	0	1	0.0197086	0.138992
AVE	0	1	0.0795509	0.270598
ALV	0	1	0.0327281	0.177919
LUM	0	1	0.0395366	0.194861
BEA	0	1	0.010989	0.104256

BEL	0	1	0.0372671	0.189419
ARE	0	1	0.0249642	0.156019
CRN	0	1	0.00764453	0.0870951
MAU	0	1	0.406474	0.491164
RAZ	0	1	0.205686	0.40421
MED	0	1	0.283923	0.450891
BOM	0	1	0.0788342	0.26949
MBOM	0	1	0.0250836	0.156371
AR	35	400	117.068	64.61

1.4 Perceptron layers

The number of perceptron layers in the neural network is 2. The following table depicts the size of each layer and its corresponding activation function.

	Inputs number	Neurons number	Activation function
1	30	3	HyperbolicTangent
2	3	1	Linear

1.5 Unscaling layer

The size of the unscaling layer is 1, the number of outputs. The unscaling method for this layer is the Minimum - Maximum. The following table shows the values which are used for scaling the inputs, which include the minimum, maximum, mean and standard deviation.

	Minimum	Maximum	Mean	Deviation
VAL	1364	11992	5047.97	1807.82

1.6 Outputs table

The number of outputs is 1. The next table depicts some basic information about them, including the name, the units and the description.

	Name
1	VAL

2 Training strategy

2.1 Training strategy

The procedure used to carry out the learning process is called training(or learning) strategy. The training strategy is applied to the neural network in order to obtain the best possible loss.

2.2 Loss index

The loss index defines the task the neural network is required to do and provides a measure of the quality of the representation required to learn. When setting a loss index, two different terms must be chosen: an

error term and a regularization term.

The error term evaluates quantitatively how the neural network fits the data set. There are several error methods, and the choice of an appropriate one depends on the particular application. In this case, the Normalized Squared Error (MSE) is selected. The normalized squared error has a value of one when the outputs from the neural network are equal to the mean values of the target variables, while a value of zero means perfect prediction of the data.

The regularization term measures the values of the parameters in the neural network. Adding it to the error will cause the neural network to have smaller weights and biases, which will force its response to be smoother, i.e., to avoid overfitting. In this case, L2 regularization method is applied. It consists of the squared sum of all the parameters in the neural network.

2.3 Optimization algorithm

The quasi-Newton method is used here as optimization algorithm. It is based on Newton's method, but does not require calculation of second derivatives. Instead, the quasi-Newton method computes an approximation of the inverse Hessian at each iteration of the algorithm, by only using gradient information.

	Description	Value
Inverse hessian approximation method	Method used to obtain a suitable training rate.	BFGS
Learning rate method	Method used to calculate the step for the quasi-Newton training direction.	learning_rate_method
Learning rate tolerance	Maximum interval length for the learning rate.	0.001000
Minimum parameters increment norm	Norm of the parameters increment vector at which training stops.	0.000000
Minimum loss decrease	Minimum loss improvement between two successive epochs.	0.000000
Loss goal	Goal value for the loss.	0.000000
Gradient norm goal	Goal value for the norm of the objective function gradient.	0.000000
Maximum selection error increases	Maximum number of epochs at which the selection error increases.	100
Maximum epochs number	Maximum number of epochs to perform the training.	1000
Maximum time	Maximum training time.	3600.000000
Reserve training error history	Plot a graph with the loss of each epoch.	true
Reserve selection error history	Plot a graph with the selection error of each epoch.	false

3 Training

3.1 Task description

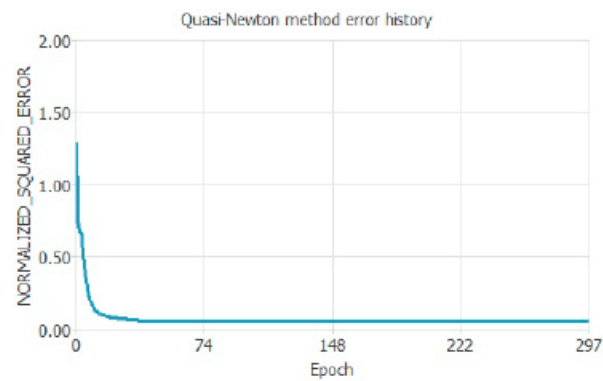
The procedure used to carry out the learning process is called training(or learning) strategy. The training strategy is applied to the neural network in order to obtain the best possible loss. The type of training is determined by the way in which the adjustment of the parameters in the neural network takes place.

3.2 Optimization algorithm

The quasi-Newton method is used here for training. It is based on Newton's method, but does not require calculation of second derivatives. Instead, the quasi-Newton method computes an approximation of the inverse Hessian at each iteration of the algorithm, by only using gradient information.

3.3 Quasi-Newton method error history

The following chart shows how the error decreases with the epochs during the training process. The initial value is 1.28138, and the final value after 297 epochs is 0.055885.



3.4 Quasi-Newton method results

The next table shows the training results by the quasi-Newton method. They include some final states from the neural network, the loss functional and the optimization algorithm.

	Value
Final parameters norm	1.5
Final training error	0.0559
Final selection error	6.47e-27
Final gradient norm	0.00196
Epochs number	297
Elapsed time	00:00:01
Stopping criterion	Minimum parameters increment norm

4 Model selection

4.1 Task description

Model selection is applied to find a neural network with a topology that optimizes the error on new data.

There are two different types of algorithms for model selection: Order selection algorithms and input selection algorithms. Order selection algorithms are used to find the optimal number of hidden neurons in the network. Inputs selection algorithms are responsible for finding the optimal subset of input variables.

