PROJETO DE REDES NEURAIS

APLICAÇÃO DO MODELO DE REGRESSÃO UTILIZANDO WEKA PARA PREDIÇÃO

Aluno: Marlon de Lima Castro

Prof. : Meuser Valença

Recife, 25 de junho de 2019.

# LISTA DE FIGURAS

[Figura 1 – Exemplo de uma Regressão linear 4](#_Toc12313393)

[Figura 2 - Dados brutos recebidos 7](#_Toc12313394)

[Figura 3 - Dados normalizados 8](#_Toc12313395)

[Figura 4 - Utiilizando base de de treinamento 9](#_Toc12313396)

[Figura 5 - Validação cruzada do Modelo 10](#_Toc12313397)

[Figura 6 - Testes do Modelo de Regressão Linear 11](#_Toc12313398)

[Figura 7 - Dependência da variável x6 12](#_Toc12313399)

# SUMÁRIO

[LISTA DE FIGURAS 2](#_Toc12313631)

[SUMÁRIO 3](#_Toc12313632)

[1. INTRODUÇÃO 4](#_Toc12313633)

[1.1. OBJETIVOS GERAIS 5](#_Toc12313634)

[1.2. OBJETIVOS ESPECÍFICOS 5](#_Toc12313635)

[2. SOFWARES UTILIZADOS 6](#_Toc12313636)

[2.1. DESCRIÇÃO DOS SOFTWARES 6](#_Toc12313637)

[2.1.1. WEKA 6](#_Toc12313638)

[2.1.2. MICROSOFT OFFICE 2010 (EXCEL) 6](#_Toc12313639)

[3. PROCEDIMENTOS REALIZADOS 7](#_Toc12313640)

[3.1. DOWNLOAD DO SOFTWARE WEKA 7](#_Toc12313641)

[3.2. OBTENÇÃO DOS DADOS A SEREM ANALISADOS 7](#_Toc12313642)

[3.3. NORMALIZAÇÃO DOS DADOS 7](#_Toc12313643)

[3.4. CRIAR UM MODELO UTILIZANDO O WEKA 9](#_Toc12313644)

[3.5. PROCESSO DE VALIDAÇÃO CRUZADA DO MODELO 10](#_Toc12313645)

[3.6. TESTE DO MODELO 11](#_Toc12313646)

[4. CONCLUSÃO 12](#_Toc12313647)

[5. REFERÊNCIAS BIBLIOGRÁFICAS 13](#_Toc12313648)

# INTRODUÇÃO

Em [estatística](https://pt.wikipedia.org/wiki/Estat%C3%ADstica) ou [econometria](https://pt.wikipedia.org/wiki/Econometria), **regressão linear** é uma equação para se estimar a condicional (valor esperado) de uma variável *y*, dados os valores de algumas outras variáveis *x.*

A *regressão*, em geral, tem como objectivo tratar de um valor que não se consegue estimar inicialmente.

A regressão linear é chamada "linear" porque se considera que a relação da resposta às variáveis é uma [função linear](https://pt.wikipedia.org/wiki/Fun%C3%A7%C3%A3o_linear) de alguns parâmetros. Os modelos de regressão que não são uma função linear dos parâmetros se chamam modelos de [regressão não-linear](https://pt.wikipedia.org/wiki/Regress%C3%A3o_n%C3%A3o-linear). Sendo uma das primeiras formas de análise *regressiva* a ser estudada rigorosamente, e usada extensamente em aplicações práticas. Isso acontece porque modelos que dependem de forma linear dos seus parâmetros desconhecidos, são mais fáceis de ajustar que os modelos não-lineares aos seus parâmetros, e porque as propriedades estatísticas dos estimadores resultantes são fáceis de determinar.

