

INSTITUTO TECNOLÓGICO DE AERONÁUTICA



Matthaeus Muniz Deusdará Ferreira Lopes

**ESTIMATIVA DE CUSTO DE OBRAS COM A
UTILIZAÇÃO DE REDES NEURAIAS ARTIFICIAIS**

Trabalho de Graduação
2017

Curso de Engenharia Civil-Aeronáutica

Matthaeus Muniz Deusdará Ferreira Lopes

**ESTIMATIVA DE CUSTO DE OBRAS COM A
UTILIZAÇÃO DE REDES NEURAIIS
ARTIFICIAIS**

Orientador

Maj Eng FRANK Cabral de Freitas Amaral (CO-DCTA)

Co-orientadores

Maj Eng Robinson SAMUEL Boschetti (CO-DCTA)

ENGENHARIA CIVIL-AERONÁUTICA

SÃO JOSÉ DOS CAMPOS
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

2017

Dados Internacionais de Catalogação-na-Publicação (CIP)

Divisão de Informação e Documentação

Lopes, Matthaeus Muniz Deusdará Ferreira
Estimativa de Custo de Obras com a Utilização de Redes Neurais Artificiais / Matthaeus Muniz Deusdará Ferreira Lopes.
São José dos Campos, 2017.
68f.

Trabalho de Graduação – Engenharia Civil-Aeronáutica – Instituto Tecnológico de Aeronáutica, 2017. Orientador: Major Engenheiro Frank Cabral de Freitas Amaral. Co-orientador: Major Engenheiro Robinson Samuel Boschetti.

1. Redes Neurais Artificiais. 2. Estimativa Paramétrica de Obras. 3. *Python*. I. Instituto Tecnológico de Aeronáutica. II. Estimativa de Custo de Obras com a Utilização de Redes Neurais Artificiais

REFERÊNCIA BIBLIOGRÁFICA

LOPES, Matthaeus Muniz Deusdará Ferreira. **Estimativa de Custo de Obras com a Utilização de Redes Neurais Artificiais**. 2017. 68f. Trabalho de Conclusão de Curso. (Graduação em Engenharia Civil-Aeronáutica) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

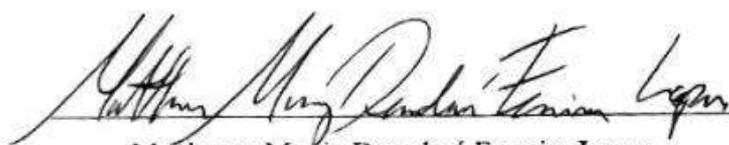
NOME DO AUTOR : Matthaeus Muniz Deusdará Ferreira Lopes
TÍTULO DO TRABALHO : ESTIMATIVA DE CUSTO DE OBRAS COM A UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS
TIPO DO TRABALHO/ANO : Graduação / 2017

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho de graduação pode ser reproduzida sem a autorização do autor.


Matthaeus Muniz Deusdará Ferreira Lopes
Rua H8B, apto 232
12228-460, São José dos Campos - SP

ESTIMATIVA DE CUSTO DE OBRAS COM A UTILIZAÇÃO DE REDES NEURAIIS ARTIFICIAIS

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação



Maíthaeus Muniz Deusdará Ferreira Lopes

Autor



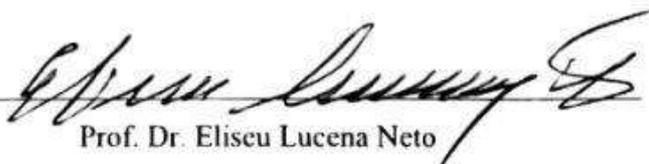
Maj Eng Frank Cabral de Freitas Amaral (CO-DCTA)

Orientador



Maj Eng Robinson SAMUEL Boschetti (CO-DCTA)

Coorientador



Prof. Dr. Eliseu Lucena Neto

Coordenador do Curso de Engenharia Civil-Aeronáutica

São José dos Campos, 09 de Novembro de 2017

Dedico este trabalho a todas as pessoas que passaram pela minha vida e me tornaram uma pessoa melhor.

Agradecimentos

Agradeço a colaboração de toda a equipe do CO-DCTA, onde eu pude viver o dia a dia da engenharia na Força Aérea Brasileira, ao lado de excelentes profissionais, e em especial ao Major Frank que me orientou sobre as possibilidades na carreira militar, e ao Major Samuel, que me orientou durante este trabalho.

Agradeço à toda a equipe do CEPE, principalmente ao Coronel Carvalho e ao Tenente Castro, e à equipe do Major Telechi, do CISCEA, pela disponibilidade, energia e profissionalismo na discussão deste trabalho e fornecimento dos projetos das respectivas unidades.

Agradeço à Equipe da DAE, sobretudo à professora Cristiane Lacaz, que me auxiliou em momentos chave na minha trajetória neste Instituto.

Agradeço aos professores que forneceram o substrato técnico para a elaboração deste TG:

- Professora Juliana, por durante as cadeiras de CES-10 e CCI-22 ter me mostrado quão poderosa a programação pode ser.

- Professor Paulo Ivo, pela nossa conversa sobre um possível substituto do Matlab, momento em que descobri a linguagem Python, a qual faço uso neste trabalho.

- Professor Tenente Coronel Márcio, por dar a oportunidade de me colocar em situação que era necessário o exercício da habilidade de programação nessa linguagem no estágio militar entre o terceiro e quarto ano.

- Professor Alex, que me apresentou o conceito de Redes Neurais Artificiais e suas utilidades na disciplina de EDI-37.

"Só existe um caminho para o sucesso:
viver a vida do seu próprio jeito".
(Christopher Morley)

Resumo

Na fase de planejamento e projeto de uma obra de engenharia, todos os custos previstos para sua realização compõem o seu orçamento, peça técnica fundamental que baliza a tomada de decisão e a própria contratação em si. Contudo, nem sempre os gestores dispõem de tempo hábil para revisar todos os itens que integram o orçamento estimativo da obra e, por vezes, pequenas falhas, como erros em composições de custo unitário ou em quantitativos, implicam enormes transtornos para a administração, culminando eventualmente até mesmo com danos ao erário, caso não sejam percebidos e saneados oportunamente. Tais prejuízos são ainda mais significativos quando tratamos de obras de grande e médio porte. Nesse sentido, o presente trabalho analisou comparativamente uma amostra de obras, em termos dos respectivos orçamentos, buscando identificar parâmetros representativos do custo e propôs um modelo, baseado em Redes Neurais Artificiais, com o intuito de facilitar a revisão sistemática a nível macro de planilhas orçamentárias, fornecendo um parâmetro para verificação da assertividade de orçamentos de futuras obras, de sorte a mitigar possíveis desvios.

Palavras-chave: Orçamento, Revisão, Python, TensorFlow, Keras, Scikit-learn, Redes Neurais Artificiais, Estimativa Paramétrica de Obras

Abstract

At the planning and project costing phase, all of those planned for the completion of a work make up your budget, which is leads the contracted price. The managers do not always have the ability to revise the workability of parts of the estimated budget of the work, and sometimes small failures in compositions of unit cost imply enormous damage to the treasury, due to the gain of scale, particularly in the case of works the objective of this work is to analyze comparatively a bunch of works, in terms of the respective analytical budgets, in order to identify representative parameters and provide an Artificial Neural Networks model, with the purpose of easing a systematic review on macro level of the budget spreadsheets to mitigate possible deviations.

Keywords: Budget, Review, Python, TensorFlow, Keras, Scikit-learn, Artificial Neural Networks, Parameter Estimate

Lista de Figuras

Figura 1 Passo-a-passo do TG.....	15
Figura 2 Ilustração da capacidade de influência no custo total da obra de cada fase.....	17
Figura 3 Exemplo de tabela para consulta do CUB.	18
Figura 4 Representação de um Neurônio Biológico.....	26
Figura 5 Neurônio Computacional	28
Figura 6 Exemplo de Rede Neural Artificial.....	29
Figura 7 Representação de Rede Neural Feedforward	35
Figura 8 Resultado da Configuração 01	38
Figura 9 Resultado da Configuração 02	39
Figura 10 Resultado da Configuração 03	39
Figura 11 Regressão Linear (Excel)	41
Figura 12 Ilustração do overfitting	43
Figura 13 Alojamento de Atletas - UNIFA	46
Figura 14 Telas de Abertura e de Estimativa, respectivamente	48

Lista de Tabelas

Tabela 1 Relação de parâmetros construtivos com componente relacionado.	23
Tabela 2 Relação de parâmetros construtivos com componente relacionado.	23
Tabela 3 Funções de ativação mais comuns	27
Tabela 4 Resultados dos primeiros testes.	40
Tabela 5 Segunda bateria de testes: com área e perímetro como entradas	41
Tabela 6 Terceira bateria de testes: com todas as variáveis disponíveis	42
Tabela 7 Quarta rodada de testes: influência da área, em BD2	44
Tabela 8 Quinta rodada de testes: todas variáveis de BD2	44
Tabela 9 Última rodada de testes: todas variáveis de BD3	45
Tabela 10 Estimativas do prédio da UNIFA	47
Tabela 11 Compilado do Banco de dados BD1	53
Tabela 12 Compilado do Banco de dados BD2	58
Tabela 13 BD3: Composição das amostras de BD1 e BD2	60

Lista de Abreviaturas e Siglas

CUB	Custo Unitário Básico
ITA	Instituto Tecnológico de Aeronáutica
TensorFlow	Biblioteca de opensource para aprendizagem de máquinas
Keras	Biblioteca de rede neural de código aberto escrita em Python
MATLAB	Software de computação numérica da empresa MathWorks
TG	Trabalho de Graduação
INCC	Índice Nacional da Construção Civil
BD1	Banco de Dados 1
BD2	Banco de Dados 2
BD3	Banco de Dados 3
Sinduscon	Sindicato da Indústria da Construção Civil
IGP-M	Índice Geral de Preços do Mercado
CUPE	Custos Unitários PINI de Edificações
SINAPI	Sistema Nacional de Pesquisa de Custos e Índices da Construção Civil
COMARA	Comissão de Aeroportos da Região Amazônica
CLA	Centro de Lançamento de Alcântara
CEPE	Centro de Estudos e Projetos de Engenharia
CISCEA	Comissão de Implantação do Sistema de Controle do Espaço Aéreo
TCU	Tribunal de Contas da União
bots	Em um jogo, um adversário com recurso a inteligência artificial

Sumário

1	INTRODUÇÃO	14
1.1	Importância.....	14
1.2	Estruturação do trabalho e metodologia.....	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Atuais Métodos para Estimar o Custo de uma Obra.....	16
2.1.1	Estimativa por Área Superficial: CUB e NBR 12721	18
2.1.2	Método da Unidade	20
2.1.3	Método da Estimativa por Volume.....	20
2.1.4	Estimativa do Envolvimento dos Pavimentos.....	21
2.1.5	Método Paramétrico	22
2.2	Conceitos Básicos de Redes Neurais Artificiais	26
3	COLETA DE DADOS.....	30
3.1	Escolha das Variáveis Elegíveis.....	30
3.2	Tratamento dos Dados	32
4	DESENVOLVIMENTO.....	33
4.1	Criação da Rede Neural.....	33
4.1.1	Pré-Processamento dos Dados.....	33
4.1.2	Determinação da Arquitetura da Rede Neural.....	34
4.1.3	Treinamento da rede	35
4.1.4	Métricas para avaliação dos resultados	36
5	ANÁLISE DOS RESULTADOS.....	38
5.1	Resultados e Interpretação	38
5.2	Análise em caso particular	46
5.3	Confecção de Aplicativo para Smartphones	47
5.4	Críticas e Limitações	48
6	CONCLUSÃO.....	50
7	REFERÊNCIAS BIBLIOGRÁFICAS	51
	APÊNDICE A – BANCOS DE DADOS INDIVIDUAIS	53
	APÊNDICE B – BANCO DE DADOS UNIFICADO.....	59
	APÊNDICE C – CÓDIGOS UTILIZADOS	61

1 Introdução

1.1 Importância

O desenvolvimento econômico de uma nação tem como premissa a existência de uma adequada infraestrutura, com a redução dos custos logísticos, e a melhoria da qualidade dos serviços públicos. Nesse contexto, as obras públicas devem ser tratadas com especial atenção por terem o papel estruturante no desenvolvimento de áreas relevantes para o país, como educação, saúde, transporte e moradia.

Assim sendo, o processo de confecção do orçamento é parte fundamental da elaboração do projeto da obra. Sua correta elaboração pode poupar muito tempo e dinheiro. Um grande problema, entretanto, é que a verificação dos cálculos do orçamento pode ser demorada.

Com esta visão, este trabalho analisará orçamentos e contratos de obras públicas e levantará parâmetros de comparação, de acordo com estudo de literatura, facilitando a revisão macro do orçamento e correção de erros pontuais na obra, ou seja, este trabalho visa à atenção à um método que pode gerar um ferramental de verificação e planejamento de orçamentos com base em resultados observados na prática pela Força Aérea Brasileira.

A metodologia desenvolvida se fundamentará no conceito de Rede Neural Artificial, para o qual será feito antes um levantamento das principais variáveis que afetam os custos, as quais servirão de dados de entrada nesse modelo.

1.2 Estruturação do trabalho e metodologia

Este trabalho começa por uma revisão bibliográfica dos métodos mais utilizados atualmente para estimativas conceituais, e introduz os conceitos básicos de Redes Neurais Artificiais.

Em seguida, prosseguimos com a coleta dos dados: escolha dos campos solicitados às fontes, metodologia de tratamento do banco de dados de projetos e formatação da referência adotada.

A etapa seguinte é o desenvolvimento do modelo. Isso envolve a utilização de variáveis independentes e representativas dos projetos, que servem de entrada para o treinamento da rede neural. O critério de seleção de parâmetros deve ser mais que somente

uma validação estatística, mas deve ser baseada em lógica, considerações teóricas e processo experimental.

Ainda, faz-se necessário estabelecer as limitações do modelo, o qual é desenvolvido a partir de um banco de dados limitado e, portanto, só é válido sob determinadas condições.



Figura 1 Passo-a-passo do TG

2 Revisão Bibliográfica

Para a compreensão deste trabalho, se faz necessária a contextualização sobre os principais métodos utilizados na precificação dos custos de uma obra, a fim de entender as variáveis envolvidas e podermos escolher apropriadamente quais delas irão compor nosso modelo.

Cabe a reflexão sobre o que seria uma boa variável. As variáveis que gerariam o resultado mais próximo do real seriam cada quantitativo da edificação e a eficiência da construtora em executar cada etapa da obra pelo custo e, principalmente, quantitativos estimados. Mas isso demora e é sujeito a falhas, em especial se não tivermos os detalhes do projeto, que forneceria esses quantitativos, o que ocorre costumeiramente nas etapas iniciais de uma obra. O que queremos aqui é uma forma mais rápida que essa, se possível mais precisa que métodos de estimativa por taxas simples, utilizando menos variáveis, as quais em contrapartida são gerais e impactantes no custo da construção como, por exemplo, o pé-direito e a área edificada.

Além disso, para a compreensão dos princípios básicos do método numérico utilizado nesta tese, a Rede Neural Artificial, elaborou-se um pequeno resumo dos conceitos-chave ao entendimento dessa ferramenta.

2.1 Atuais Métodos para Estimar o Custo de uma Obra

No momento de estimarmos o custo nos estágios iniciais de uma obra, pode-se recorrer a vários métodos, ou a combinação deles, e a escolha do qual utilizar dependerá da disponibilidade dos dados e de sua confiabilidade. De forma geral, podemos agrupá-las da seguinte forma:

Métodos de Estimativa Pré-proposta:

Estimativas Conceituais

Estimativa por Área Superficial

Estimativa pelo Método da Unidade

Estimativa por Volume

Estimativa Por Pavimento

Estimativa de Etapas posteriores

Estimativa Aproximada

Estimativa Elemental

Modelagem do Custo

Estimativa de Design/Engenharia

Estimativa detalhada de Execução

Método da Quantidade de Unidade

Método da Quantidade Total

Focaremos abaixo no primeiro grupo, o de Estimativas Conceituais porque são importantes ferramentas na administração dos custos de um projeto e vai nos fornecer as variáveis mais simples, gerais e de fácil obtenção, de métodos já empiricamente validados.