4.2 Inputs selection algorithm

The inputs selection algorithm chosen for this application is growing inputs. With this method, the inputs are added progressively based on their correlations with the targets.

	Description	Value
Trials number	Number of trials for each neural network.	3
Tolerance	Tolerance for the selection error in the training of the algorithm.	0.01
Selection loss goal	Goal value for the selection error.	0
Maximum selection failures	Maximum number of iterations at which the selection error increases.	100
Maximum inputs number	Maximum number of inputs in the neural network.	30
Minimum correlations	Minimum value for the correlations to be considered.	0
Maximum correlation	Maximum value for the correlations to be considered.	1
Maximum iterations number	Maximum number of iterations to perform the algorithm.	1000
Maximum time	Maximum time for the inputs selection algorithm.	3600
Plot training loss history	Plot a graph with the training errors of each iteration.	true
Plot selection error history	Plot a graph with the selection errors of each iteration.	true

4.3 Order selection algorithm

The order selection algorithm chosen for this application is growing neurons. This method start with the minimum order and adds a given number of perceptrons in each iteration.

	Description	Value
Minimum neurons	Minimum number of hidden perceptrons to be evaluated.	3
Maximum neurons	Maximum number of hidden perceptrons to be evaluated.	10
Step	Number of hidden perceptrons added in each iteration.	1
Trials number	Number of trials for each neural network.	3
Tolerance	Tolerance for the selection error in the trainings of the algorithm.	0
Selection loss goal	Goal value for the selection error.	0
Maximum selection failures	Maximum number of iterations at which the selection error increases.	100
Maximum iterations number	Maximum number of iterations to perform the algorithm.	1000
Maximum time	Maximum time for the order selection algorithm.	3600
Plot training error history	Plot a graph with the training error of each iteration.	true
Plot selection error history	Plot a graph with the selection error of each iteration.	true

5 Order selection

5.1 Task description

The best selection is achieved by using a model whose complexity is the most appropriate to produce an adequate fit of the data. The order selection algorithm is responsible of finding the optimal number of neurons in the network.

5.2 Neurons selection algorithm

Growing neurons algorithm is used here as order selection algorithm in the model selection.

5.3 Growing order results

The next table shows the order selection results by the growing neurons algorithm. They include some final states from the neural network, the error functional and the order selection algorithm.

	Value
Optimal order	3
Optimum training error	0.0558359
Optimum selection error	-4.17181e+30
Iterations number	1
Elapsed time	00:00:03

6 Neural network outputs

6.1 Task description

A neural network produces a set of outputs for each set of inputs applied. The outputs depend, in turn, on the values of the parameters.

6.2 Inputs-outputs table

The next table shows the input values and their corresponding output values. The input variables are EST, SVC, CMP, PEF, OLI, ALC, CAM, SAM, MIS, MAR, SAC, ARR, BEM, SAN, SDM, PAN, AJU, AVE, ALV, LUM, BEA, BEL, ARE, CRN, MAU, RAZ, MED, BOM, MBOM and AR; and the output variable is VAL.

	Value
EST	1
SVC	0
CMP	0
PEF	0
OLI	0
ALC	0
CAM	0

SAM	0
MIS	0
MAR	0
SAC	0
ARR	0
BEM	0
SAN	0
SDM	0
PAN	0
AJU	0
AVE	0
ALV	0
LUM	0
BEA	0
BEL	0
ARE	0
CRN	0
MAU	1
RAZ	0
MED	0
BOM	0
MBOM	0
AR	117.068001
VAL	5065.80127

7 Directional output

7.1 Task description

It is very useful to see the how the outputs vary as a function of a single input, when all the others are fixed. This can be seen as the cut of the neural network model along some input direction and through some reference point.

7.2 Reference point table

The next table shows the reference point for the plots.

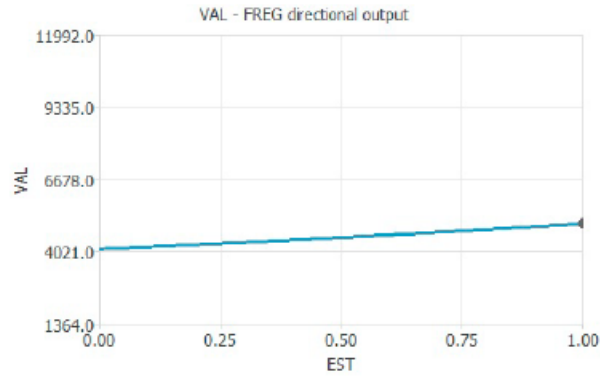
	Value
--	-------

8

EST	1
SVC	0
CMP	0
PEF	0
OLI	0
ALC	0
CAM	0
SAM	0
MIS	0
MAR	0
SAC	0
ARR	0
BEM	0
SAN	0
SDM	0
PAN	0
AJU	0
AVE	0
ALV	0
LUM	0
BEA	0
BEL	0
ARE	0
CRN	0
MAU	1
RAZ	0
MED	0
BOM	0
MBOM	0
AR	117

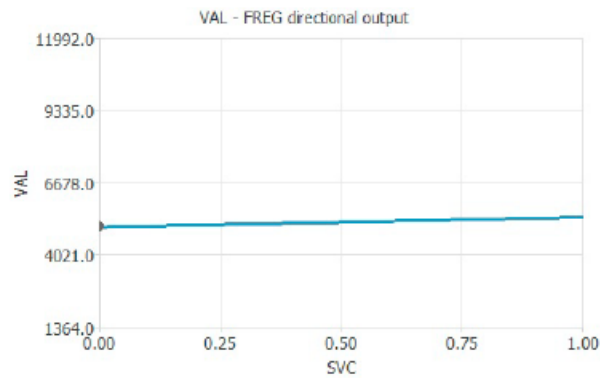
7.3 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