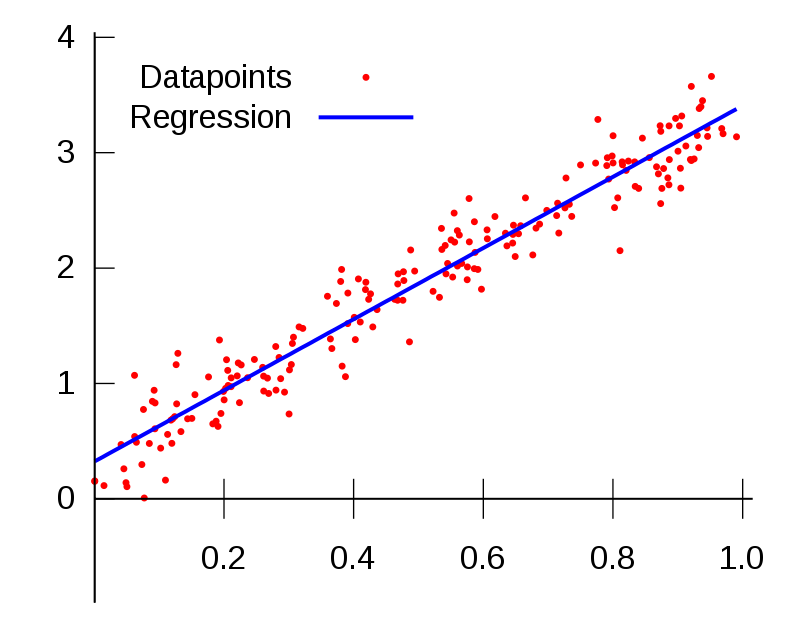


Figura 1 – Exemplo de uma Regressão linear

Modelos de regressão linear são frequentemente ajustados usando a abordagem dos mínimos quadrados, mas que também pode ser montada de outras maneiras, tal como minimizando a "falta de ajuste" em alguma outra norma (com menos desvios absolutos de regressão), ou através da minimização de uma penalização da versão dos mínimos quadrados. Por outro lado, a abordagem de mínimos quadrados pode ser utilizado para ajustar a modelos que não são modelos lineares. Assim, embora os termos "mínimos quadrados" e "modelo linear" estejam intimamente ligados, eles não são sinônimos.

## OBJETIVOS GERAIS

O Propósito deste trabalho é apresentar um exemplo de utilização de um Modelo de regressão linear utilizando o Software WEKA. As Explicações sobre o funcionamento deste modelo e sobre o seu processo de predição não serão escopo deste texto.

## OBJETIVOS ESPECÍFICOS

Efetuar todos os passos necessários para Modelagem de uma Regressão Linear, a exemplo, pode-se citar:

* Download do software necessário.
* Obtenção dos dados a serem analisados *(repassados pelo Prof Meuser)*;
* Normalizar os dados de entrada.
* Criar o modelo utilizando o WEKA.
* Efetuar o treinamento do modelo;
* Teste do modelo;

# SOFWARES UTILIZADOS

## DESCRIÇÃO DOS SOFTWARES

### WEKA

O pacote de software **Weka** (*Waikato Environment for Knowledge Analysis*) começou a ser escrito em 1993, usando [Java](https://pt.wikipedia.org/wiki/Java_(linguagem_de_programa%C3%A7%C3%A3o)), na Universidade de Waikato, Nova Zelândia sendo adquirido posteriormente por uma empresa no final de 2006. O Weka encontra-se licenciado ao abrigo da [*General Public License*](https://pt.wikipedia.org/wiki/GNU_General_Public_License) sendo portanto possível estudar e alterar o respectivo código fonte.

O Weka tem como objectivo agregar algoritmos provenientes de diferentes abordagens/paradigmas na sub-área da [inteligência artificial](https://pt.wikipedia.org/wiki/Intelig%C3%AAncia_artificial) dedicada ao estudo de aprendizagem de máquina.

Essa sub-área pretende desenvolver algoritmos e técnicas que permitam a um computador "aprender" (no sentido de obter novo conhecimento) quer indutiva quer dedutivamente.

O Weka procede à análise computacional e estatística dos dados fornecidos recorrendo a técnicas de [mineração de dados](https://pt.wikipedia.org/wiki/Minera%C3%A7%C3%A3o_de_dados)tentando, indutivamente, a partir dos padrões encontrados gerar hipóteses para soluções e no extremos inclusive teorias sobre os dados em questão.

### MICROSOFT OFFICE 2010 (EXCEL)

**Microsoft Office 2010**, também chamado de **Office 14**, é uma [suíte de escritório](https://pt.wikipedia.org/wiki/Su%C3%ADte_de_escrit%C3%B3rio) ou uma Informática Intermediária para [Microsoft Windows](https://pt.wikipedia.org/wiki/Microsoft_Windows), sucessor do [Microsoft Office 2007](https://pt.wikipedia.org/wiki/Microsoft_Office_2007). O Microsoft Office 2010 oferece compatibilidade estendida a diversos [formatos de arquivos](https://pt.wikipedia.org/wiki/Formato_de_arquivo), atualizações de [interface do usuário](https://pt.wikipedia.org/wiki/Interface_gr%C3%A1fica_do_utilizador) e uma requintada experiência do usuário.

Neste projeto, embora o Office tenha uma suíte vasta de aplicações, somente serão utilizados os software **MICROSOFT WORD E EXCEL**.

# PROCEDIMENTOS REALIZADOS

## DOWNLOAD DO SOFTWARE WEKA

O Download do Softwa pode ser feito no site <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>, no link descrito é possível baixar as verões para qualquer Sistema Operacional, até por que o arquivo principal do Software está em formato “.jar” oriundo de programação produzida em JAVA.