A estimativa conceitual é vital para decisões nas etapas iniciais do projeto, fase em que as alterações mais impactam as fases subseqüentes da obra, e para o planejamento e controle de custos.

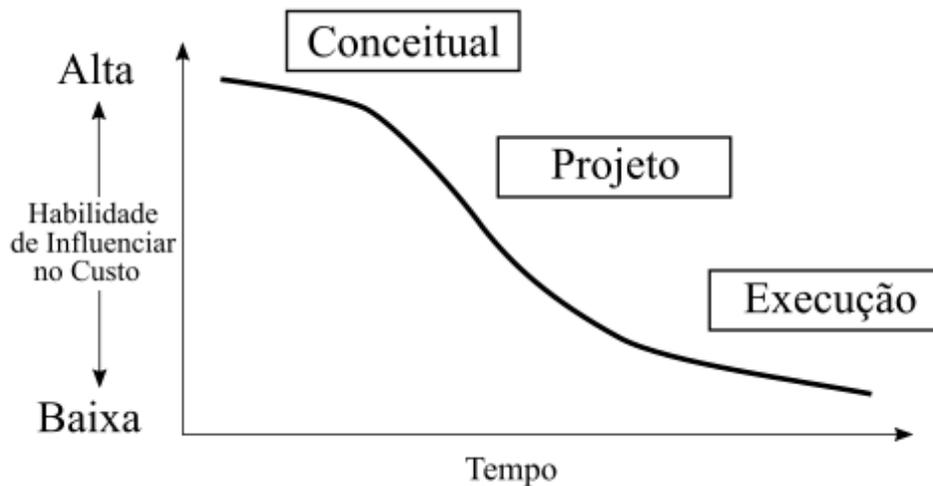


Figura 2 Ilustração da capacidade de influência no custo total da obra de cada fase.

(DASCHBACH, 1988)

Uma estimativa conceitual também é conhecida como uma estimativa top-down, ordem de magnitude, viabilidade, análoga ou preliminar. É o primeiro esforço sério feito na tentativa de prever o custo do projeto. Uma estimativa conceitual geralmente é realizada como parte da análise de viabilidade do projeto no início do projeto. Desta forma, a estimativa é feita com informações limitadas sobre o escopo do projeto, e geralmente é feita sem os

detalhamentos e dados de engenharia. Espera-se que a faixa de precisão seja de + 50% a -30% (PHAOBUNJONG, 2002).

Há de se atentar para as limitações de cada ferramenta dessa classe de estimativas. Cada projeto tem um método conceitual mais apropriado, que resulta em um custo realista. As peculiaridades de cada estimativa serão abordadas a seguir.

2.1.1 Estimativa por Área Superficial: CUB e NBR 12721

A sigla CUB significa custo por metro quadrado de construção do projeto-padrão considerado, e seu valor é calculado de acordo com metodologia estabelecida pelos Sindicatos da Indústria da Construção Civil, e serve de base para avaliação de parte dos custos de construção das edificações.

Basicamente, existe uma tabela de custo por metro quadrado, que leva em conta a função (residencial ou comercial) e o padrão de acabamento. Pode-se estimar o custo da construção fazendo a simples multiplicação da área construída pelo custo por metro quadrado.

CUB/RS do mês de SETEMBRO/2015 - NBR 12.721- Versão 2006

PROJETOS	Padrão de acabamento	Código	Custo R\$/m ²	Variação %		
				Mensal	Anual	12 meses
RESIDENCIAIS						
R - 1 (Residência Unifamiliar)	Baixo	R 1-B	1.238,23	0,41	7,81	8,13
	Normal	R 1-N	1.544,87	0,41	9,10	9,54
	Alto	R 1-A	1.927,81	0,42	8,95	9,60
PP (Prédio Popular)	Baixo	PP 4-B	1.127,45	0,33	7,54	8,26
	Normal	PP 4-N	1.475,30	0,39	8,33	8,76
R - 8 (Residência Multifamiliar)	Baixo	R 8-B	1.069,57	0,30	7,49	8,21
	Normal	R 8-N	1.274,39	0,35	8,54	9,06
	Alto	R 8-A	1.566,16	0,39	8,58	9,26
R - 16 (Residência Multifamiliar)	Normal	R 16-N	1.238,63	0,33	8,58	9,15
	Alto	R 16-A	1.601,82	0,31	7,72	8,32
PIS (Projeto de Interesse Social)		PIS	861,46	0,28	8,26	8,77
RPQ1 (Residência Popular)		RP1Q	1.279,08	0,28	9,75	10,61
COMERCIAIS						
CAL- 8 (Comercial Andar Livres)	Normal	CAL 8-N	1.505,33	0,31	8,10	8,71
	Alto	CAL 8-A	1.658,73	0,33	8,52	9,41
CSL- 8 (Comercial Salas e Lojas)	Normal	CSL 8-N	1.260,24	0,26	8,07	8,46
	Alto	CSL 8-A	1.450,64	0,28	8,78	9,59
CSL- 16 (Comercial Salas e Lojas)	Normal	CSL 16-N	1.684,54	0,28	8,07	8,49
	Alto	CSL 16-A	1.933,70	0,30	8,75	9,58
GI (Galpão Industrial)		GI	672,45	0,06	8,53	8,88

Figura 3 Exemplo de tabela para consulta do CUB.

(SINDUSCON-MG)

Contudo, o CUB não engloba todos os serviços e materiais de uma construção, como, por exemplo, a execução de fundações, muros de arrimo, compra de equipamentos extras tais como elevador e bomba de recalque. Há também a remuneração da construtora e incorporadora que deve ser adicionada a posteriori.

Além disso, como o CUB se mede em reais por metro quadrado, teremos que enfrentar o problema de referência monetária: o dinheiro por si só não tem significado se não estiver atrelado a uma data, porque o poder de compra muda com o tempo. Por exemplo, R\$ 100,00 hoje não conseguem comprar a mesma quantidade de cimento que compravam há 10 anos ou daqui a 10 anos.

Para resolvermos essa dificuldade, devemos então estabelecer uma data de referência padrão, para a qual devemos transportar os dados obtidos considerando a inflação no período de acordo com algum índice representativo, como o Índice Geral de Preços de Mercado (IGP-M) ou o Índice Nacional da Construção Civil (INCC). A data que será referência monetária é arbitrária, de forma que, neste trabalho, será escolhido 01 de setembro de 2017.

Outra questão que surge é que o CUB varia com a localidade.

O CUB pode ser de várias fontes distintas, com sistemas de classificação de obras diferentes:

- CUB – SIDUSCON
- CUPE (Custos Unitários PINI de Edificações).
- CUB – SINAPI

Ainda devemos atentar se o CUB é desonerado ou não.

Lidaremos com essa problemática de referências mais a frente, na seção “Tratamento dos Dados”.

Além dessas considerações, a NBR 12721:2006 apresenta ajuda no cálculo dos custos de obras pelo método da "área equivalente de construção", um aperfeiçoamento em relação a simplesmente multiplicar o CUB pela área de construção, pois leva em consideração que as obras não são constituídas por partes com o mesmo padrão de acabamento, ou seja, as garagens são diferentes dos terraços, que são diferentes dos banheiros, das salas, dos halls de entrada, dos halls dos andares, das caixas de escada, etc. Assim, a norma determina a transformação das áreas de diferentes padrões de acabamento em áreas equivalentes de construção de uma dessas áreas, tomadas como padrão. Por exemplo, para uma obra que possui apartamentos, garagens e terraços sociais, etc, toma-se como padrão os apartamentos e transformam-se as demais em áreas equivalentes de construção do padrão dos apartamentos.

A área total, então, será a soma das áreas tomadas como padrão com as áreas calculadas como equivalentes à padrão. Calcula-se o custo total da mesma maneira, como sendo o produto da área total pelo CUB considerando-se o padrão escolhido.

2.1.2 Método da Unidade

O método da unidade consiste na escolha de uma unidade padronizada apropriada para o projeto em questão e multiplicando-a por um custo aproximado por unidade.

$$\text{Método da Unidade} = \text{Unidades Padrão} \times \text{Custo Unitário}$$

É o fundamento lógico do método do CUB, em que a unidade padrão é a área construída e o custo unitário é o CUB. Generaliza-se esse raciocínio.

Este método é usado para estimativas de projeto e de licitações. A estimativa de custo é obtida multiplicando o número de alojamentos para uma nova edificação com o custo por unidade do alojamento baseado em uma outra edificação apropriada. A atual unidade de alojamento pode ser obtida facilmente da planta preliminar ou de informação com o cliente.

O tipo de unidade de alojamento varia conforme a função da edificação. Por exemplo:

Escolas: Custo por Sala de Aula

Hospitais: Custo por leito

Rodovias: Custo por quilômetro

Estacionamentos: Custo por vaga

Com este método, pode-se gerar uma estimativa bruta relativamente rápida, mas a falta de precisão a faz não tão usada no planejamento preliminar. Entretanto, é frequente seu uso na primeira noção de custo nas primeiras discussões sobre o projeto. Só se pode usar esse raciocínio para comparar mesmo tipo de obras.

2.1.3 Método da Estimativa por Volume

Método de estimativa de taxa única (que toma como hipótese o comportamento linear das variáveis) baseado no conteúdo cúbico de uma construção. Custo por metro cúbico é

específico para projetos de construções e se propõe a superar o método do CUB, que não leva em conta possíveis variações na altura dos pavimentos.

O método se tornou muito popular em alguns países da Europa como Alemanha e Suíça, onde os custos de construção são frequentemente expressos por metro cúbico. O custo total do projeto seria dado por:

$$\text{Estimativa} = \text{Volume (m}^3\text{)} \times \text{Custo Unitário (R\$ / m}^3\text{)}$$

Para usar o método, o volume de construção deve ser primeiramente avaliado e existem regras explícitas em alguns países para esse fim. Construções com diferentes fins têm volumes correspondentes avaliados distintamente. Além disso, como no CUB, trabalhos específicos como escavação e fundações devem ser considerados separadamente.

Custos por metro cúbico podem ser difíceis de achar em países onde esse não é o principal método de estimativa conceitual. Tais custos dependem de variáveis como tipo da construção, proporção de área de parede por área de piso, qualidade do acabamento, dentre outras. Além disso, o cálculo do volume está sujeito a determinadas regras, como, por exemplo, a medida das paredes é feita de faces externas e a altura da construção é tomada do topo da fundação.

2.1.4 Estimativa do Envoltório dos Pavimentos

A estimativa por Envoltório dos Pavimentos é um método de taxa única, atualmente pouco utilizada, que mede a área de paredes externas, pisos e coberturas (efetivamente criando uma envoltória sobre a edificação) e as multiplica por coeficientes apropriados.

Leva em conta a diferença na forma plana total da área do piso, a posição vertical dos pisos, altura total, altura do pavimento, custos extras de áreas abaixo do solo.

Áreas de vários pisos, telhados e paredes externas são calculadas, e cada uma é multiplicada por um coeficiente e são somadas, resultando em um número de unidades de encapsulamento dos pavimentos que será utilizado para a estimativa total, multiplicando-o por uma taxa de encapsulamento obtida de análise prévia de projetos anteriores.

Quando este método foi sugerido, sua ideia era superar os problemas detectados nos métodos de taxa única, por levar em consideração variações no formato plano e na altura do pavimento. O método nunca foi totalmente adotado por profissionais da construção porque

requer muito mais cálculos que os outros métodos disponíveis e porque as taxas necessárias não podem ser extraídas de dados históricos.

2.1.5 Método Paramétrico

A estimativa paramétrica é um método de taxa múltipla e um dos mais usados em estimativas conceituais. Estimativas paramétricas têm fundamentos bastante gerais e são aplicados em vários ramos do conhecimento.

O que o método paramétrico faz é utilizar várias variáveis ao mesmo tempo, com contribuições diferentes, em sua predição, o que basicamente é o que este trabalho se propõe a fazer, mas com outra ferramenta matemática/computacional. Tendo em vista isso, seu conhecimento é importante, principalmente na escolha das variáveis do modelo de rede neural visado.

A obtenção de uma equação pelo método paramétrico é obtida através da regressão multivariável, uma ferramenta estatística, com base em um banco de dados extenso sobre do problema. Seu formato geral é $y = \sum_{k=0}^n b_k v_k + e$ em que b_k é o k-ésimo coeficiente do modelo e v_k a k-ésima variável e e é o erro da predição. Há abordagens que ainda se considera um expoente em cada variável.

No que diz respeito à construção civil, a literatura sugere a relação de alguns parâmetros individuais da construção com o custo dos componentes da construção (PHAOBUNJONG, 2002), conforme abaixo:

Parâmetros de Construção	Componentes Relacionados ao Custo da Construção Relacionados
1. Área Construída	a. Fundações b. Subestrutura
2. Área de pisos suspensos	a. Superestrutura do Piso
3. Área da Cobertura	a. Superestrutura da Cobertura b. Telhado
4. Área da Parede Externa	Revestimento Externo
5. Área, Portas Exteriores e Janelas	Portas e Janelas Externas
6. Área Total de Acabamento	Construção Interna

7. Área Bruta do Sítio	Trabalho do Sítio
8. Volume Total	Mecânica
9. Classificação do Transformador	Elétrica

Tabela 1 Relação de parâmetros construtivos com componente relacionado.

Parâmetros de Construção	Componentes Relacionados ao Custo da Construção Relacionados
1. Área Construída	a. Pisos b. Forros c. Partições d. Mecânica e. Elétrica
2. Perímetro Construído	a. Revestimento Externo b. Beirais e Parapeitos
3. Número de Ocupantes	a. Elevadores b. Encanamento

Tabela 2 Relação de parâmetros construtivos com componente relacionado.

Esses parâmetros são usados primariamente no desenvolvimento de estimativas baseadas no método unitário, isto é, o custo dos componentes é baseado nas quantidades propostas no projeto e custos unitários pré-estabelecidos. Por exemplo, o custo das fundações pode ser calculado baseado na área construída e custo por metro quadrado pré-estabelecido, como R\$ 9,00/m² de área construída. Determinando o custo de cada um desses elementos construtivos e sua respectiva quantidade, obtém-se o custo total estimado.

Alternativamente, extensas revisões da literatura descobriram três estudos de método paramétrico aplicados ao projeto de construção de edifícios.

No estudo de Kouskoulas (1974), o método de estimativa paramétrica foi desenvolvido para estimar o custo unitário da construção, em dólares por metro quadrado. A análise por regressão foi aplicada aos dados coletados de 38 edificações entre 1963 e 1972. Os parâmetros significantes identificados foram:

- a. Região nos Estados Unidos onde a construção estava localizada
- b. Ano de construção

- c. Uso da construção
- d. Número de pavimentos
- e. Qualidade do acabamento e materiais especificados

O segundo estudo identificado (KARSHENAS, 1984) usou regressão na confecção da equação paramétrica do custo de uma construção para estimar o custo total de uma edificação. O estudo focou no custo de construção de edifícios de escritórios multi-pavimentos. Os dados usados no estudo consistiram de 24 edificações construídas entre 1961 e 1979. Só dois parâmetros construtivos foram usados no estudo:

- a. Área do Piso Típica
- b. Altura da edificação

O terceiro estudo (BOWLBY, 1986) usou regressão para desenvolver um modelo de estimativa paramétrica para a previsão do custo de construção em dólares por metro quadrado. Os autores tiveram acesso ao banco de dados de contratos da consultora norte-americana de construção civil F.W. Dodge, e conseqüentemente um total de 157855 projetos executados entre 1972 e 1982 foram utilizados na regressão. O estudo identificou os seguintes parâmetros como significantes:

- a. Número de pavimentos
- b. Área total da Edificação
- c. Localização (Uma variável dummy que identifica se a área é metropolitana ou rural)
- d. Se uma parte da obra aconteceu no inverno
- e. Uso final do edifício
- f. Modelo Construtivo
- g. Região dos EUA onde a edificação está localizada

Como pode ser visto nos modelos paramétricos acima, numerosos parâmetros são sugeridos pelos estudos. Entretanto, é evidente que o método unitário necessita de um banco de dados mais detalhado (nível micro) que o método paramétrico. Os dados usados no método paramétrico tendem a ser mais gerais (nível macro).