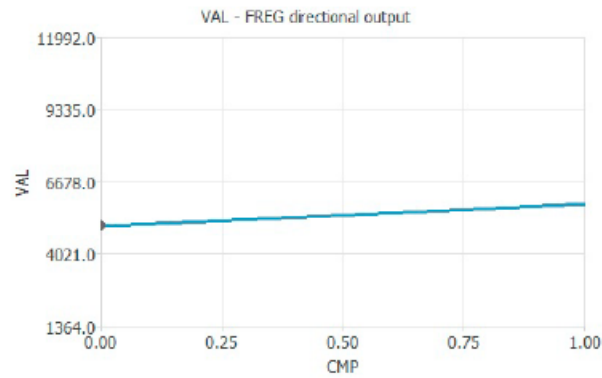
7.4 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



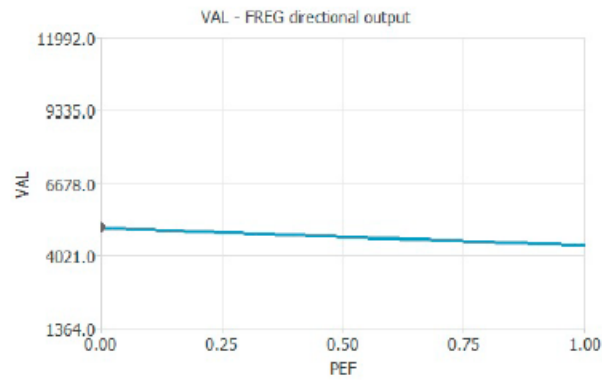
7.5 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



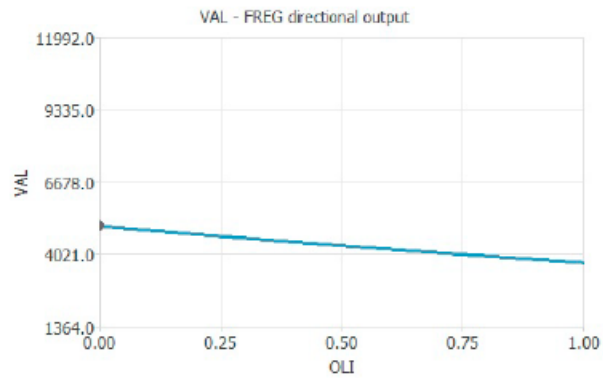
7.6 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



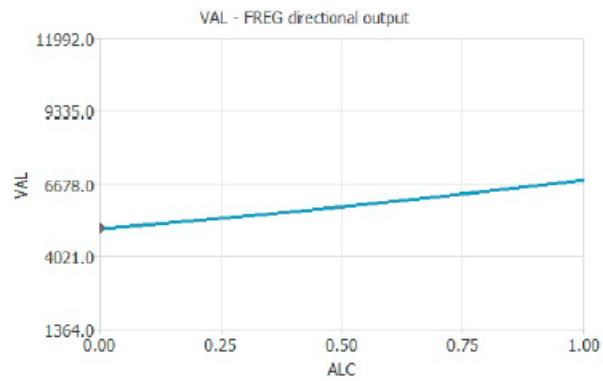
7.7 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



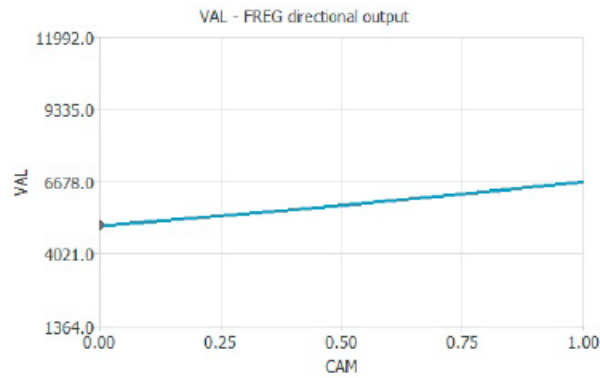
7.8 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



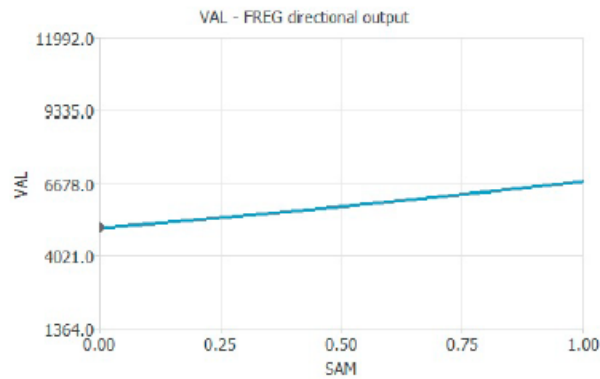
7.9 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



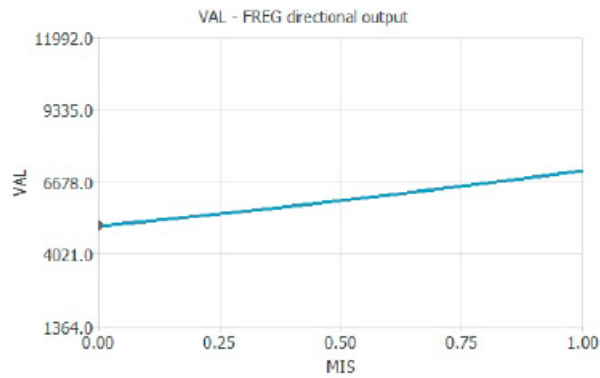
7.10 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