## OBTENÇÃO DOS DADOS A SEREM ANALISADOS

Os dados contém 17.125 linhas diferentes constituídas de 7 Colunas dos quais 6 colunas são os atributos e a ultima é denominada de Classe, estes dados foram repassados pelo professor Meuser Valença e têm a estrutura conforme imagem abaixo.

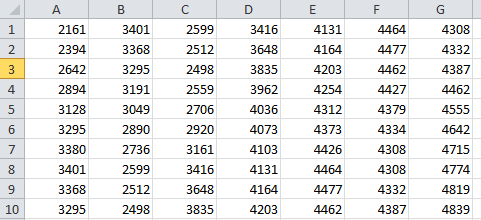


Figura 2 - Dados brutos recebidos

## NORMALIZAÇÃO DOS DADOS

A Normalização dos dados foi feita conforme formula repassada pelo Prof Meuser em aulas e segue segundo a equação:

Onde

Efetuando a devida normalização nos dados previamente obtidos obtém-se um conjunto de valores conforme a imagem abaixo.

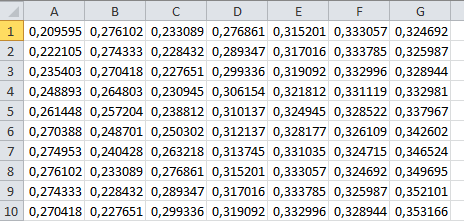


Figura 3 - Dados normalizados

## CRIAR UM MODELO UTILIZANDO O WEKA

Foi testado um algoritmo de regressão baseado em Máquina de Vetores de suporte, o qual não obteve desempenho satisfatório. Sendo feito posteriormente uma **Regressão Linear** que teve um erro médio menor. Observe a figura abaixo.