Vários parâmetros foram identificados por muitos estudos e pesquisas em estimativas conceituais de custos baseadas no método unitário e no modelo paramétrico. A natureza dos

parâmetros identificados é distinta em função das diferenças fundamentais entre os dois métodos, qual seja, o método da unidade é uma abordagem mais simples na qual o custo unitário (parâmetros e relação entre os custos) são pré-estabelecidas de forma geral. Em contraste, os métodos paramétricos não fazem uso, a priori, de quaisquer relações pré-determinadas ou regras de bolso. Pelo método paramétrico, os parâmetros significativos, os quais são conhecidos como drivers de custo são identificados, e a relação entre os parâmetros e o custo é estabelecida pela base de dados histórica. Desta maneira o método paramétrico é mais um processo intensivo de dados e demanda uma quantidade grande de dados de projetos para garantir validade estatística. Para desenvolver um banco de dados de projetos com um número suficiente de projetos e com consistência de dados, o nível de detalhe deve ser sacrificado. Isso explica por que os parâmetros usados no método paramétrico tendem a não serem tão detalhados.

A primeira questão identificada relacionada ao método paramétrico é, portanto, os dados históricos, ou a falta deles. As estimativas paramétricas são pouco utilizadas porque não existem bancos de dados paramétricos históricos sistematizados. Em vez disso, as estimativas de custo derivadas pelo método da unidade são mais utilizadas (PARK, 1999). Essa falta de dados é uma das razões que contribuem para a pesquisa limitada e a aplicação do método paramétrico na estimativa de custo conceitual do projeto de construção de edifícios.

A segunda questão identificada decorre de pesquisa limitada realizada no método paramétrico. A pesquisa limitada levou a um conhecimento muito pequeno sobre a implementação do método paramétrico e a identificação limitada dos parâmetros que podem ser usados com sucesso no desenvolvimento de um modelo de estimativa de custo paramétrico. A terceira questão identificada é que o método paramétrico tem o potencial de ser a metodologia ideal para a estimativa conceitual de custos de projetos de construção de edifícios. Embora seja necessária uma quantidade suficiente de dados para o seu desenvolvimento, o modelo de estimativa de custo paramétrico, uma vez desenvolvido, fornece os seguintes recursos desejáveis:

- a. Baseado numa base histórica significativamente grande, isso reflete as experiências e o grau de especialização na construção dos projetos envolvidos;
- b. Permite estimativa conceitual de custos com alguns parâmetros de projeto de alto nível, facilitando o processo de estimativa e levando a um método eficiente e rápido de produção da estimativa; e
- c. Uma abordagem sistemática e quantitativa minimiza a subjetividade no processo de estimativa e garante a consistência nas estimativas geradas.

2.2 Conceitos Básicos de Redes Neurais Artificiais

Uma das máquinas mais poderosa que a humanidade conhece é o próprio cérebro humano. Desde o século passado muito esforço vem sendo feito para conseguirmos simular a mente humana em um computador, e para isso foram desenvolvidas as Redes Neurais Artificiais, hoje bastante popularizadas devido a aplicações como reconhecimento de escrita, de rostos e de vozes, na navegação autônoma, na identificação de padrões em tecidos humanos e previsões no mercado financeiro.

Esta teoria é inspirada no comportamento do cérebro dos animais. Toda rede neural, seja ela biológica ou computacional, tem como elemento fundamental o neurônio.

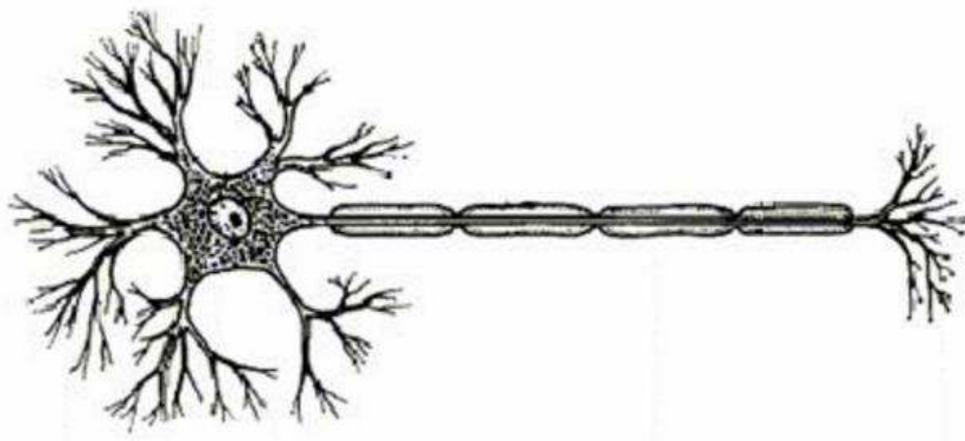


Figura 4 Representação de um Neurônio Biológico

Os neurônios biológicos são células nervosas que desempenham o papel de conduzir os impulsos nervosos. Eles possuem três partes principais: os dendritos, onde ocorre a recepção das informações e é a parte receptora do neurônio, o corpo celular, responsável pela integração das informações, e o axônio, onde ocorre o transporte do impulso nervoso de um neurônio para outra célula.

O neurônio computacional tem sua forma geral bem simples e parecida com o neurônio biológico: os sinais de entrada (inputs) recebidos são integrados resultando em um sinal de saída (output). Essa integração ocorre pela ponderação dos sinais, por um peso, e aplicados a uma função de ativação, que é a ferramenta responsável pela finalização da integração dessas entradas já ponderadas, para que os resultados do modelo matemático sejam

próximos ao biológico. Ainda é possível acrescentar uma constante (chamada de bias) na soma ponderada das entradas, antes de aplicá-la à função de ativação.

$$\text{output}_i = f(\text{net}_i) = f\left(\sum_1 w_i x_i + \text{bias}_i\right)$$

Equação 1 Fórmula do output de um neurônio artificial

Nesta fórmula, para o i-ésimo input, w_i é o peso, x_i é o valor do input, bias_i a entrada constante para melhorar a convergência, e f é a função de ativação, que pode ser qualquer função, sendo as mais adotadas listadas abaixo.

ReLU (Rectified linear unit)	$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$	
Tangente Hiperbólica	$f(x) = \tanh(x)$	
Logística	$f(x) = \frac{1}{1 + e^{-x}}$	

Tabela 3 Funções de ativação mais comuns

Uma função também muito utilizada é chamada softmax, entretanto ela não é aconselhada para problemas de regressão como o tratado neste trabalho.

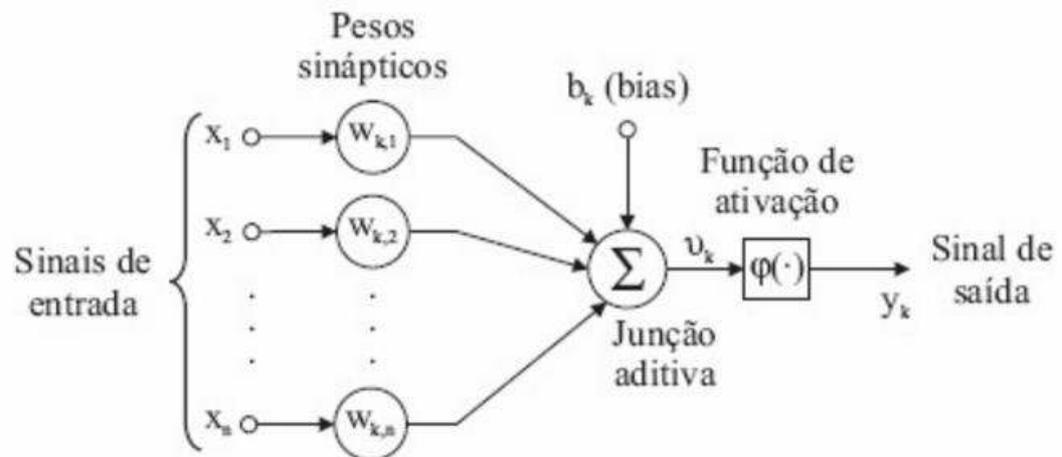


Figura 5 Neurônio Computacional

Um conjunto de neurônios ligados entre si constitui uma rede neural (artificial ou biológica), conforme ilustra a Figura 6. Seus elementos são:

- Valores de Entrada: são as entradas numéricas.
- Nós: as unidades de processamentos, ou seja, os neurônios artificiais.
- Pesos Sinápticos: consistem nos valores que calibram cada ligação entre nós para melhorar as previsões.
- Valores de Saídas: os resultados que o último conjunto de neurônios retorna.

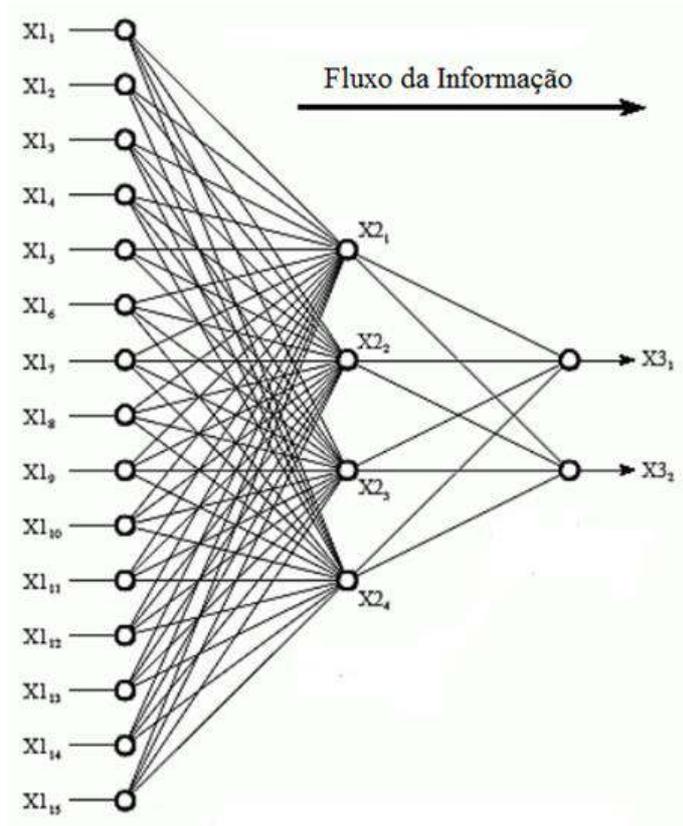


Figura 6 Exemplo de Rede Neural Artificial

Na Figura 6 temos a topologia de rede neural mais comum: três camadas completamente interconectadas. Os nós da camada dos valores de entradas são passivos, isto é, não fazem nada além de repassar os valores que recebem para os nós da camada imediatamente posterior. Eles podem ser entendidos como neurônios com função de ativação identidade. As outras camadas são ativas, modificando os sinais recebidos de acordo com sua função de ativação.

Elegidas variáveis do problema, ou seja, valores de entrada e os valores alvo que a rede tentará alcançar, estas serão passadas para a rede neural, com uma topologia apropriada. Os pesos serão ajustados para que o erro nas previsões seja o menor possível. Este processo de minimização do erro é chamado de treinamento.

Elementos mais específicos da solução proposta neste trabalho, como qual topologia escolher, quantos neurônios por camada utilizar e qual a métrica de erro mais apropriada, serão discutidos na etapa de desenvolvimento da rede.

3 Coleta de Dados

Serão utilizados dois bancos de dados, conforme suas informações disponíveis.

O primeiro banco de dados (BD1) deste trabalho é composto de diversas fontes, mas em sua essência do DT-INFRA-MANAUS – DESTACAMENTO DE INFRAESTRUTURA DA AERONÁUTICA DE MANAUS, COMARA – COMISSÃO DE AEROPORTOS DA REGIÃO AMAZÔNICA, CLA - CENTRO DE LANÇAMENTO DE ALCÂNTARA, CEPE - CENTRO DE ESTUDOS E PROJETOS DE ENGENHARIA e CISCEA – COMISSÃO DE IMPLANTAÇÃO DO SISTEMA DE CONTROLE DO ESPAÇO AÉREO.

Esse compilado de dados é composto de 24 projetos, datados de abril de 2012 a agosto de 2017, e organizados por pastas com projeto básico, contratos, plantas, orçamentos, cronogramas, e outros arquivos auxiliares como os Relatórios Mensais de Obras.

O segundo banco de dados (BD2) deste trabalho foi obtido com o TCU – Tribunal de Contas da União, em um de seus estudos (TRIBUNAL DE CONTAS DA UNIÃO, 2013), por meio de solicitação do banco de dados utilizado, com base na Lei do Acesso à Informação (Lei Nº 12.527 de 18 de novembro de 2011).

Nem todos os itens puderam ser aproveitados, de forma que somente 46 obras se encaixavam nos requisitos deste trabalho. Estas datam de junho de 2011 a dezembro de 2007. Algumas informações tiveram que ser colhidas individualmente de cada um dos contratos, também fornecidos pelo TCU.

3.1 Escolha das Variáveis Elegíveis

Das estimativas conceituais apresentadas anteriormente, abaixo segue um resumo das principais variáveis, por cada método, que impactam no custo total:

Método da Área:

Área construída

Padrão de acabamento

Tipo da Construção (residencial ou comercial, e seus casos particulares)

Localidade

Data da Construção

Método da Unidade:

Tipo de Construção (Escola, Hospital, Creche, etc.)

Quantidade de unidades representativas

Método Volumétrico:

Área do Piso

Altura da Edificação

Tipo da Construção

Proporção Área de Parede por Área de Piso

Padrão de acabamento

Envoltório dos Pavimentos:

Área de Paredes Externas

Área de Pisos

Área de Coberturas

Método Paramétrico:

Várias, a depender do estudo utilizado

Desta forma, observando o exposto acima, e tendo em vista apenas as informações disponíveis pelas fontes dos dados, optou-se por filtrar os seguintes campos no primeiro banco de dados:

- Valor do contrato
- Data de referência do contrato
- Prazo de execução (dias)
- Localidade (Estado)
- Pé direiro (metros)
- Número de pavimentos
- Área Edificada (m²)
- Perímetro Edificado (m)

No segundo banco de dados, coletou-se:

- Valor do contrato
- Data de referência do contrato
- Prazo de execução (dias)
- Área Edificada (m²)
- Localidade (Estado)
- Se a obra foi feita em área urbana ou em interior

3.2 Tratamento dos Dados

De posse dessas informações, a etapa seguinte é estruturá-las de forma a ser corretamente interpretada pela rede, incorporando ao banco de dados apenas as obras com todas as informações que necessitamos.

As variáveis que envolvem valores monetários devem ser trazidas a um mesmo referencial temporal, padronizado como setembro de 2017, utilizando um índice de inflação, o INCC – Índice Nacional da Construção Civil.

Os resultados de BD1 e BD2 se encontram no Apêndice A. Esses dois bancos de dados foram unidos utilizando-se a intersecção dos dados disponíveis, o que gerou o BD3, que está no Apêndice B.

4 Desenvolvimento

Nesta seção será discutida a escolha dos principais parâmetros que caracterizam a rede neural, desde a arquitetura da rede, o pré-processamento de variáveis categóricas, até o processo de otimização do treinamento escolhido.

4.1 Criação da Rede Neural

Elegidas as variáveis do problema, as quais já tratadas serão passadas como entrada para a Rede Neural, com uma arquitetura inicial apropriada. Tal rede será treinada em um processo de aprendizagem supervisionada¹, isto é, os pesos serão ajustados de tal forma que para um dado conjunto de valores de entrada cheguemos próximos o suficiente dos valores de saída. O erro a ser minimizado recomendado para regressões para averiguar quão bom está tal configuração de peso será o Erro Médio Quadrático.

Toda a simulação da rede neural será feita utilizando a linguagem de programação Python 3.5 com auxílio das bibliotecas Scikit-Learn, Keras e TensorFlow. Os dados serão incluídos por meio da biblioteca Pandas.

O banco de dados com os inputs para cada projeto ficará armazenado sumariamente em uma planilha.

A questão, portanto, é selecionarmos apropriadamente nossos dados de entrada, valores alvo, a arquitetura e parâmetros da rede neural.

4.1.1 Pré-Processamento dos Dados

Alguns valores devem ser pré-processados, de forma a aumentar a eficiência computacional da rede e não sobrepor à importância de uma variável sobre outra.