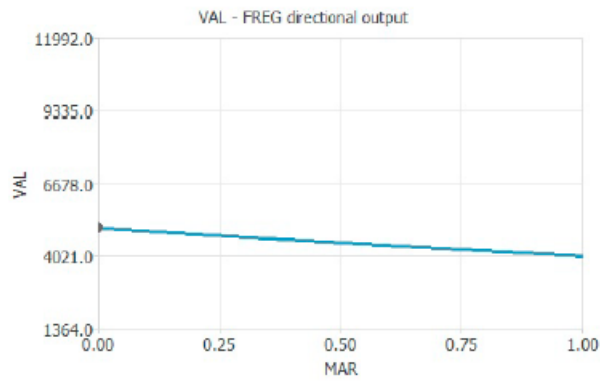
7.11 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



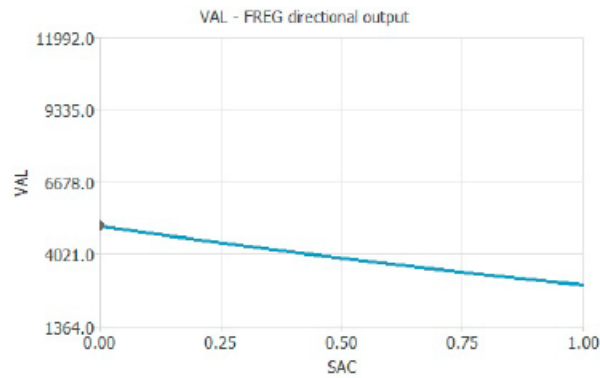
7.12 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



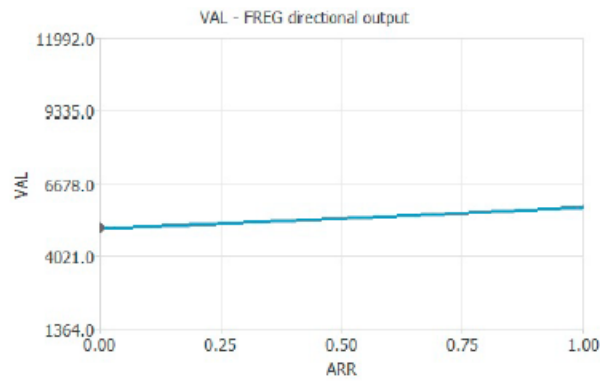
7.13 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



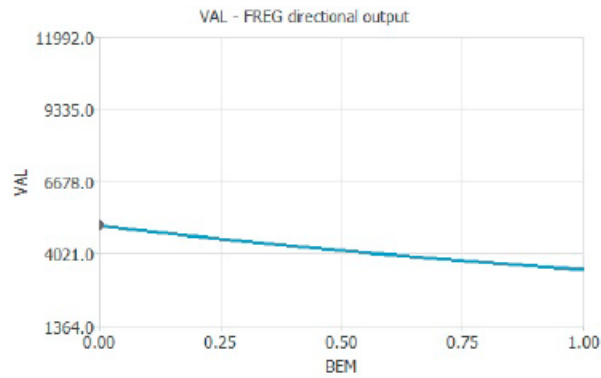
7.14 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



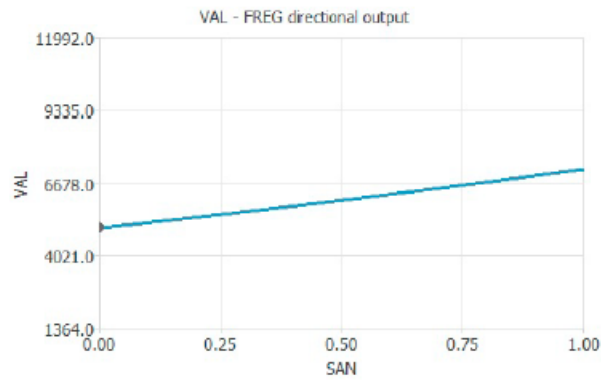
7.15 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



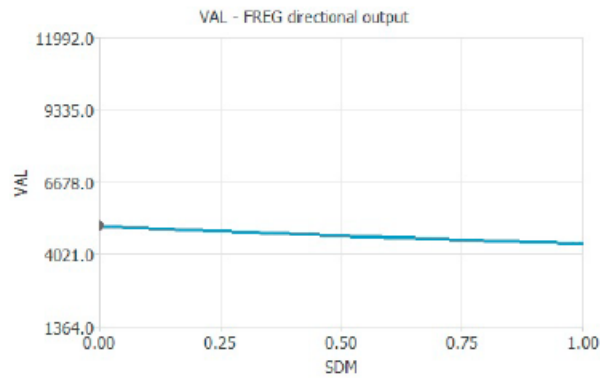
7.16 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



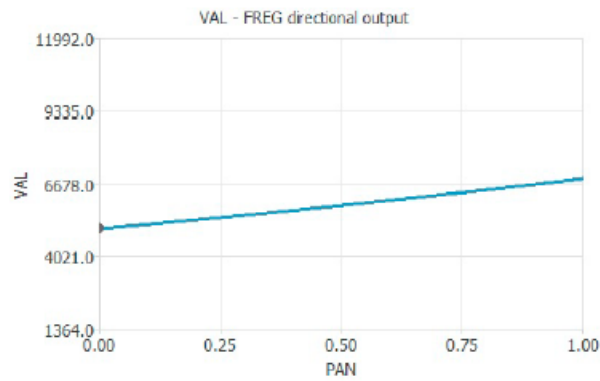
7.17 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