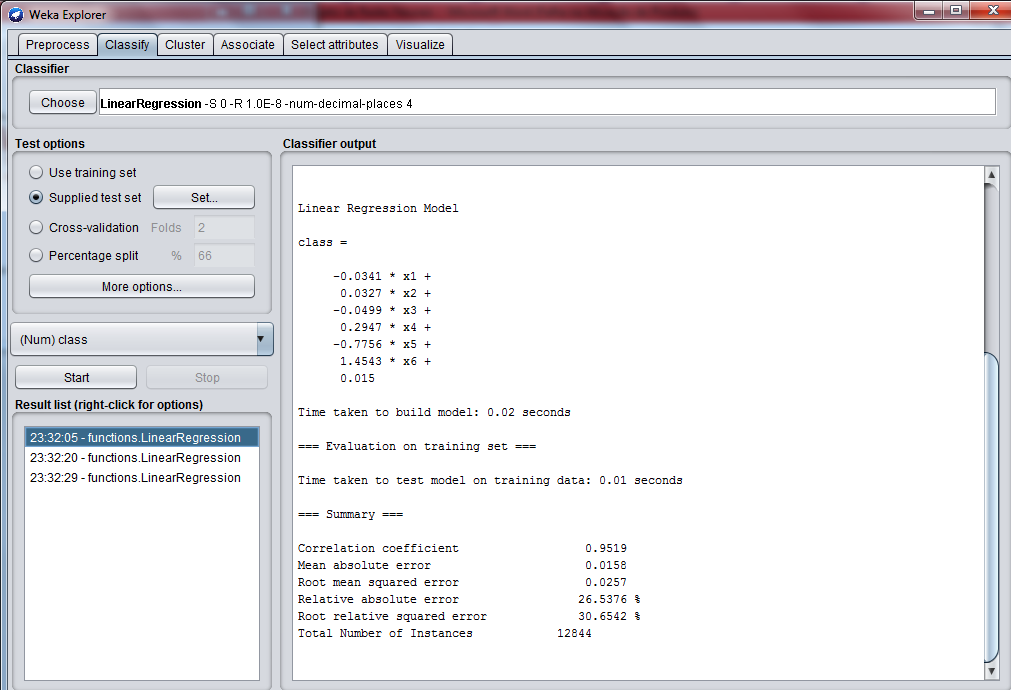


Figura 4 - Utiilizando base de de treinamento

## PROCESSO DE VALIDAÇÃO CRUZADA DO MODELO

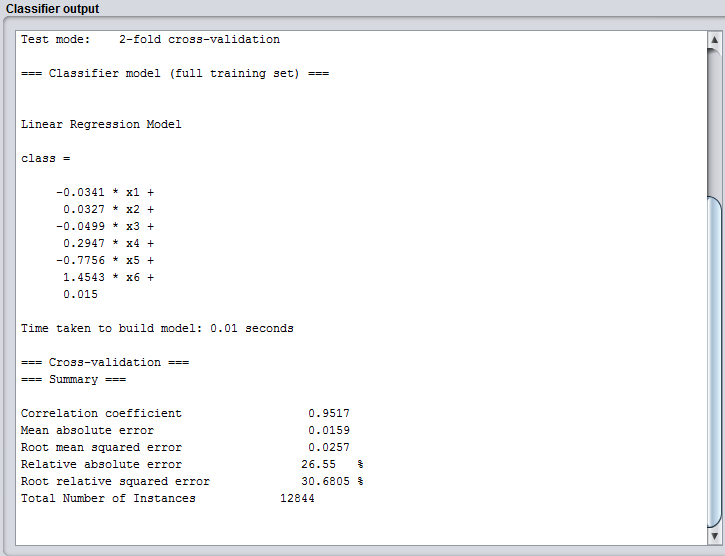


Figura 5 - Validação cruzada do Modelo

## TESTE DO MODELO

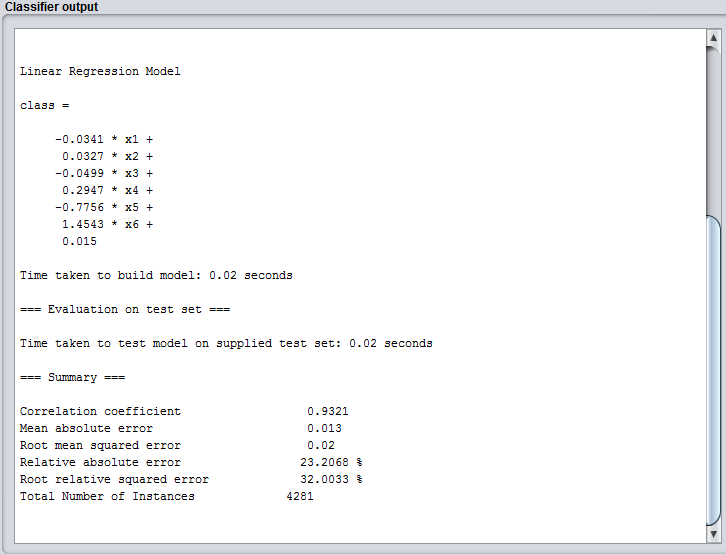


Figura 6 - Testes do Modelo de Regressão Linear

# CONCLUSÃO

O Desenvolvimento desse projeto permitiu verificar que hoje em dia existe a facilidade de implementação de algoritmos de IA, dado o fato que as ferramentas estão populares, e qualquer pessoa com curiosidade o suficiente para aprender pode implementar Modelo que se adeque aos propósitos de um problema especifico.

Os resultados do teste mostraram que existem a variável de saída possui uma maior dependência da variável **x6**, conforme gráfico abaixo.

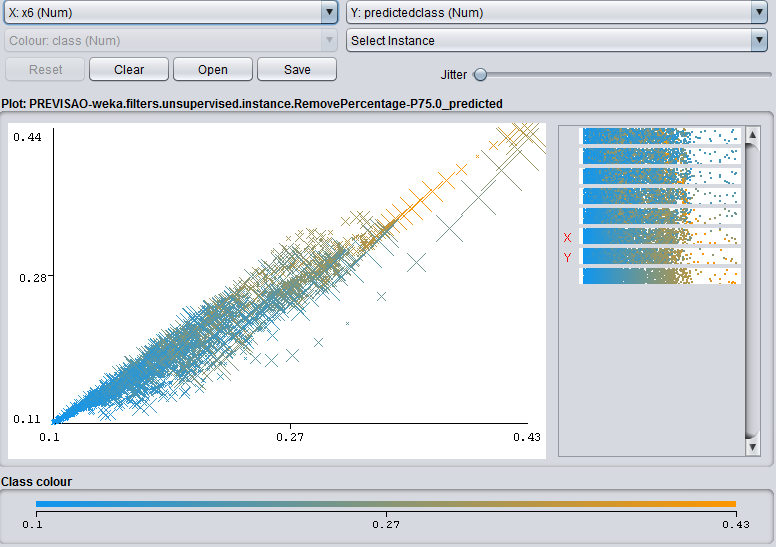


Figura 7 - Dependência da variável x6

# REFERÊNCIAS BIBLIOGRÁFICAS

[1] Site: <http://www.ufjf.br/epd042/files/2009/02/previsao1.pdf>, acessado em 22/06/2019.

[2] Apresentação: Pedro Bragioni, <https://homepages.dcc.ufmg.br/~pedro.lascasas/aula_2_minera%c3%a7%c3%a3o_de_dados_aplicada_weka.pdf>, acessado em 24/06/2019.

[3] Youtube: <https://www.youtube.com/watch?v=iJ7dSdwv2mE>, utilizando weka para separação de dados de treino e teste.