¹ A aprendizagem de uma rede neural é dividida em quatro tipos:

- Supervisionada: quando temos os valores alvo (os valores que queremos prever). A rede tenta aprender uma função que prediz corretamente esses valores com base em um banco de dados. Exemplos típicos: classificação e regressão.
- Não supervisionada: quando não se tem os valores alvo para serem preditos. Em vez disso, a rede aprende um modelo que pode gerar o banco de dados. Exemplos típicos: agrupamento (clustering) e redução de ruído.
- Reforçado: é uma configuração em que temos um problema de decisão sequencial. Tomar uma decisão agora influencia as decisões que podemos tomar no futuro. Uma função de recompensa é fornecida que diz à rede neural quão bom são certos estados. Muito utilizado na criação de bots para jogos.
- Ativo: é quando o algoritmo pode pedir um especialista para o rótulo de destino correto de um exemplo. Isso permite que o algoritmo se concentre em áreas onde o modelo é o menos certo.

Primeiramente transformaremos variáveis categóricas (tipo de edificação e localização) em um conjunto de variáveis do tipo dummy (0 ou 1). Isso funciona da seguinte forma: se uma variável pode assumir apenas quatro valores (norte, sul, leste e oeste, por exemplo), deverão ser criadas quatro variáveis (“é Norte?”, “é Sul?”, “é Leste?” e “é Oeste?”) que assumirão valores de 0 ou 1 (verdadeiro ou falso). E, atentando-se para a dummy variable trap (Cope), ou seja, uma das incógnitas é redundante, exclui-se uma delas, restando três. De modo geral, se há n categorias, trabalharemos com $n-1$ variáveis tipo dummy.

Em seguida dividiremos o banco de dados em um conjunto para treinamento e outro para teste. A proporção inicial escolhida é arbitrada como 5 para 1.

Além disso, o ajuste de escala nas variáveis ainda deve ser feito, isto é, temos que dar o mesmo grau de importância a uma variável que varia de 0 a 1 e uma que o menor valor é 1.000.000,00 e o maior 10.000.000,00. Isso é importante por que uma variável não deve dominar numericamente sobre a outra, além de melhorar o processo de convergência no treinamento da rede por diminuir o esforço computacional. Há diversas formas de fazer isso (RASCHKA, 2014), e cada caminho resulta em performances próprias, de forma que se optou por escalar as variáveis de 0 a 1, após os testes convergirem mais rapidamente.

4.1.2 Determinação da Arquitetura da Rede Neural

A escolha da arquitetura afeta substancialmente o sucesso da rede neural e não há uma regra fixa para a escolha dos melhores parâmetros para a rede, o qual se dá por experimentação.

Entretanto, algumas características mínimas podem ser estabelecidas como se a rede será Feedforward (comum ou estrita) ou Feedback. Essa classificação diz se as ligações só ocorrerão em camadas sucessivas ou com alguma camada anterior.

Para o problema de regressão, é indicada a topologia Feedforward do tipo Multilayer Perceptron, ou Perceptron Multicamadas (SARLE, 1997).

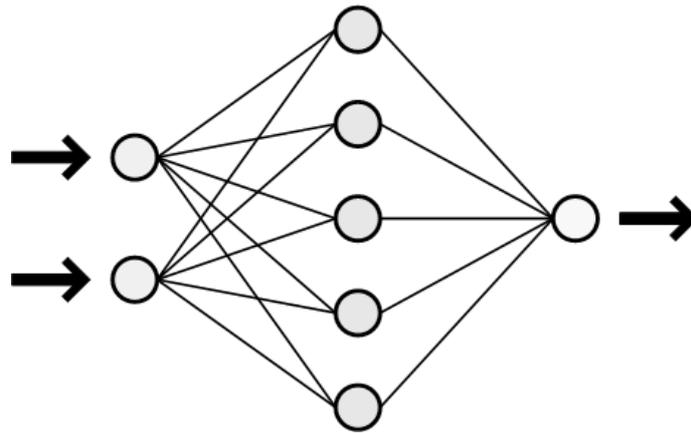


Figura 7 Representação de Rede Neural Feedforward

A primeira camada terá a quantidade de nós igual à quantidade de variáveis do banco de dados, e a última camada, por se tratar de um problema de regressão, será composta de apenas um neurônio por estarmos fazendo a previsão de um valor específico que é o preço da obra (WARREN S. SARLE, 2002).

A quantidade de camadas escondidas e o número de neurônios, assim como suas respectivas funções de ativação, serão determinados por testes manuais, entretanto haverá a limitação de três camadas, pois o backprop, algoritmo de treinamento da rede que será discutido mais adiante, funciona com qualquer número de camadas intermediárias (aquelas que não são nem entrada nem saída), mas consegue-se abstrair, praticamente, qualquer tipo de conhecimento com no máximo duas camadas intermediárias. Problemas mais simples são facilmente resolvidos com duas camadas. Os mais complexos exigem três camadas. O tempo final de treinamento também cresce com o aumento do número de camadas.

4.1.3 Treinamento da rede

No treinamento das redes, será utilizado o algoritmo chamado backpropagation, ou backprop, que funciona calculando a contribuição no erro de cada neurônio depois do processamento de um lote de dados (batch of data). O tamanho do lote de dados para nossa amostra que trouxe melhor resultado foi 4.

O backprop calcula o gradiente da função de perda, e é comumente usado no algoritmo de otimização gradiente descendente, que procura o mínimo local indo na direção contrária ao gradiente (daí o nome gradient descent). Também é chamado de backward

propagation of erros, porque o erro é calculado no output e distribuído para os neurônios pelas camadas da rede. É considerado um método de treinamento supervisionado e normalmente é tratado com o uso de um otimizador, sendo que o mais utilizado é o Adam, ou Adaptive Moment Estimation (KINGMA, 2014).

Uma taxa de aprendizagem (η) também tem que ser informada ao modelo, sendo que após várias experimentações, o valor que propiciou um maior número casos de convergência foi de 0,01. Este parâmetro é análogo ao “passo” no método do passo, e é importante na convergência. Se η for muito grande, o ponto de minimização do erro pode passar despercebido pelo gradient descent. Se η for muito pequeno, o treinamento irá demorar demais.

Existem várias funções erro possíveis (MATA, 2017) específicas para cada tipo de problema, de forma que a mais apropriada no caso de regressões com amostras com características não exponenciais é o custo quadrático.

$$CMST(W,B,S_r,E_r)=0.5\sum_j(a_{Lj}-E_{rj})^2$$

Equação 2 Custo Quadrático

$$\nabla_a CMST=(a_L-E_r)$$

Equação 3 Gradiente do Custo Quadrático, a ser utilizado no backprop

Os pesos serão inicializados de forma aleatória segundo uma seed = 7, isto é, o elemento gerador da aleatoriedade nos valores do peso. Isso é importante para a reprodutibilidade dos resultados alcançados neste trabalho, ou seja, se for feita a mesma análise por outra pessoa em outro computador, mas utilizando o mesmo seed, em teoria chega-se aos mesmos valores.

Todo o código se encontra no Apêndice C.

4.1.4 Métricas para avaliação dos resultados

Os resultados obtidos da previsão de cada modelo devem ser comparados, e as métricas disponíveis mais adequadas (SCIKIT LEARN, 2017) para a regressão de redes neurais artificiais são mostradas abaixo:

$$\text{explained_variance}(y, \hat{y}) = 1 - \frac{\text{Var}\{y - \hat{y}\}}{\text{Var}\{y\}}$$

Equação 4 Variância Explicada

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{samples}}-1} (y_i - \bar{y})^2}$$

Equação 5 Coeficiente de Determinação, ou R^2

$$\bar{R}^2 = 1 - \frac{n-1}{n-(k+1)}(1-R^2)$$

Equação 6 Coeficiente de Determinação Ajustado

$$\text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i|$$

Equação 7 Erro Absoluto Médio

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2$$

Equação 8 Erro Quadrático Médio

$$\text{MedAE}(y, \hat{y}) = \text{median}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$$

Equação 9 Mediana do Erro Absoluto

Usaremos essencialmente R^2 , Variância Explicada, MSE, MAE e MedAE.

5 Análise dos Resultados

Utilizando todo o substrato dos itens acima, procedeu-se à análise, tendo em vista que os resultados serão analisados sob a óptica das métricas discutidas em 4.1.4.

5.1 Resultados e Interpretação

De todas as variáveis consideradas, sabidamente a que tem mais influência no preço de uma obra é a área. Sendo assim, primeiro fizemos uma análise isolada da área da edificação versus o valor orçamentado.

Os parâmetros alterados inicialmente foram: neurônios da primeira camada, neurônios da segunda camada, neurônios da terceira camada, respectivas funções de ativação e número de épocas.

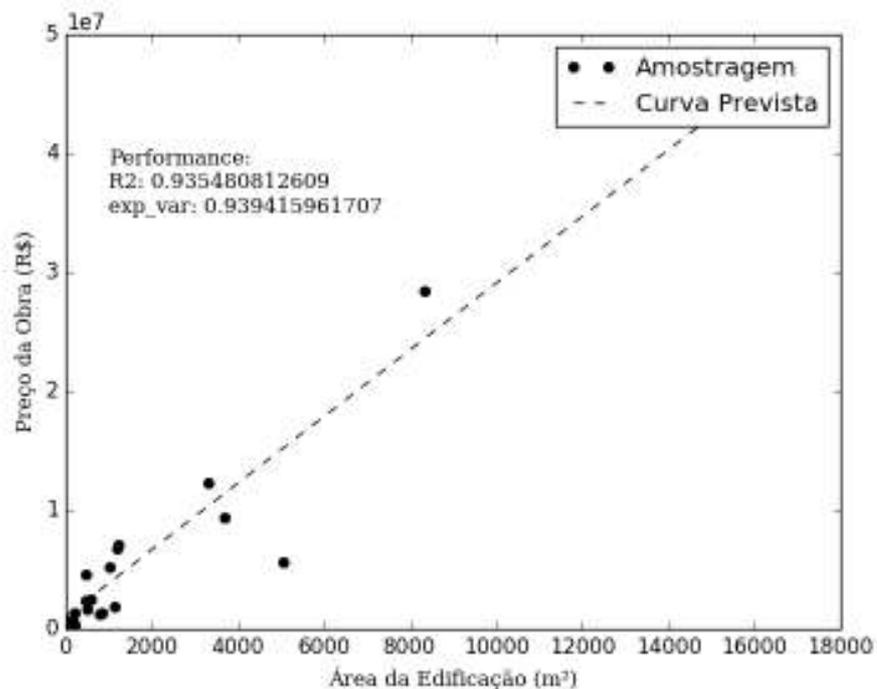


Figura 8 Resultado da Configuração 01

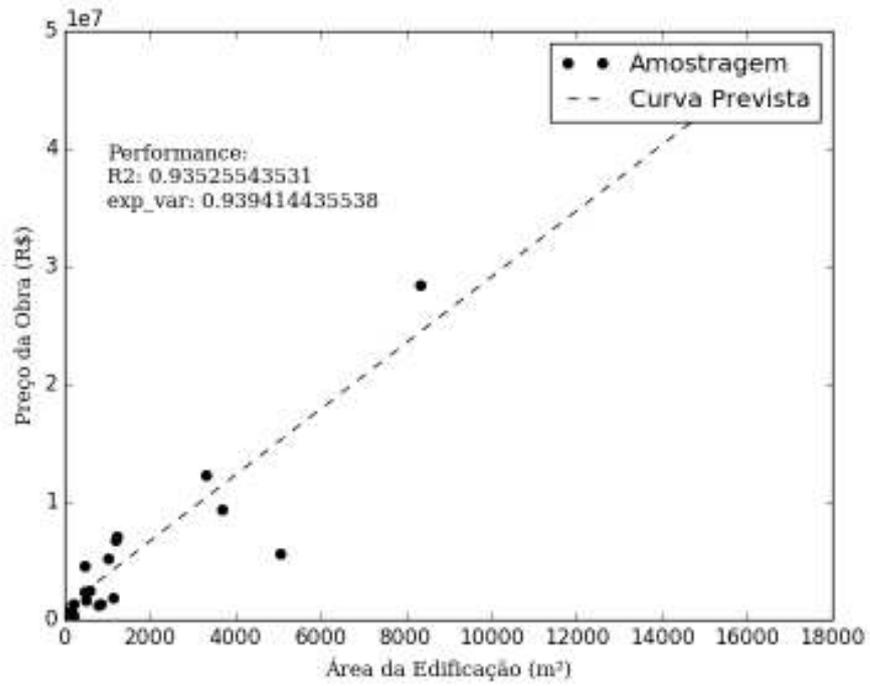


Figura 9 Resultado da Configuração 02

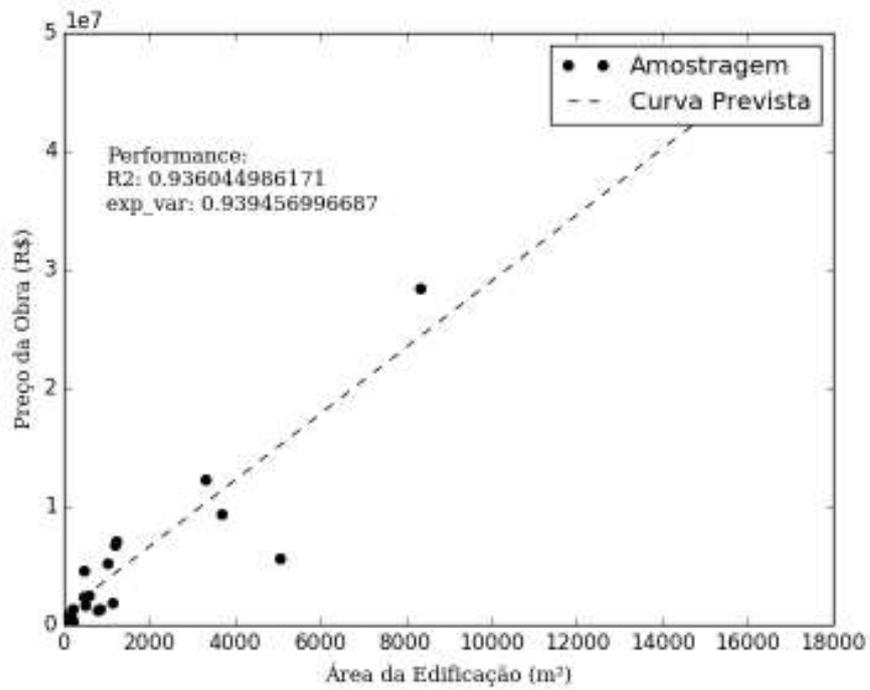


Figura 10 Resultado da Configuração 03

Sumariamente, temos:

	Configuração 1	Configuração 2	Configuração 3	Configuração 4	Configuração 5
Neurônios 1ª camada	1	2	3	2	2
Neurônios 2ª camada	0	0	0	1	2
Neurônios 3ª camada	1	1	1	1	1
Função de Ativação 1ª camada	relu	relu	relu	relu	relu
Função de Ativação 2ª camada	relu	relu	relu	relu	relu
Função de Ativação 3ª camada	relu	relu	relu	relu	relu
Número de Épocas	1000	2000	3000	2000	2000
R²	0.9355	0.9353	0.9360	-0.3224	-0.3224
Variância Explicada	0.9394	0.9394	0.9395	0.0000	0.0000
Erro Absoluto Médio	0.0404	0.0406	0.0399	0.1272	0.1272
Erro Quadrático Médio	0.0032	0.0033	0.0032	0.0664	0.0664
Mediana do Erro Absoluto	0.0285	0.0285	0.0285	0.0383	0.0383

Tabela 4 Resultados dos primeiros testes.

Percebe-se que o gráfico não se alterou significativamente para quando consideramos mais neurônios na primeira camada, como mostra a tendência nos testes 01, 02 e 03, entretanto piorou com a adição de neurônios na camada secundária.