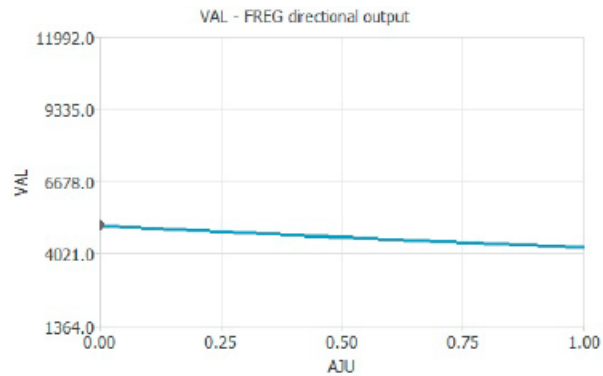
7.18 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



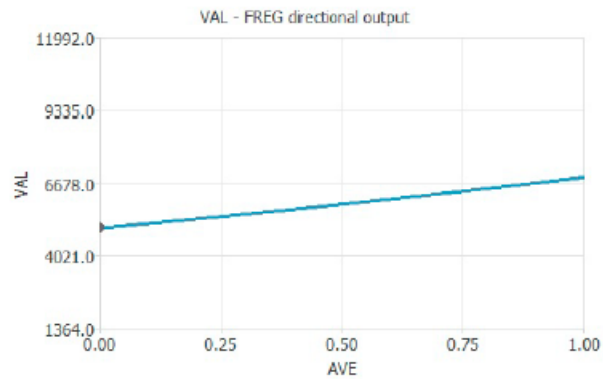
7.19 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



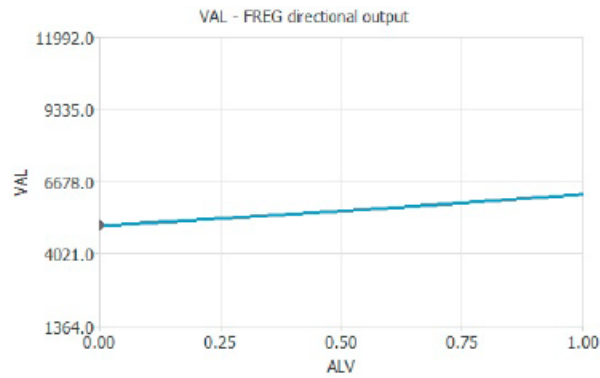
7.20 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



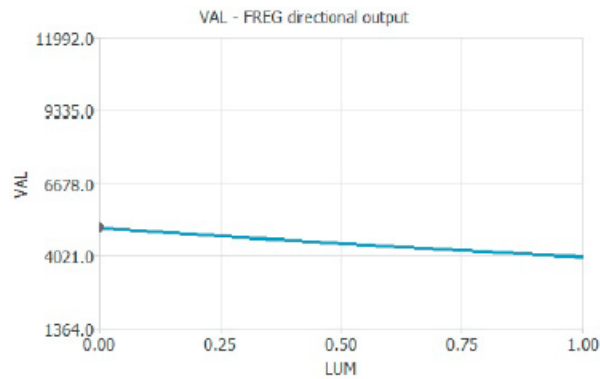
7.21 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



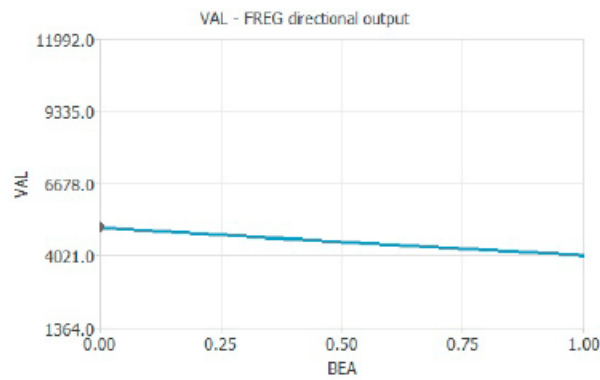
7.22 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



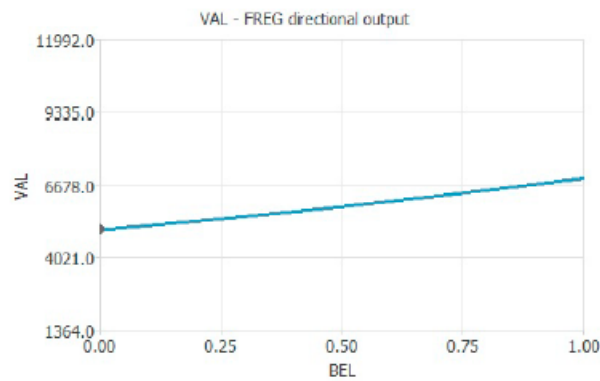
7.23 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



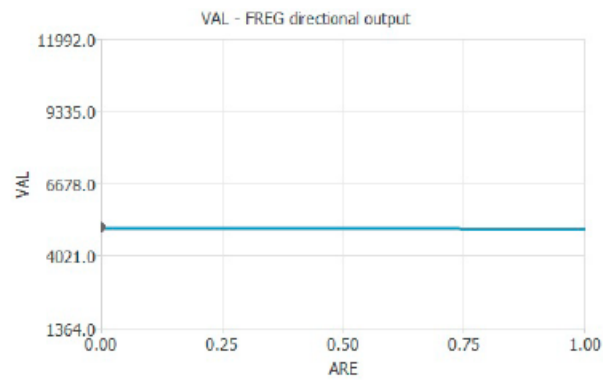
7.24 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