Estes resultados foram ao encontro do obtido pelo Excel em uma regressão linear:

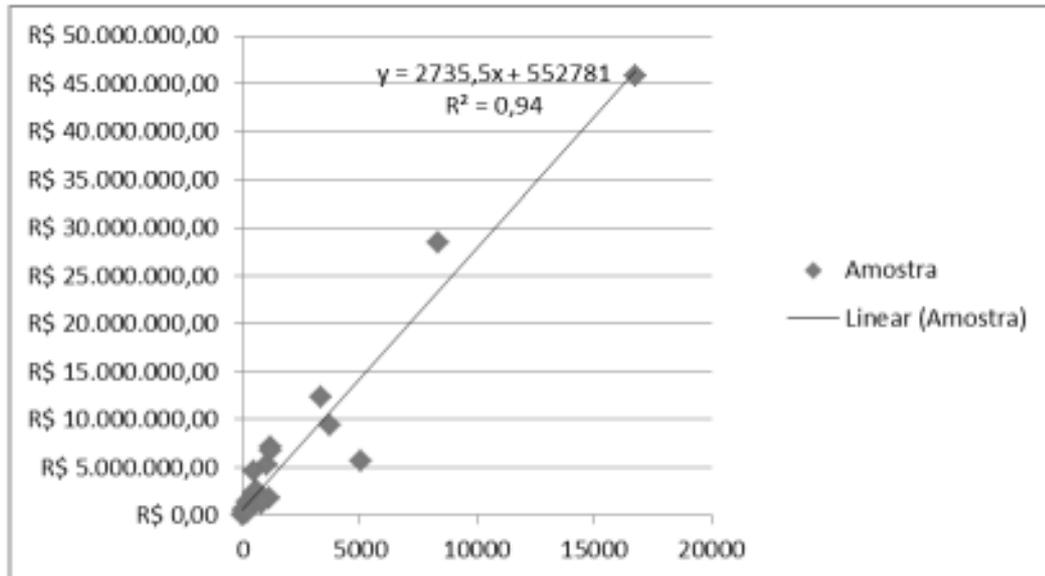


Figura 11 Regressão Linear (Excel)

Agora vamos incorporar ao modelo o perímetro. Por ser um gráfico tridimensional, a visualização das simulações pode ser melhor entendida pela tabela abaixo, ao invés de um gráfico.

Configuração	1	2	3	4	5	6	7
Neurônios 1ª camada	4	4	4	5	3	3	4
Neurônios 2ª camada	0	0	0	0	0	1	0
Neurônios 3ª camada	1	1	1	1	1	1	1
Função de Ativação 1ª camada	relu	relu	relu	relu	relu	relu	relu
Função de Ativação 2ª camada	relu	relu	relu	relu	relu	relu	relu
Função de Ativação 3ª camada	relu	relu	relu	relu	relu	relu	tanh
Número de Épocas	1000	2000	3000	2000	2000	2000	2000
R ²	0.9644	0.9618	0.9621	-0.3224	-0.3224	0.9622	0.9017
Variância Explicada	0.9664	0.9644	0.9644	0.0000	0.0000	0.9649	0.9058
Erro Absoluto Médio	0.0314	0.0322	0.0319	0.1272	0.1272	0.0320	0.0387
Erro Quadrático Médio	0.0018	0.0019	0.0019	0.0664	0.0664	0.0019	0.0049
Mediana do Erro Absoluto	0.0252	0.0259	0.0252	0.0383	0.0383	0.0263	0.0132

Tabela 5 Segunda bateria de testes: com área e perímetro como entradas

Algumas conclusões podem ser listadas a partir desses ensaios:

- O aumento no número de épocas nas configurações 1 a 3 não impactou tanto na qualidade da previsão. Já nas configurações 4 e 5 não houve convergência.
- Com uma variável a mais, e mais épocas, é possível uma previsão melhor, vide configuração 1, que resultou em um $R^2=0,9644$, mais que uma simples regressão linear da Figura 11;
- A escolha da função de ativação tangente hiperbólica (tanh) na terceira camada piorou, como esperado, os resultados (configuração 7 versus configuração 2);
- A inclusão de uma segunda camada ajudou na convergência, se compararmos as configurações 5 com a 6;
- A configuração 7 evidencia que a utilização de uma função de ativação diferente da relu na camada de saída resultou em uma previsão precária.
- Para mais de 4 neurônios na primeira camada, a rede não convergiu.

Vamos então incluir todas as variáveis disponíveis em BD1, isto é, área, perímetro, localidade, pé-direito e se o sítio é em área urbana ou de interior.

Configuração	1	2	3	4	5	6	7	8	9	10
Neurônios 1ª camada	4	4	4	4	6	6	8	8	4	4
Neurônios 2ª camada	0	0	0	2	2	4	2	4	2	2
Neurônios 3ª camada	1	1	1	1	1	1	1	1	1	1
Função de Ativação 1ª camada	relu	relu	relu	relu	relu	relu	relu	relu	relu	tanh
Função de Ativação 2ª camada	relu	relu	relu	relu	relu	relu	relu	relu	tanh	relu
Função de Ativação 3ª camada	relu	relu	relu	relu	relu	relu	relu	relu	relu	relu
Número de Épocas	1000	2000	5000	2000	2000	2000	2000	2000	2000	2000
R^2	0.9605	0.9496	0.7782	0.9628	0.9550	-0.3224	0.9539	0.9138	0.9392	0.9332
Variância Explicada	0.9642	0.9530	0.7926	0.9649	0.9588	0.0000	0.9562	0.9173	0.9426	0.9358
Erro Absoluto Médio	0.0216	0.0212	0.0309	0.0229	0.0202	0.1272	0.0229	0.0297	0.0235	0.0239
Erro Quadrático Médio	0.0020	0.0025	0.0111	0.0019	0.0023	0.0664	0.0023	0.0043	0.0031	0.0034
Mediana do Erro Absoluto	0.0091	0.0066	0.0075	0.0139	0.0070	0.0383	0.0139	0.0189	0.0089	0.0116

Tabela 6 Terceira bateria de testes: com todas as variáveis disponíveis

A única disposição que não convergiu foi a 6. Já o segundo pior resultado (configuração 3) mostra algo contra intuitivo: com um aumento no número de épocas, obteve-se um R^2 pior. Uma possível explicação para este fato é que, conforme já discutido em 4.1.3, durante a busca pode-se ter caído em um mínimo local diferente do global, ou ainda a taxa de aprendizado não estava do tamanho ideal.

Agora cabe observar que, conforme discutido em 3.2, os pontos são poucos e mal distribuídos, de forma que não necessariamente o R^2 mais alto é devido a uma previsão melhor. Cabe uma interpretação qualitativa do problema e dos resultados.

Ainda, adicionar variáveis indiscriminadamente pode acarretar resultados aparentemente melhores mas que na prática são piores, devido ao fenômeno chamado overfitting, que acontece quando temos muitas graus de liberdade na regressão em relação à quantidade de pontos da amostra.

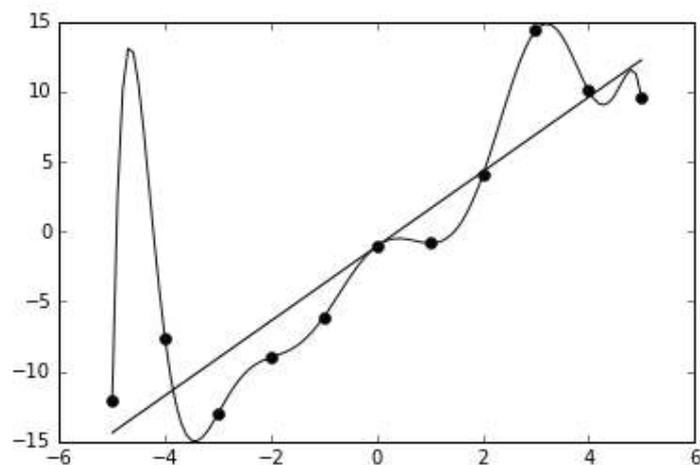


Figura 12 Ilustração do overfitting

Na figura acima, os dados com ruído aproximadamente lineares são ajustados a uma função linear e a uma função polinomial. Embora a função polinomial seja um ajuste perfeito, a função linear pode ser generalizada melhor: se as duas funções fossem usadas para extrapolar além dos dados de ajuste, a função linear faria melhores previsões.

Tendo isso em vista e utilizando a Navalha de Occam, enunciada como “Se em tudo o mais forem idênticas as várias explicações de um fenômeno, a mais simples é a melhor”, analisemos como o preço previsto varia apenas com a área, utilizando agora o segundo banco de dados BD2.

Configuração	1	2	3	4	5	6
Neurônios 1ª camada	10	8	4	8	4	1
Neurônios 2ª camada	0	0	0	2	2	0
Neurônios 3ª camada	1	1	1	1	1	1
Função de Ativação 1ª camada	relu	relu	relu	relu	relu	relu
Função de Ativação 2ª camada	relu	relu	relu	relu	relu	relu
Função de Ativação 3ª camada	relu	relu	relu	relu	relu	relu
Número de Épocas	1000	1000	1000	1000	1000	1000
R ²	0.8159	0.8159	0.8138	-0.6969	0.8153	0.8155
Variância Explicada	0.8160	0.8160	0.8143	0.0000	0.8153	0.8155
Erro Absoluto Médio	0.0736	0.0736	0.0787	0.2305	0.0747	0.0744
Erro Quadrático Médio	0.0140	0.0140	0.0142	0.1294	0.0141	0.0141
Mediana do Erro Absoluto	0.0252	0.0252	0.0295	0.0885	0.0293	0.0266

Tabela 7 Quarta rodada de testes: influência da área, em BD2

Via de regra são resultados bons, mas observa-se que a configuração 4 não convergiu, mesmo variando-se parâmetros como a função do erro e número de épocas.

Utilizando todas as variáveis disponíveis em BD2 (área, estado, urbana/interior), obteve-se:

Configuração	1	2	3	4	5	6	7
Neurônios 1ª camada	10	8	6	4	2	1	4
Neurônios 2ª camada	0	0	0	0	0	0	2
Neurônios 3ª camada	1	1	1	1	1	1	1
Função de Ativação 1ª camada	relu						
Função de Ativação 2ª camada	relu						
Função de Ativação 3ª camada	relu						
Número de Épocas	2000	2000	2000	2000	2000	2000	2000
R ²	0.8249	0.5991	0.6936	0.8021	0.7664	0.9267	-0.6969
Variância Explicada	0.8249	0.6067	0.6941	0.8021	0.7667	0.9271	0.0000
Erro Absoluto Médio	0.0479	0.0526	0.0566	0.0505	0.0532	0.0412	0.2305
Erro Quadrático Médio	0.0134	0.0306	0.0234	0.0151	0.0178	0.0056	0.1294
Mediana do Erro Absoluto	0.0106	0.0053	0.0117	0.0122	0.0167	0.0208	0.0885

Tabela 8 Quinta rodada de testes: todas variáveis de BD2

O melhor resultado da quinta rodada de testes com um R^2 de 0,9271 foi o da configuração 6, embora, como veremos mais adiante, neste caso particular não podemos esperar previsões tão boas como a dos outros modelos. A grande variação nos R^2 das outras configurações que convergiram (1, 2, 3, 4 e 5) desse mesmo teste já deveria nos alertar para um possível overfitting. De fato, a variável de localidade se transforma em muitas variáveis dummy, já que há uma quantidade muito grande de estados neste banco de dados.

Por fim, com a mesma abordagem de analisar todas as variáveis disponíveis, atentemos agora ao banco de dados BD3. O tempo de execução não é considerado como input, mas sim como output, e foi coletado apenas para originalmente extrair outliers.

Configuração	1	2	3	4	5	6
Neurônios 1ª camada	10	8	6	4	2	1
Neurônios 2ª camada	0	0	0	0	0	0
Neurônios 3ª camada	1	1	1	1	1	1
Função de Ativação 1ª camada	relu	relu	relu	relu	relu	relu
Função de Ativação 2ª camada	relu	relu	relu	relu	relu	relu
Função de Ativação 3ª camada	relu	relu	relu	relu	relu	relu
Número de Épocas	2000	2000	2000	2000	2000	2000
R^2	0.8849	0.9169	0.8736	0.8809	0.8768	0.8766
Variância Explicada	0.8867	0.9235	0.8756	0.8827	0.8782	0.8790
Erro Absoluto Médio	0.0376	0.0365	0.0430	0.0436	0.0463	0.0455
Erro Quadrático Médio	0.0090	0.0065	0.0099	0.0093	0.0096	0.0097
Mediana do Erro Absoluto	0.0176	0.0133	0.0155	0.0206	0.0208	0.0194

Tabela 9 Última rodada de testes: todas variáveis de BD3

5.2 Análise em caso particular

Para ilustrarmos a capacidades dos modelos faremos uma análise do caso particular do Alojamento de Atletas – UNIFA, localizado no Rio de Janeiro.

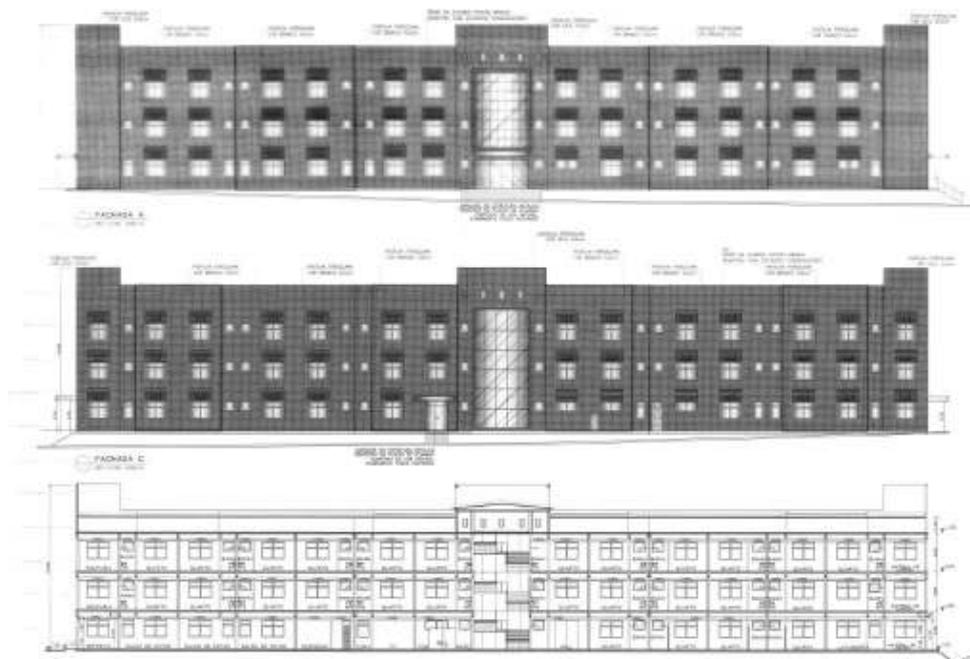


Figura 13 Alojamento de Atletas - UNIFA

A edificação tem as seguintes características: área edificada de 3312,70 m², pé direito de 3,5 m, perímetro edificado de 532,62 m e obra orçada em R\$ 12.312.184,23 (valor já trazido para a referência temporal deste estudo).

Utilizando as melhores configurações de cada teste, obtivemos os resultados exibidos na tabela seguinte.

Modelo	Previsão	Erro Absoluto	Erro Percentual Relativo
Teste 01	R\$ 10.377.371,00	R\$ 1.934.813,23	15,71%
Teste 02	R\$ 10.136.761,00	R\$ 2.175.423,23	17,67%
Teste 03	R\$ 12.524.821,00	R\$ 212.636,77	1,73%

Teste 04	R\$ 8.627.796,00	R\$ 3.684.388,23	29,92%
Teste 05	R\$ 2.290.627,50	R\$ 10.021.556,73	81,40%
Teste 06	R\$ 8.632.756,00	R\$ 3.679.428,23	29,88%

Tabela 10 Estimativas do prédio da UNIFA

Com exceção do Teste 05, onde ocorre o fenômeno do overfitting conforme já discutido anteriormente, obtivemos resultados satisfatórios, em que o Teste 03 apresentou um erro desprezível de 1,73% e o Teste 04 um erro de 29,92% abaixo do efetivamente orçado, ficando dentro da faixa de precisão admissível de + 50% a -30% para estimativas conceituais (PHAOBUNJONG, 2002).

5.3 Confecção de Aplicativo para Smartphones

Para a criação de um modelo prático para a estimativa conceitual, criou-se um aplicativo para ser utilizado em celulares, programado totalmente em Python com a utilização da biblioteca Kivy. Trata-se de uma primeira versão, a qual deve ser melhorada conforme aumente a quantidade de projetos no banco de dados.

Lembrando sempre que tal ferramenta serve apenas para uma noção preliminar do preço de uma obra, estando sujeita a possíveis erros não identificados previamente, de forma que sua utilização serve apenas como apoio para estimativa, devendo-se o orçamentista utilizar outros métodos em paralelo.

Esta versão 1.0 faz estimativas com 3 redes neurais diferentes, as melhores obtidas nos testes que 1, 4 e 6, e seu uso é bem intuitivo, conforme mostra as capturas de telas abaixo.



Figura 14 Telas de Abertura e de Estimativa, respectivamente

O valor da área edificada é imputado na primeira linha e, após apertar o botão “Estimar Preço (R\$)” na segunda linha, são feito os cálculos para obtenção do preço.

5.4 Críticas e Limitações

Uma análise manual mais cuidadosa teve de ser feita, levando em conta aspectos qualitativos, devido ao tamanho do dataset. Por exemplo, a variável de localidade no BD2 foi prejudicial às previsões, pois gerou várias variáveis dummy causando overfitting quando utilizada na regressão (Teste 05).

Os resultados do modelo adotado são decorrentes diretamente do banco de dados disponível, de forma que até a topologia da rede foi escolhida de forma a melhorar as previsões. Ou seja, existem limitações decorrentes do fato de o modelo ser desenvolvido a partir de um banco de dados limitado e, portanto, válido sob determinadas condições.

Sendo assim, deve-se atentar para quando for feita previsão de gastos para obras muito diferentes da amostra geradora da rede, como:

- obras com área edificada muito maior que 20.000 m²;
- obras com área menor que 100 m²;
- obras de natureza distinta, como as de infraestruturas ou reformas;

- quando houver elementos isolados que encareçam a obra como um todo, como uma esquadria grande de alumínio (neste caso o que poderia ser feito é a estimativa sem o item em questão e depois adicioná-lo ao valor previsto);

- obras executadas com métodos construtivos excêntricos;

Por fim, alguns pontos poderiam ser considerados outliers, se o tempo de obra fosse discrepante aos demais, ou a área muito distante das demais, mas optou-se por mantê-los devido ao tamanho já reduzido das amostras.

6 Conclusão

A regressão na rede neural se aproximou de um comportamento linear, semelhante ao do método do CUB. Este estudo conclui que a rede neural como ferramenta para previsão de orçamentos é promissora e pode ser melhorada com a utilização de um banco de dados mais extenso.

Embora este procedimento seja considerado uma “caixa-preta” por muitos, ele tem potencial de agregar a importância de variáveis que o CUB não leva em consideração, como a altura da edificação e o perímetro.

O campo estudado neste trabalho se chama deep-learning e está em expansão atualmente. Vários trabalhos vêm sendo feitos e cada um é importante e contribui para o conhecimento humano, principalmente por este campo estar muito sujeito a experimentações.

Assim, este trabalho poderia ser continuado e aperfeiçoado fazendo-se testes como:

- expansão do banco de dados;
- variação dos inputs;
- novas configurações dos parâmetros da rede;
- inclusão do percentual de gasto em cada elemento da obra na análise (fundações, serviços preliminares, arquitetura, etc);
- avaliação do impacto de variáveis isoladas na previsão da obra;

Com relação à coleta de mais projetos, o CEPE está elaborando uma plataforma para inclusão de obras no SILOMS – Sistema Integrado de Logística de Material e de Serviços, que compila várias informações das obras acompanhadas pela FAB, além de variáveis importantes, como o valor efetivamente executado. Quando este sistema estiver em plena operação, será interessante a utilização de seu banco de dados para estudos como este, fornecendo informação que facilite a tomada de decisão de gestores e engenheiros.

7 Referências Bibliográficas

KINGMA, Diederik; BA, Jimmy. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

BOWLBY, Roger L.; SCHRIVER, William R. Observations on productivity and composition of building construction output in the United States, 1972–82. **Construction Management and Economics**, v. 4, n. 1, p. 1-18, 1986.

SCIKIT LEARN. **Compare the effect of different scalers on data with outliers**. Disponível em: <http://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#sphx-gl-auto-examples-preprocessing-plot-all-scaling-py>. Acesso em: 29 out. 2017.

CONSTRUCTION TUTS. **Type of Estimates**. Disponível em: <<https://www.constructiontuts.com/type-of-estimates/>>. Acesso em: 29 out. 2017.

COPE, G. Dummy Variable Trap in Regression Models. **ALGOSOME**. Disponível em: <<http://www.algosome.com/articles/dummy-variable-trap-regression.html>>. Acesso em: 29 out. 2017.

DASCHBACH, J. M.; APGAR, Henry. Design analysis through techniques of parametric cost estimation. **Engineering Costs and Production Economics**, v. 14, n. 2, p. 87-93, 1988.

JANTSCH, R. H. Engenharia de avaliações : análise da viabilidade econômica de empreendimento de base imobiliária. **LUME - Repositório Digital UFRGS**, 2015. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/127722>>. Acesso em: 02 nov. 2017.

PHAOBUNJONG, K. **Parametric cost estimating model for conceptual cost estimating of building construction projects**. 2002. Tese de Doutorado – The University of Texas at Austin. Austin, p. 203. 2002.

KARSHENAS, S. Predesign cost estimating method for multistory buildings. **Journal of Construction Engineering and Management**, v. 110, n. 1, p. 79-86, 1984.

KOUSKOULAS, V.;KOEHN, E. Predesign Cost-Estimation Function for Buildings. **Journal of the Construction Division**, v. 100, p. 589-604, Dezembro 1974. ISSN CO4.

KRIESEL, D. A Brief Introduction to Neural Networks. **dkriesel**, 2009. Disponível em: <http://www.dkriesel.com/en/science/neural_networks>. Acesso em: 29 out. 2017.

MATA, D. **An Extensive Introduction to Deep Neural Networks**. Disponível em: <<https://davidmatablog.wordpress.com/2017/06/12/an-extensive-introduction-to-neural-networks-part1/>>. Acesso em: 30 out. 2017.

PARK, Chansik; CHOI, Seok-In; KI-HONG, Kim. An efficient quantity takeoff method for the preliminary design phase. **AACE International Transactions**, p. T91, 1999.

RASCHKA, R. **About Feature Scaling and Normalization and the effect of standardization for machine learning algorithms**. Disponível em: <http://sebastianraschka.com/Articles/2014_about_feature_scaling.html>. Acesso em: 29 out. 2017.

SCIKIT LEARN. **Model evaluation: quantifying the quality of predictions**. Disponível em: <http://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics>. Acesso em: 31 out. 2017.

SINDUSCON-MG, Sindicato da Indústria da Construção Civil no Estado de Minas Gerais. **Custo Unitário Básico (CUB/m²)**. Belo Horizonte: Sinduscon - MG, 2013. Disponível em: <<http://www.cub.org.br/static/web/download/cartilha-saiba-mais-cub.pdf>>.

TRIBUNAL DE CONTAS DA UNIÃO. **Estudo sobre taxas referenciais de BDI de obras públicas e de materiais**. Tribunal de Contas da União. Brasília, DF. 2013.

SARLE, W. S. **Neural Network FAQ, part 1 of 7: Introduction**. Disponível em: <<ftp://ftp.sas.com/pub/neural/FAQ.html>>. 1997. Acesso em: 29 out. 2017.

Apêndice A – Bancos de Dados Individuais

PROJETO	VALOR DO CONTRATO (R\$ em Setembro de 2017)	PRAZO DE EXECUÇÃO (DIAS)	LOCALIDADE	PÉ-DIREITO (m)	NÚMERO DE PAVIMENTOS	ÁREA EDIFICADA (m ²)	PERÍMETRO DA EDIFICAÇÃO (m)	URBANA / INTERIOR
P1	R\$ 12.312.184,23	360	RJ	3,5	3	3312,7	532,62	URBANA
P2	R\$ 546.776,59	90	RN	2,78	2	178,224	86,40	URBANA
P3	R\$ 473.612,94	90	SP	2,78	2	178,224	86,40	INTERIOR
P4	R\$ 21.702,96	150	RN	2,5	1	5,29	9,20	URBANA
P5	R\$ 136.212,36	150	RN	2,66	5	78,4	0,00	URBANA
P6	R\$ 4.535.333,35	360	RJ	2,4	1	455,1	85,33	URBANA
P7	R\$ 45.851.950,63	900	DF	2,5	8	16750	2376,00	URBANA
P8	R\$ 22.117,94	60	RJ	2,4	1	11,7	25,82	URBANA
P9	R\$ 1.393.344,14	210	RJ	2,46	1	212,57	72,04	URBANA
P10	R\$ 28.428.557,23	480	RN	2,8	7	8321,41	1896,00	INTERIOR
P11	R\$ 5.676.968,52	180	MA	3	1	5053	398,90	INTERIOR
P12	R\$ 605.564,67	180	MA	3,35	1	53,2	29,20	INTERIOR
P13	R\$ 7.106.901,45	180	MA	3	1	1219,56	212,04	INTERIOR
P14	R\$ 6.753.469,92	180	MA	3,2	1	1180,01	187,25	INTERIOR
P15	R\$ 1.823.399,10	150	SP	3	2	1127,32	252,92	URBANA
P16	R\$ 1.689.381,84	360	SP	6,8	1	495	96,00	URBANA
P17	R\$ 2.345.764,98	360	SP	3	1	475,15	107,97	URBANA
P18	R\$ 5.192.075,31	180	AM	3	1	1004,96	176,32	INTERIOR
P19	R\$ 2.548.813,08	180	AM	3	1	569,09	111,62	INTERIOR
P20	R\$ 9.351.156,25	390	SC	3	4	3700	643,14	URBANA
P21	R\$ 181.857,48	180	SP	2,8	1	111,03	42,15	URBANA
P22	R\$ 334.232,52	180	SP	2,8	1	204,06	57,14	URBANA
P23	R\$ 1.286.646,15	180	SP	2,8	1	785,54	112,11	URBANA
P24	R\$ 1.394.895,82	180	SP	2,8	1	851,63	116,73	URBANA

Tabela 11 Compilado do Banco de dados BD1

PROJETO	DESCRIÇÃO DO OBJETO	VALOR DO CONTRATO (R\$ em Setembro de 2017)	PRAZO DE EXECUÇÃO (DIAS)	URBANA / INTERIOR	DATA DA PUBLICAÇÃO DO CONTRATO	ÁREA EDIFICADA (m ²)	LOCALIDADE
P25	CONTRATAÇÃO DE SERVIÇO DE ENGENHARIA PARA CONSTRUÇÃO DO CENTRO DE BIOLOGIA - CEBIO- DA UFRR.	R\$ 561.309,57	240	URBANA	27/01/2009	398,88	RR
P26	EMPREITADA POR PREÇO GLOBAL PARA A CONSTRUÇÃO DO PREDIO DA VARA DO TRABALHO DE CEARA-MIRIM.	R\$ 652.509,70	120	INTERIOR	07/12/2007	578,42	CE
P27	CONTRATAÇÃO DA EMPRESA CONSTRUTORA ZAMBONI LTDA PARA EXECUÇÃO DOS SERVIÇOS DE CONSTRUÇÃO DO PREDIO PRODUÇÃO VEGETAL, NO CENTRO DE CIÊNCIAS AGRÁRIAS, CAMPUS DE ALEGRE, CIDADE DE ALEGRE, ES.	R\$ 581.886,84	180	URBANA	18/03/2009	360,00	ES
P28	EXECUÇÃO DA OBRA DE CONSTRUÇÃO DO PREDIO DA EDITORA E GRÁFICA NO CAMPUS CARREIROS DA CONTRATANTE.	R\$ 859.343,55	270	URBANA	13/12/2010	448,20	RG
P29	CONTRATAÇÃO DE EMPRESA DE ENGENHARIA PARA EXECUTAR OBRA DE CONSTRUÇÃO DO PREDIO DA APS DE TAUÁ, SITUADO À RUA JORNALISTA HELDER FEITOSA, TAUÁZINHO, TAUÁ/CE.	R\$ 1.140.510,50	270	INTERIOR	12/03/2009	479,74	CE
P30	CONTRATAÇÃO DE EMPRESA DE ENGENHARIA PARA EXECUTAR OBRA DE CONSTRUÇÃO DO PREDIO DA AGÊNCIA DA PREVIDÊNCIA SOCIAL EM PARAMBU, LOCALIZADO À RUA DA MATRIZ, S/Nº, CENTRO, PARAMBU/CE.	R\$ 1.274.528,23	180	INTERIOR	19/02/2010	330,00	CE
P31	CONTRATAÇÃO DE EMPRESA ESPECIALIZADA DE ENGENHARIA PARA EXECUTAR OBRA DE CONSTRUÇÃO DA AGÊNCIA DA PREVIDÊNCIA SOCIAL (INSS) NO MUNICÍPIO DE PRESIDENTE FIGUEIREDO - LOCALIZADO NA RODOVIA FEDERAL BR 174, KM 107 MARGEM DIREITA - PRESIDENTE FIGUEIREDO/AM.	R\$ 1.479.054,51	180	INTERIOR	10/12/2009	330,00	AM
P32	CONSTRUÇÃO DO PREDIO DE BIOTECNOLOGIA DO CAMPUS UNIVERSITÁRIO UFPA-BELEM.	R\$ 1.393.692,23	180	URBANA	04/02/2011	850,00	PA
P33	CONTRATAÇÃO DE EMPRESA PARA CONSTRUÇÃO DA APS EM NOVA MUTUM/MT.	R\$ 1.629.539,08	180	INTERIOR	18/01/2010	1082,47	MT
P34	CONSTRUÇÃO DO PREDIO ADMINISTRATIVO DO CAMPUS DE CAMETA.	R\$ 1.703.095,57	180	URBANA	04/04/2011	615,96	PA
P35	CONSTRUÇÃO DO MÓDULO DE SAÚDE ESTADO MS	R\$ 1.942.538,84	300	URBANA	14/01/2009	345,80	MS

P36	CONTRATAÇÃO DE EMPRESA DE ENGENHARIA PARA EXECUTAR A OBRA DE CONSTRUÇÃO DA AGENCIA DA PREVIDENCIA SOCIAL EM CANTO DO BURITI/PI.	R\$ 1.893.780,99	180	INTERIOR	08/02/2011	362,57	PI
P37	CONSTRUÇÃO DO PREDIO FISILOGIA DO EXERCICIO DO CEFD/UFES LOCALIZADO NO CAMPUS DE GOIABEIRAS, VITORIA / ES COM FORNECIMENTO DE MAO-DE-OBRA E MATERIAL.	R\$ 2.182.265,09	360	INTERIOR	17/12/2009	854,60	ES
P38	CONSTRUÇÃO DE PREDIO ACADEMICO JUNTO AO IF FARROUPILHA - CAMPUS SANTO AUGUSTO, COM AREA TOTAL DE 1.074,46 M², SITO NA RUA FABIO JOAO SINDOLHE, 1100, BAIRRO FLORESTA, SANTO AUGUSTO/RS.	R\$ 2.048.323,92	300	INTERIOR	17/11/2010	620,00	RS
P39	PRESTAÇÃO DOS SERVIÇOS DE CONSTRUÇÃO DO PREDIO SEDE DA UNIDADE AVANÇADA DE ARAGUAINA, MUNICIPIO DE ARAGUAINA/TO, NO IMÓVEL SITUADO NA AV. NEIEF MURAD, LOTE 5-A, QUADRA 01, DESMEMBRADO DA CHACARA Nº 47-A - ARAGUAINA/TO, COM AREA TOTAL DE 1.689,00 M² E AREA A SER CONSTRUÍDA DE 835,75 M², REFERENTE AO EDITAL DE TOMADA DE PREÇOS Nº 01/2010 E SEUS ANEXOS.	R\$ 2.101.353,61	270	INTERIOR	23/12/2010	835,75	TO
P40	O PRESENTE CONTRATO TEM COMO OBJETO A CONSTRUÇÃO DE UM PREDIO EM ALVENARIA, MEDINDO 1.520,85 M² (CONFORME DISPOE O ANEXO III DO EDITAL DE CONCORRÊNCIA Nº 03/2009, CONSTANTE AFLS. 167 A 185 DO PROCESSO Nº 23107.004027/2009-25 - VOL. I), DESTINADO AO ARQUIVO GERAL DAUFAC NO CAMPUS UNIVERSITARIO DA UNIVERSIDADE FEDERAL DO ACRE, DE ACORDO COM A DOCUMENTAÇÃO /PROJETOS ELABORADOS PELA SUPERINTENDENCIA DO CAMPUS.	R\$ 2.311.855,03	300	URBANA	29/12/2009	1520,85	AC
P41	CONSTRUÇÃO (1 ETAPA) DO PREDIO DO FORUM ELEITORAL DA 4 ZONA, EM CRUZEIRO DO SUL, COM FORNECIMENTO DE MATERIAL E MAO DE OBRA, CONFORME ESPECIFICAÇÕES CONSTANTES DO EDITAL E ANEXOS.	R\$ 2.308.608,35	150	URBANA	30/09/2010	2392,00	AC