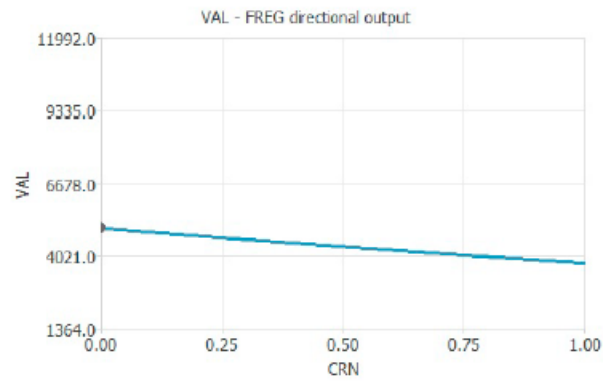
7.25 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



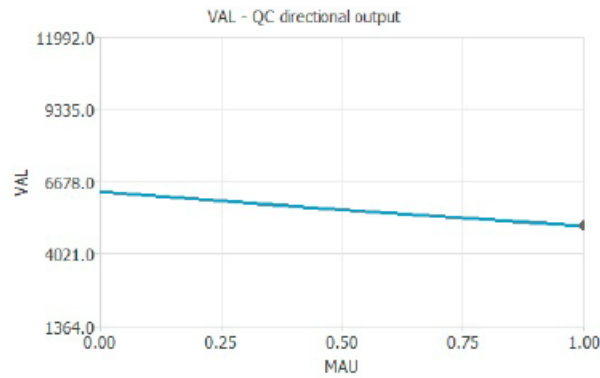
7.26 VAL - FREG directional output

The next plot shows the output VAL as a function of the input FREG. The x and y axes are defined by the range of the variables FREG and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



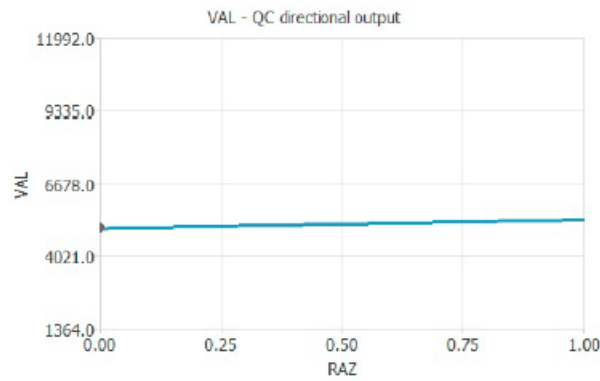
7.27 VAL - QC directional output

The next plot shows the output VAL as a function of the input QC. The x and y axes are defined by the range of the variables QC and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



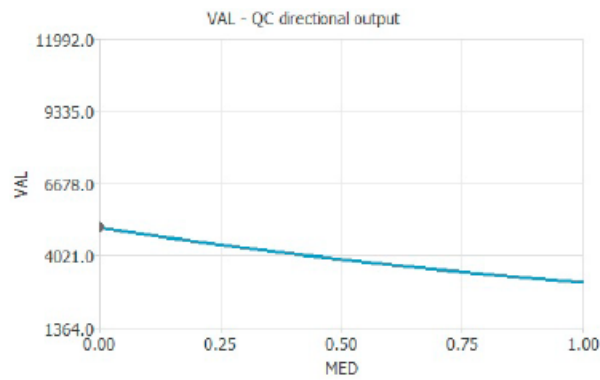
7.28 VAL - QC directional output

The next plot shows the output VAL as a function of the input QC. The x and y axes are defined by the range of the variables QC and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



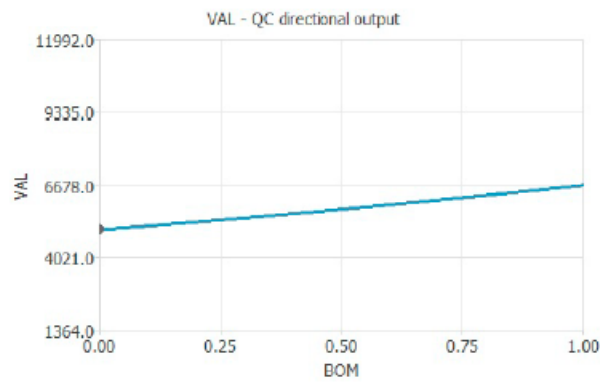
7.29 VAL - QC directional output

The next plot shows the output VAL as a function of the input QC. The x and y axes are defined by the range of the variables QC and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



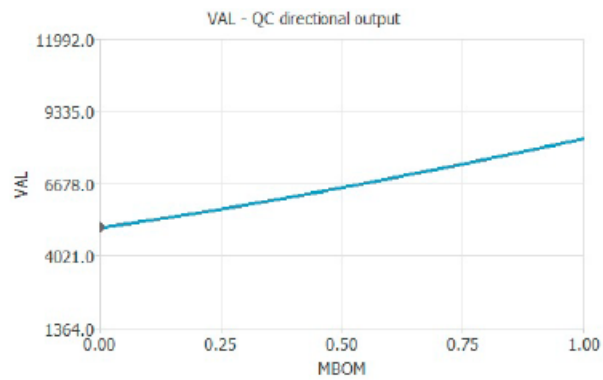
7.30 VAL - QC directional output

The next plot shows the output VAL as a function of the input QC. The x and y axes are defined by the range of the variables QC and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



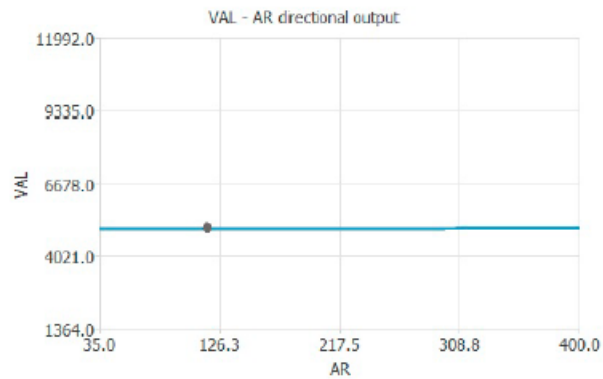
7.31 VAL - QC directional output

The next plot shows the output VAL as a function of the input QC. The x and y axes are defined by the range of the variables QC and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



7.32 VAL - AR directional output

The next plot shows the output VAL as a function of the input AR. The x and y axes are defined by the range of the variables AR and VAL, respectively. The gray point represents the reference point. Note that some directional outputs fall outside the range of VAL, and therefore they are not plotted.



8 Mathematical expression

8.1 Task description

The predictive model takes the form of a function of the outputs with respect to the inputs. The mathematical expression represented by the model can be used to embed it into another software, in the so called production mode.

8.2 Expression

The mathematical expression represented by the neural network is written below. It takes the inputs EST, SVC, CMP, PEF, OLI, ALC, CAM, SAM, MIS, MAR, SAC, ARR, BEM, SAN, SDM, PAN, AJU, AVE, ALV, LUM, BEA, BEL, ARE, CRN, MAU, RAZ, MED, BOM, MBOM, AR, to produce the output VAL. In approximation models, the information is propagated in a feed-forward fashion through the scaling layer, the perceptron layers and the unscaling layer.

```

scaled_EST = EST*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_SVC = SVC*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_CMP = CMP*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_PEF = PEF*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_OLI = OLI*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_ALC = ALC*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_CAM = CAM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_SAM = SAM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_MIS = MIS*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_MAR = MAR*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_SAC = SAC*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_ARR = ARR*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_BEM = BEM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_SAN = SAN*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_SDM = SDM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_PAN = PAN*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_AJU = AJU*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_AVE = AVE*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_ALV = ALV*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_LUM = LUM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_BEA = BEA*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_BEL = BEL*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_ARE = ARE*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_CRN = CRN*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_MAU = MAU*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_RAZ = RAZ*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_MED = MED*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_BOM = BOM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_MBOM = MBOM*(1+1)/(1-(0))-0*(1+1)/(1-0)-1;
scaled_AR = AR*(1+1)/(400-(35))-35*(1+1)/(400-35)-1;

perceptron_layer_0_output_0 = tanh[ -0.0273545 + (scaled_EST*0.
145978)+ (scaled_SVC*0.0484602)+ (scaled_CMP*0.0903133)+
(scaled_PEF*-0.0837136)+ (scaled_OLI*-0.157173)+ (scaled_ALC*0.
23075)+ (scaled_CAM*0.175239)+ (scaled_SAM*0.197303)+
(scaled_MIS*0.21412)+ (scaled_MAR*-0.121374)+ (scaled_SAC*-0.
226487)+ (scaled_ARR*0.122935)+ (scaled_BEM*-0.258941)+
(scaled_SAN*0.200095)+ (scaled_SDM*-0.135994)+ (scaled_PAN*0.
195899)+ (scaled_AJU*-0.122241)+ (scaled_AVE*0.181573)+
(scaled_ALV*0.135663)+ (scaled_LUM*-0.147329)+ (scaled_BEA*-0.
089875)+ (scaled_BEL*0.236099)+ (scaled_ARE*-0.0819445)+
(scaled_CRN*-0.143974)+ (scaled_MAU*-0.145977)+ (scaled_RAZ*0.
0155515)+ (scaled_MED*-0.321467)+ (scaled_BOM*0.194483)+
(scaled_MBOM*0.336823)+ (scaled_AR*-0.00182084) ];
perceptron_layer_0_output_1 = tanh[ -0.0215225 + (scaled_EST*0.