P42	CONSTRUCAO DO PREDIO DO RESTAURANTE UNIVERSITARIO E PRACA DE SERVICOS DA UFVJM-CAMPUS AVANÇADO DO MUCURI-TO/MG.	R\$ 3.819.134,41	365	INTERIOR	01/06/2011	3702,60	MG
P43	CONTRATAÇÃO DE EMPRESA JURÍDICA PARA A CONSTRUÇÃO DO PREDIO DA BIBLIOTECA DO CAMPUS ALTO PARAÓPEBA - CAP DA UFSJ	R\$ 4.697.557,78	240	INTERIOR	18/10/2010	2182,73	MG
P44	SERVICOS DE CONSTRUÇÃO DO EDIFÍCIO MUL-TIDISCIPLINAR NO CAMPUS DE CATALÃO-GO/UGF.	R\$ 4.992.576,40	365	INTERIOR	11/02/2010	3372,00	GO
P45	EXECUÇÃO TOTAL DA OBRA DA UNED DE TIMONMA, DESTE CEFET-MA.	R\$ 5.605.830,47	180	INTERIOR	12/09/2008	4086,16	MA
P46	EXECUÇÃO DA OBRA DE EDIFICAÇÃO DO CAMPUS DE CAXIAS-MA.	R\$ 5.461.318,43	360	INTERIOR	30/09/2010	5711,23	MA
P47	CONSTRUÇÃO DA UNIVERSIDADE ESTADUAL DE RORAIMA NO MUNICÍPIO DE CARACARAI	R\$ 6.846.070,06	180	INTERIOR	02/01/2008	1326,00	RR
P48	CONTRATAÇÃO DE EMPRESA DE ENGENHARIA ESPECIALIZADA PARA A EXECUÇÃO DA 1ª ETAPA DAS OBRAS DE CONSTRUÇÃO DA SEDE (PREDIO PRINCIPAL) DO GALPAO DO CURSO DE MECANICA DO IFES CAMPUS SMATEUS.	R\$ 5.991.864,31	240	INTERIOR	24/08/2009	3922,01	ES
P49	EXECUÇÃO DE SERVIÇOS DE CONSTRUÇÃO TOTAL DA OBRAS DE EDIFICAÇÕES DA UNIDADE DE ENSINO DESCENTRALIZADA - UNED - DE ALMENARA.	R\$ 6.971.516,83	180	INTERIOR	09/01/2009	3922,01	MG
P50	EXECUÇÃO TOTAL DA OBRA DA UNIDADE DE ENSINO DO CEFET-MA NO MUNICÍPIO DE PINHEIRO-MA.	R\$ 9.571.127,65	360	INTERIOR	23/01/2009	5486,36	MA
P51	PRESTAÇÃO DOS SERVIÇOS DE EXECUÇÃO DE OBRA DE CONSTRUÇÃO DO ANEXO DO EDIFÍCIO-SEDE DA SUBSEÇÃO JUDICIÁRIA DE CAMPINA GRANDE/PB	R\$ 10.890.597,89	540	URBANA	19/12/2008	3196,00	PB
P52	OBRAS DE CONSTRUÇÃO DE IMÓVEL DESTINADO A ABRIGAR O FORUM TRABALHISTA DE RIBEIRÃO PRETO	R\$ 12.673.940,39	420	INTERIOR	19/03/2008	4588,82	SP
P53	EXECUÇÃO DA OBRA DE CONSTRUÇÃO DE PREDIO DESTINADO À UNIDADE DE ENSINO E DOCÊNCIA - UED DA FACULDADE UNB GAMA.	R\$ 11.791.899,35	240	URBANA	30/12/2009	4485,00	DF

P54	OBRA DE CONSTRUCAO D PREDIO DESTINADO A UNIDADE ACADEMICA - UAC, DA FACULDADE UNB PLANALTINA, LOCALIZADA NA AREA UNIVERSITARIA N. 1, PLANALTINA/DF.	R\$ 13.405.531,69	240	URBANA	27/04/2009	4795,00	DF
P55	LICITACAO DO TIPO MENOR PRECO PARA EXECUCAO DA OBRA DE CONSTRUCAO DO INSTITUTO DE CULTURA E ARTE - ICA - CAMPUS DO PICI/UFC, SOB A FORMA DE EMPREITADA POR PRECO GLOBAL, TUDO DE ACORDO COM OS ELEMENTOS TECNICOS DISCRIMINADOS NO EDITAL DE CONCORRENCIA PUBLICA NACIONAL Nº 15/2010 E DEMAIS ESPECIFICACOES EXPRESSAS, QUE TAMBEM PASSAM A FAZER PARTE INTEGRANTE DESTE INSTRUMENTO.	R\$ 14.724.519,79	330	URBANA	08/04/2011	4355,91	CE
P56	CONSTRUCAO DE UM PREDIO DE 05 (CINCO) PAVIMENTOS, DO TIPO UFASA, A SER DENOMINADO COM BLOCO P, PARA ABRIGAR AS INSTALACOES DE SALAS DE AULA DO INSTITUTO DE CIENCIAS HUMANAS E FILOSOFIA - ICHF.	R\$ 16.094.396,87	360	INTERIOR	18/06/2010	5144,00	RJ
P57	CONSTRUCAO DO CENTRO DE CONVENCOES	R\$ 17.810.492,58	900	URBANA	21/01/2009	2100,00	PB
P58	O PRESENTE INSTRUMENTO DE CONTRATO TEM POR OBJETO A CONSTRUCAO DO CAMPUS DE MACAPA, DOIFAP, NO REGIME DE EXECUCAO EMPREITADA POR PRECO GLOBAL FORNECIMENTO TOTAL DE MATERIAIS DE MAODE OBRA, CONFORME DETALHAMENTO CONSTANTE NO EDITAL DA CONCORRENCIA Nº 001/2009 E SEUS ANEXOS.	R\$ 21.134.445,45	420	URBANA	23/10/2009	10925,10	PA
P59	CONSTRUCAO DO HOSPITAL DAS CLINICAS DE RORAIMA	R\$ 34.268.515,95	450	URBANA	16/01/2008	9863,08	RR
P60	CONSTRUCAO DA PENITENCIARIA ESTADUAL DE GUAIBA	R\$ 38.121.517,08	360	INTERIOR	18/01/2008	14390,71	RS
P61	CONSTRUCAO DE DOIS PREDIOS, TIPO UFASA, COM 05 (CINCO) PAVIMENTOS EM ESTRUTUTA DE CONCRETO ARMADO PARA ABRIGAR SALAS DE AULA DA FAC. DE ADMINISTRACAO.	R\$ 34.651.158,00	540	INTERIOR	24/06/2010	10288,00	RJ
P62	CONSTRUCAO DO ANEXO AO EDIFICIO-SEDE DO TRE DA BAHIA.	R\$ 37.175.736,77	540	URBANA	14/08/2009	7360,17	BA

P63	EXECUCAO TOTAL DAS OBRAS DO EDIFICIO DA FUTURA SEDE DA DELEGACIA DA RECEITA FEDERAL DO BRASIL EM BELO HORIZONTE, INCLUINDO A ELABORACAO DO PROJETO EXECUTIVO CONCOMITANTEMENTE À OBRA, CONFORME ESPECIFICACOES DO PROJETO BASICO, CRONOGRAMA E PROPOSTA.	R\$ 39.306.172,89	1080	URBANA	24/12/2009	13537,66	MG
P64	EXECUCAO DA OBRA DE EDIFICACAO DO EDIFICIO-SEDE DA PROCURADORIA DA REPUBLICA NA PARAIBA. A EDIFICACAO TERA, APROXIMADAMENTE, 9.600 M2 DE AREA CONSTRUIDA.	R\$ 37.700.213,05	540	URBANA	10/01/2011	9600,00	PB
P65	CONSTRUCAO DE UNIDADE DE SAUDE	R\$ 48.134.672,10	365	URBANA	16/01/2008	11656,00	AC
P66	CONTRATACAO DE EMPRESA DE ENGENHARIA PARA CONSTRUCAO DO EDIFICIO SEDE E SEUS ANEXOS DASUPERINTENDENCIA REGIONAL DO DPF EM RORAIMA.	R\$ 44.489.615,67	912	URBANA	22/12/2009	10241,66	RR
P67	Obra de Construção do Bloco I do Quartel General do Exército, no Setor Militar Urbano, em Brasília/DF.	R\$ 55.564.074,20	540	URBANA	01/06/2008	19850,00	DF
P68	EXECUCAO DAS OBRAS DE EDIFICACAO DO EDIFICIO-SEDE DO DEPARTAMENTO DE POLICIA RODOVIARIA FEDERAL.	R\$ 63.863.682,90	540	URBANA	31/05/2007	12253,46	DF
P69	CONSTRUCAO HOSPITAL DA CRIANCA	R\$ 64.549.671,58	545	INTERIOR	03/07/2008	10983,00	BA
P70	CONSTRUCAO DA PENITENCIARIA FEMININA EM TREMEMBE	R\$ 72.515.058,93	300	INTERIOR	14/01/2008	13800,00	SP

Tabela 12 Compilado do Banco de dados BD2

Apêndice B – Banco de Dados Unificado

PROJETO	VALOR DO CONTRATO (R\$ em Setembro de 2017)	PRAZO DE EXECUÇÃO (DIAS)	ÁREA EDIFICADA (m ²)	LOCALIDADE	URBANA / INTERIOR
P1	R\$ 12.312.184,23	360	3312,70	RJ	URBANA
P2	R\$ 546.776,59	90	178,22	RN	URBANA
P3	R\$ 473.612,94	90	178,22	SP	INTERIOR
P4	R\$ 21.702,96	150	5,29	RN	URBANA
P5	R\$ 136.212,36	150	78,40	RN	URBANA
P6	R\$ 4.535.333,35	360	455,10	RJ	URBANA
P7	R\$ 45.851.950,63	900	16750,00	DF	URBANA
P8	R\$ 22.117,94	60	11,70	RJ	URBANA
P9	R\$ 1.393.344,14	210	212,57	RJ	URBANA
P10	R\$ 28.428.557,23	480	8321,41	RN	INTERIOR
P11	R\$ 5.676.968,52	180	5053,00	MA	INTERIOR
P12	R\$ 605.564,67	180	53,20	MA	INTERIOR
P13	R\$ 7.106.901,45	180	1219,56	MA	INTERIOR
P14	R\$ 6.753.469,92	180	1180,01	MA	INTERIOR
P15	R\$ 1.823.399,10	150	1127,32	SP	URBANA
P16	R\$ 1.689.381,84	360	495,00	SP	URBANA
P17	R\$ 2.345.764,98	360	475,15	SP	URBANA
P18	R\$ 5.192.075,31	180	1004,96	AM	INTERIOR
P19	R\$ 2.548.813,08	180	569,09	AM	INTERIOR
P20	R\$ 9.351.156,25	390	3700,00	SC	URBANA
P21	R\$ 181.857,48	180	111,03	SP	URBANA
P22	R\$ 334.232,52	180	204,06	SP	URBANA
P23	R\$ 1.286.646,15	180	785,54	SP	URBANA
P24	R\$ 1.394.895,82	180	851,63	SP	URBANA
P25	R\$ 561.309,57	240	398,88	RR	URBANA
P26	R\$ 652.509,70	120	578,42	CE	INTERIOR
P27	R\$ 581.886,84	180	360,00	ES	URBANA
P28	R\$ 859.343,55	270	448,20	RG	URBANA
P29	R\$ 1.140.510,50	270	479,74	CE	INTERIOR
P30	R\$ 1.274.528,23	180	330,00	CE	INTERIOR
P31	R\$ 1.479.054,51	180	330,00	AM	INTERIOR
P32	R\$ 1.393.692,23	180	850,00	PA	URBANA
P33	R\$ 1.629.539,08	180	1082,47	MT	INTERIOR
P34	R\$ 1.703.095,57	180	615,96	PA	URBANA
P35	R\$ 1.942.538,84	300	345,80	MS	URBANA
P36	R\$ 1.893.780,99	180	362,57	PI	INTERIOR
P37	R\$ 2.182.265,09	360	854,60	ES	INTERIOR
P38	R\$ 2.048.323,92	300	620,00	RS	INTERIOR
P39	R\$ 2.101.353,61	270	835,75	TO	INTERIOR

P40	R\$ 2.311.855,03	300	1520,85	AC	URBANA
P41	R\$ 2.308.608,35	150	2392,00	AC	URBANA
P42	R\$ 3.819.134,41	365	3702,60	MG	INTERIOR
P43	R\$ 4.697.557,78	240	2182,73	MG	INTERIOR
P44	R\$ 4.992.576,40	365	3372,00	GO	INTERIOR
P45	R\$ 5.605.830,47	180	4086,16	MA	INTERIOR
P46	R\$ 5.461.318,43	360	5711,23	MA	INTERIOR
P47	R\$ 6.846.070,06	180	1326,00	RR	INTERIOR
P48	R\$ 5.991.864,31	240	3922,01	ES	INTERIOR
P49	R\$ 6.971.516,83	180	3922,01	MG	INTERIOR
P50	R\$ 9.571.127,65	360	5486,36	MA	INTERIOR
P51	R\$ 10.890.597,89	540	3196,00	PB	URBANA
P52	R\$ 12.673.940,39	420	4588,82	SP	INTERIOR
P53	R\$ 11.791.899,35	240	4485,00	DF	URBANA
P54	R\$ 13.405.531,69	240	4795,00	DF	URBANA
P55	R\$ 14.724.519,79	330	4355,91	CE	URBANA
P56	R\$ 16.094.396,87	360	5144,00	RJ	INTERIOR
P57	R\$ 17.810.492,58	900	2100,00	PB	URBANA
P58	R\$ 21.134.445,45	420	10925,10	PA	URBANA
P59	R\$ 34.268.515,95	450	9863,08	RR	URBANA
P60	R\$ 38.121.517,08	360	14390,71	RS	INTERIOR
P61	R\$ 34.651.158,00	540	10288,00	RJ	INTERIOR
P62	R\$ 37.175.736,77	540	7360,17	BA	URBANA
P63	R\$ 39.306.172,89	1080	13537,66	MG	URBANA
P64	R\$ 37.700.213,05	540	9600,00	PB	URBANA
P65	R\$ 48.134.672,10	365	11656,00	AC	URBANA
P66	R\$ 44.489.615,67	912	10241,66	RR	URBANA
P67	R\$ 55.564.074,20	540	19850,00	DF	URBANA
P68	R\$ 63.863.682,90	540	12253,46	DF	URBANA
P69	R\$ 64.549.671,58	545	10983,00	BA	INTERIOR
P70	R\$ 72.515.058,93	300	13800,00	SP	INTERIOR