```

25

```
00150197)+ (scaled_SVC*0.0144937)+ (scaled_CMP*0.00424944)+  
(scaled_PEF*0.017386)+ (scaled_OLI*0.04919)+ (scaled_ALC*0.  
0265654)+ (scaled_CAM*0.00475437)+ (scaled_SAM*0.00137881)+  
(scaled_MIS*0.0043579)+ (scaled_MAR*0.0444562)+ (scaled_SAC*0.  
112635)+ (scaled_ARR*0.0216203)+ (scaled_BEM*0.0387804)+  
(scaled_SAN*0.0285445)+ (scaled_SDM*0.00836119)+ (scaled_PAN*0.  
00671644)+ (scaled_AJU*0.0237183)+ (scaled_AVE*0.0193724)+  
(scaled_ALV*0.00271408)+ (scaled_LUM*0.0372086)+ (scaled_BEA*0.  
0533647)+ (scaled_BEL*0.0246722)+ (scaled_ARE*0.00998793)+  
(scaled_CRN*0.0513657)+ (scaled_MAU*0.0385684)+ (scaled_RAZ*0.  
0181201)+ (scaled_MED*0.024089)+ (scaled_BOM*0.00115592)+  
(scaled_MBOM*0.0198193)+ (scaled_AR*0.0114953) ];  
perceptron_layer_0_output_2 = tanh[ 0.000105209 + (scaled_EST*0.  
0431895)+ (scaled_SVC*0.0127711)+ (scaled_CMP*0.04477)+  
(scaled_PEF*0.0308506)+ (scaled_OLI*0.101905)+ (scaled_ALC*0.  
0114294)+ (scaled_CAM*0.0674428)+ (scaled_SAM*0.0505643)+  
(scaled_MIS*0.0844668)+ (scaled_MAR*0.067452)+ (scaled_SAC*0.  
208866)+ (scaled_ARR*0.00916232)+ (scaled_BEM*0.0918329)+  
(scaled_SAN*0.131406)+ (scaled_SDM*0.0160158)+ (scaled_PAN*0.  
0764724)+ (scaled_AJU*0.0251097)+ (scaled_AVE*0.103947)+  
(scaled_ALV*0.0422868)+ (scaled_LUM*0.055665)+ (scaled_BEA*0.  
0941168)+ (scaled_BEL*0.0218961)+ (scaled_ARE*0.0845755)+  
(scaled_CRN*0.0999329)+ (scaled_MAU*0.0330524)+ (scaled_RAZ*0.  
050392)+ (scaled_MED*0.1529)+ (scaled_BOM*0.0393776)+  
(scaled_MBOM*0.0979794)+ (scaled_AR*0.00865001) ];  
  
perceptron_layer_1_output_0 = [ 0.168256 +  
(perceptron_layer_0_output_0*0.91462)+  
(perceptron_layer_0_output_1*0.192488)+  
(perceptron_layer_0_output_2*0.445754) ];  
  
unscaling_layer_output_0 =  
perceptron_layer_1_output_0*(11992-1364)/  
(1+1)+1364+1*(11992-1364)/(1+1);
```

9 Output data

9.1 Task description

The output data defines a set of output vectors from a set of input vectors, the input data. The output data results have been saved in the following file: C:/Users/bruno/OneDrive - Instituto Superior de Engenharia de Lisboa/ISEL/TFM/Amostras/NEURAL DESIGNER SMPL/Ensaios de RNA com ND/L2_+indice de perda/output_Test_V24.csv.

9.2 Output statistics

The table below shows the minimums, maximums, means and standard deviations of all the output predicted.

	Minimum	Maximum	Mean	Deviation
VAL	1757	1.084e+4	5102	1729

9.3 VAL histogram

The following chart shows the histogram for the output VAL. The abscissa represents the bins numbers, and the ordinate their corresponding frequencies. The maximum frequency is 551, which corresponds to the bin with center 4027. The minimum frequency is 31, which corresponds to the bin with center 1.038e+4.



10 Training strategy

10.1 Training strategy

The procedure used to carry out the learning process is called training(or learning) strategy. The training strategy is applied to the neural network in order to obtain the best possible loss.

10.2 Loss index

The loss index defines the task the neural network is required to do and provides a measure of the quality of the representation required to learn. When setting a loss index, two different terms must be chosen: an error term and a regularization term.

The error term evaluates quantitatively how the neural network fits the data set. There are several error methods, and the choice of an appropriate one depends on the particular application. In this case, the Normalized Squared Error (MSE) is selected. The normalized squared error has a value of one when the outputs from the neural network are equal to the mean values of the target variables, while a value of zero means perfect prediction of the data.

The regularization term measures the values of the parameters in the neural network. Adding it to the error will cause the neural network to have smaller weights and biases, which will force its response to be smoother, i.e., to avoid overfitting. In this case, L2 regularization method is applied. It consists of the squared sum of all the parameters in the neural network.

10.3 Optimization algorithm

The optimization algorithm chosen for this application is gradient descent. With this method, the neural parameters are updated in the direction of the negative gradient of the loss function.

	Description	Value
Learning rate method	Training rate algorithm for one-dimensional minimization.	BrentMethod
Learning rate tolerance	Maximum interval length for the training rate.	0.001000
Minimum parameters increment norm	Norm of the parameters increment vector at which training stops.	0.000000
Minimum loss decrease	Minimum loss improvement between two successive epochs.	0.000000
Loss goal	Goal value for the loss.	0.001000
Gradient norm goal	Goal value for the norm of the objective function gradient.	0.000000
Maximum selection error increases	Maximum number of epochs at which the selection error increases.	100
Maximum epochs number	Maximum number of epochs to perform the training.	1000
Maximum time	Maximum training time.	3600.000000
Reserve training error history	Plot a graph with the loss of each epoch.	true
Reserve selection error history	Plot a graph with the selection error of each epoch.	false

11 Training

11.1 Task description

The procedure used to carry out the learning process is called training(or learning) strategy. The training strategy is applied to the neural network in order to obtain the best possible loss. The type of training is determined by the way in which the adjustment of the parameters in the neural network takes place.

11.2 Optimization algorithm

Gradient descent is used here for training. With this method, the neural parameters are updated in the direction of the negative gradient of the loss function.

11.3 Gradient descent results

The next table shows the training results by the gradient descent method. They include some final states from the neural network, the loss functional and the optimization algorithm. Note that the number of epochs needed to converge is zero, which means that the optimization algorithm did not modify the state of the neural network.

	Value
Final parameters norm	1.5
Final training error	0.0558

Final selection error	-1.8
Final gradient norm	0.00129
Epochs number	0
Elapsed time	00:00:00
Stopping criterion	Minimum parameters increment norm