Tabela 13 BD3: Composição das amostras de BD1 e BD2

Apêndice C – Códigos Utilizados

```

import keras
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from keras.models import Sequential
from keras.layers import Dense
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.preprocessing import StandardScaler, MinMaxScaler,
MaxAbsScaler, RobustScaler
import datetime
t_ini = datetime.datetime.now()

def rede_neural(parametros,dimensao_inicial):

    (neuronio_c1, neuronio_c2, neuronio_c3, f_at_1, f_at_2, f_at_3,
    metrica_erro, taxa_aprendizagem, qtde_epocas,
    tamanho_batch, otimizador) = parametros

    model = Sequential()

    model.add(Dense(units = neuronio_c1, kernel_initializer = 'uniform',
    activation = f_at_1, input_dim = dimensao_inicial))

    if neuronio_c2>0:
        model.add(Dense(units=neuronio_c2, kernel_initializer='uniform',
    activation=f_at_2))

    model.add(Dense(units=neuronio_c3, kernel_initializer='uniform',
    activation=f_at_3))

    keras.optimizers.Adam(lr=taxa_aprendizagem, beta_1=0.9, beta_2=0.999,
    epsilon=1e-08, decay=0.0)

    model.compile(loss=metrica_erro, optimizer=otimizador)

    return model

#Primeiro Teste
teste_numero = 'avulso'#4 e 5
imprime=True
imprime_grafico=True

if teste_numero == 'avulso':
    campos_selecionados_inputs = [6] # deixar ordenado
    campos_selecionados_outputs = [2] # deixar ordenado
    campos_categoricos = [] # deixar ordenado
    escalar = True #
    bd = '2' # base de dados
    casos = [(8, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.001,
    1000, 4, 'adam'),
    #(4, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error',
    0.0001, 1000, 4, 'adam'),
    ]
if teste_numero == '1':

```

```

campos_seleccionados_inputs = [6]#deixar ordenado
campos_seleccionados_outputs = [1]#deixar ordenado
campos_categoricos = []#deixar ordenado
escalar = True#
bd = '1'
casos = [(1, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
1000, 4, 'adam'),
          (2, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (3, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
3000, 4, 'adam'),
          (2, 1, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (2, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam')],
]
if teste_numero == '2':
campos_seleccionados_inputs = [6,7]#deixar ordenado
campos_seleccionados_outputs = [1]#deixar ordenado
campos_categoricos = []#deixar ordenado
escalar = True#
bd = '1'#base de dados
casos = [(4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
1000, 4, 'adam'),
          (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
3000, 4, 'adam'),
          (5, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (3, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (3, 1, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (4, 0, 1, 'relu', 'relu', 'tanh', 'mean_squared_error', 0.01,
2000, 4, 'adam')],
]
if teste_numero == '3':
campos_seleccionados_inputs = [3,4,5,6,7,8]#deixar ordenado
campos_seleccionados_outputs = [1]#deixar ordenado
campos_categoricos = [3,8]#deixar ordenado
escalar = True#
bd = '1'#base de dados
casos = [(4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
1000, 4, 'adam'),
          (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
5000, 4, 'adam'),
          (4, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (6, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (6, 4, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (8, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (8, 4, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
          (4, 2, 1, 'relu', 'tanh', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam')],
]

```

```

        (4, 2, 1, 'tanh', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
    ]
if teste_numero == '4':#Só com área
    campos_selecionados_inputs = [6]#deixar ordenado
    campos_selecionados_outputs = [2]#deixar ordenado
    campos_categoricos = []#deixar ordenado
    escalar = True#
    bd = '2'#base de dados
    casos = [(10, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.1,
1000, 4, 'adam'),
            (8, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.1,
1000, 4, 'adam'),
            (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.1,
1000, 4, 'adam'),
            (8, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error',
0.0001, 1000, 4, 'adam'),
            (4, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error',
0.0001, 1000, 4, 'adam'),
            (1, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error',
0.0001, 1000, 4, 'adam'),
    ]
if teste_numero == '5':
    campos_selecionados_inputs = [4,6,7]#deixar ordenado
    campos_selecionados_outputs = [2]#deixar ordenado
    campos_categoricos = [4,7]#deixar ordenado
    escalar = True#
    bd = '2'#base de dados
    casos = [(10, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (8, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (6, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (2, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (1, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (4, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
    ]
if teste_numero == '6':#só área
    campos_selecionados_inputs = [3]#deixar ordenado
    campos_selecionados_outputs = [1]#deixar ordenado
    campos_categoricos = []#deixar ordenado
    escalar = True#
    bd = '3'#base de dados
    casos = [(6, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
10000, 4, 'adam'),
            (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
5000, 4, 'adam'),
            (4, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (6, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (6, 4, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
    ]

```

```

        (8, 2, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
        (8, 4, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
        (4, 2, 1, 'relu', 'tanh', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
        (4, 2, 1, 'tanh', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
    ]
if teste_numero == '7':
    campos_selecionados_inputs = [3,4,5]#deixar ordenado
    campos_selecionados_outputs = [1]#deixar ordenado
    campos_categoricos = [4,5]#deixar ordenado
    escalar = True#
    bd = '3'#base de dados
    casos = [(10, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (8, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (6, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (4, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (2, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
            (1, 0, 1, 'relu', 'relu', 'relu', 'mean_squared_error', 0.01,
2000, 4, 'adam'),
    ]

dataset = pd.ExcelFile("BD.xlsx").parse('BD ' + bd )
X=dataset.iloc[:,campos_selecionados_inputs].values
y=dataset.iloc[:,campos_selecionados_outputs].values

# Tratando as variáveis categóricas
if campos_categoricos and (set(campos_categoricos) <=
set(campos_selecionados_inputs)):
    new_categ=[]
    maximos=[]
    for categ in campos_categoricos: #Transformando os labels em números
        new_posic=campos_selecionados_inputs.index(categ)
        new_categ.append(new_posic)
        labelencoder_X = LabelEncoder()
        X[:,new_posic]=labelencoder_X.fit_transform(X[:,new_posic])
        maximos.append(max(X[:,new_posic]))
    onehotencoder=OneHotEncoder(categorical_features=new_categ)
    aux=onehotencoder.fit_transform(X).toarray()
    X=aux# dummy variavel trap

deletar=np.array(maximos+[0])+np.array([0]+np.cumsum(np.array(maximos)+1).t
olist())
deletar=deletar[0:len(deletar)-1]
X=np.delete(X,deletar,1)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2,
random_state = 0)

if escalar:
    std_sc = StandardScaler()
    min_max_sc=MinMaxScaler()
    max_abs_sc=MaxAbsScaler()
    robust_sc=RobustScaler()#Vide 
```

nets/part2/section-16.html

```

sc_x=MaxAbsScaler() #max_abs_sc#max_abs_sc

X_train = sc_x.fit_transform(X_train)
X_test = sc_x.transform(X_test)
X0=X
X=sc_x.transform(X)
#y_train = sc.fit_transform(y_train)
#y_test = sc.transform(y_test) #<Tá dando merda nessas duas linhas

sc_y=MaxAbsScaler() #max_abs_sc

y_train = sc_y.fit_transform(y_train)
y_test = sc_y.transform(y_test)
y=sc_y.transform(y)

np.random.seed(7)
dimensao_inicial=X_train.shape[1]
#PARÂMETROS A VARIAR

comentarios=[]

resultados=[]

contador=0

for caso in casos:
    contador=contador+1
    modelo=rede_neural(caso,dimensao_inicial)
    (neuronio_c1, neuronio_c2, neuronio_c3, f_at_1, f_at_2, f_at_3,
    metrica_erro, taxa_aprendizagem, qtde_epocas,
    tamanho_batch, otimizador) = caso
    modelo.fit(X_train, y_train, batch_size = tamanho_batch, epochs =
    qtde_epocas)

    y_predicao=modelo.predict(X)

    #Métricas
    r2=metrics.r2_score(y, y_predicao)
    exp_var=metrics.explained_variance_score(y, y_predicao)
    mae=metrics.mean_absolute_error(y, y_predicao)
    msr=metrics.mean_squared_error(y, y_predicao)
    MedAe=metrics.median_absolute_error(y, y_predicao)

    resultados.append([neuronio_c1, neuronio_c2, neuronio_c3, f_at_1,
    f_at_2, f_at_3,qtde_epocas,'% .4f' %r2,
    '% .4f' %exp_var,'% .4f' %mae,'% .4f' %msr,'% .4f'
    %MedAe])

    if teste_numero == '1':
        if imprime_grafico == True:
            N = 500
            x1 = np.linspace(0, 1, N, endpoint=True)
            y1 = np.linspace(0, 0, N, endpoint=True)
            for i in range(0, N):
                new_prediction = modelo.predict(np.array([x1[i]]))
                y1[i] = new_prediction[0][0]
            y_0=sc_y.inverse_transform(y)
            y_p=sc_y.inverse_transform(y_predicao)
            y1 = y1.reshape(y1.shape[0], 1)

```

```

y1=sc_y.inverse_transform(y1)
x1 = x1.reshape(x1.shape[0], 1)
x1 = sc_x.inverse_transform(x1)

x_0=sc_x.inverse_transform(X)
font = {'family': 'serif',
        'color': 'black',
        'weight': 'normal',
        'size': 12,
        }
fig = plt.figure()
plt.plot(x_0,y_0,'ko',label='Amostragem')
plt.plot(x1,y1,'k--',label='Curva Prevista')
#plt.plot(X,y_predicao,'k*',label='Previsão para amostragem')
#plt.title('', fontdict=font)
texto_perf='Performance:\n'+R2: '+str(r2)+'\nexp_var:
'+str(exp_var)

#texto_param=str(neuronio_c1)+f_at_1+str(neuronio_c2)+f_at_2+str(neuronio_c
3)+f_at_3+'_'+str(qtde_epocas)
plt.text(1e3,3.5e7,texto_perf, fontdict=font)
plt.xlabel('Área da Edificação (m²)', fontdict=font)
plt.ylabel('Preço da Obra (R$)', fontdict=font)
plt.legend()
fig.savefig('Teste '+teste_numero+'_'+ ' Configuração
'+str(contador)+'_'+'.png')
if teste_numero == '4':
    if imprime_grafico == True:
        N = 500
        x1 = np.linspace(0, 1, N, endpoint=True)
        y1 = np.linspace(0, 0, N, endpoint=True)
        for i in range(0, N):
            new_prediction = model.predict(np.array([x1[i]]))
            y1[i] = new_prediction[0][0]
        y_0=sc_y.inverse_transform(y)
        y_p=sc_y.inverse_transform(y_predicao)
        y1 = y1.reshape(y1.shape[0], 1)
        y1=sc_y.inverse_transform(y1)
        x1 = x1.reshape(x1.shape[0], 1)
        x1 = sc_x.inverse_transform(x1)

        x_0=sc_x.inverse_transform(X)
        font = {'family': 'serif',
                'color': 'black',
                'weight': 'normal',
                'size': 12,
                }
        fig = plt.figure()
        plt.plot(x_0,y_0,'ko',label='Amostragem')
        plt.plot(x1,y1,'k--',label='Curva Prevista')
        #plt.plot(X,y_predicao,'k*',label='Previsão para amostragem')
        #plt.title('', fontdict=font)
        texto_perf='Performance:\n'+R2: '+str(r2)+'\nexp_var:
'+str(exp_var)

#texto_param=str(neuronio_c1)+f_at_1+str(neuronio_c2)+f_at_2+str(neuronio_c
3)+f_at_3+'_'+str(qtde_epocas)
plt.text(1e3,3.5e7,texto_perf, fontdict=font)
plt.xlabel('Área da Edificação (m²)', fontdict=font)
plt.ylabel('Preço da Obra (R$)', fontdict=font)
plt.legend()

```

```

        fig.savefig('Teste '+teste_numero+'_'+' Configuração
'+str(contador)+'_'+'.png')
        if teste_numero == '6':
            if imprime_grafico == True:
                N = 500
                x1 = np.linspace(0, 1, N, endpoint=True)
                y1 = np.linspace(0, 0, N, endpoint=True)
                for i in range(0, N):
                    new_prediction = model.predict(np.array([x1[i]]))
                    y1[i] = new_prediction[0][0]
                y_0=sc_y.inverse_transform(y)
                y_p=sc_y.inverse_transform(y_predicao)
                y1 = y1.reshape(y1.shape[0], 1)
                y1=sc_y.inverse_transform(y1)
                x1 = x1.reshape(x1.shape[0], 1)
                x1 = sc_x.inverse_transform(x1)

                x_0=sc_x.inverse_transform(X)
                font = {'family': 'serif',
                        'color': 'black',
                        'weight': 'normal',
                        'size': 12,
                        }
                fig = plt.figure()
                plt.plot(x_0,y_0,'ko',label='Amostragem')
                plt.plot(x1,y1,'k--',label='Curva Prevista')
                #plt.plot(X,y_predicao,'k*',label='Previsão para amostragem')
                #plt.title('', fontdict=font)
                texto_perf='Performance:\n'+R2: '+str(r2)+'\nexp_var:
'+str(exp_var)

#texto_param=str(neuronio_c1)+f_at_1+str(neuronio_c2)+f_at_2+str(neuronio_c
3)+f_at_3+'_'+str(qtde_epocas)
                plt.text(1e3,3.5e7,texto_perf, fontdict=font)
                plt.xlabel('Área da Edificação (m²)', fontdict=font)
                plt.ylabel('Preço da Obra (R$)', fontdict=font)
                plt.legend()
                fig.savefig('Teste '+teste_numero+'_'+' Configuração
'+str(contador)+'_'+'.png')

if imprime == True:
    linhas=pd.Index(['Neurônios 1ª camada',
                    'Neurônios 2ª camada',
                    'Neurônios 3ª camada',
                    'Função de Ativação 1ª camada',
                    'Função de Ativação 2ª camada',
                    'Função de Ativação 3ª camada',
                    'Número de Épocas',
                    'R²',
                    'Variância Explicada',
                    'Erro Absoluto Médio',
                    'Erro Quadrático Médio',
                    'Mediana do Erro Absoluto'])
    #linhas=pd.Index(['Teste ' + teste_numero])
    tabela=np.transpose(np.array(resultados))
    ind_casos=[str(x+1) for x in range(len(casos))]
    ind_casos=['Configuração ' + s for s in ind_casos]
    colunas=pd.Index(ind_casos)
    df = pd.DataFrame(tabela, columns=colunas,index=linhas)

writer = pd.ExcelWriter('resultados ' + teste_numero + '.xlsx')

```

```
df.to_excel(writer, sheet_name='Sheet1')
writer.save()

t_fin = datetime.datetime.now()
print('Tempo decorrido: ' + str(t_fin - t_ini))
```

FOLHA DE REGISTRO DO DOCUMENTO			
1. CLASSIFICAÇÃO/TIPO TC	2. DATA 24 de novembro de 2017	3. REGISTRO N° DCTA/ITA/TC-152/2017	4. N° DE PÁGINAS 68
5. TÍTULO E SUBTÍTULO: Estimativa do Custo de Obras com a utilização de Redes Neurais			
6. AUTOR(ES): Matthaeus Muniz Deusdará Ferreira Lopes			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Instituto Tecnológico de Aeronáutica – ITA			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: 1. Redes Neurais Artificiais. 2. Estimativa Paramétrica de Obras. 3. Python.			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Redes neurais; Estimativas de custos; Planejamento econômico; Tomada de decisões; Administração.			
10. APRESENTAÇÃO: <input checked="" type="checkbox"/> Nacional <input type="checkbox"/> Internacional ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Major Engenheiro Frank Cabral de Freitas Amaral; coorientador: Major Engenheiro Robinson SAMUEL Boschetti. Publicado em 2017			
11. RESUMO: Na fase de planejamento e projeto de uma obra de engenharia, todos os custos previstos para sua realização compõem o seu orçamento, peça técnica fundamental que baliza a tomada de decisão e a própria contratação em si. Contudo, nem sempre os gestores dispõem de tempo hábil para revisar todos os itens que integram o orçamento estimativo da obra e, por vezes, pequenas falhas, como erros em composições de custo unitário ou em quantitativos, implicam enormes transtornos para a administração, culminando eventualmente até mesmo com danos ao erário, caso não sejam percebidos e saneados oportunamente. Tais prejuízos são ainda mais significativos quando tratamos de obras de grande e médio porte. Nesse sentido, o presente trabalho analisou comparativamente uma amostra de obras, em termos dos respectivos orçamentos, buscando identificar parâmetros representativos do custo e propôs um modelo, baseado em Redes Neurais Artificiais, com o intuito de facilitar a revisão sistemática a nível macro de planilhas orçamentárias, fornecendo um parâmetro para verificação da assertividade de orçamentos de futuras obras, de sorte a mitigar possíveis desvios.			
12. GRAU DE SIGILO: <input checked="" type="checkbox"/> OSTENSIVO <input type="checkbox"/> RESERVADO <input type="checkbox"/> SECRETO			