Universidade de Ribeirão Preto

Engenharia da Computação

Sistemas Inteligentes

Igor Nandes - 764155

Lais Caroline Bergamo - 82548

Pedro Afonso Ferreira Haupenthal - 823974

Thiago Alcarás - 824989

RELATÓRIO PROVA PARCIAL

RIBEIRÃO PRETO

2019

**BASE DE DADOS**

Número original de registros para vinho branco: 4899

Número original de registros para vinho vermelho: 600

**ATRIBUTOS ORIGINAIS**

fixed acidity - Acidez fixada

volatile acidity - Acidez volátil (dispersível no ar)

citric acid - Acidez cítrica

residual sugar - Açúcar residual

chlorides - Cloretos presentes

free sulfur dioxide - Livre de ácido sulfúrico

total sulfur dioxide - Total presente de ácido sulfúrico

density - Densidade do composto

pH - Nível de acidez ou basicidade do composto

sulphates - Sulfatos presentes

alcohol - Nível de álcool

quality - Qualidade do composto (vinho)

**CLASSES**

É a classificação da qualidade do vinho, variando de 0 a 10, onde 0 seria um vinho muito ruim e 10 seria um vinho muito bom.

**NORMALIZAÇÃO DAS BASES DE DADOS**

Antes de criar o algoritmo que iria se basear na massa de dados disponibilizada, foi necessário fazer uma normalização dessa base, pois diversos valores encontravam-se fora do normal e despadronizados, assim resultaria em um aprendizado incorreto e consequentemente, resultados imprecisos.

A normalização das duas bases de dados (vinho branco e vermelho) foram feitas da seguinte forma:

Inicialmente constatou-se que não havia uma padronização de separação das casas decimais, ou seja, para a mesma linha ou coluna eram utilizados valores com “,” e “.”, para isso todos os valores foram padronizados com vírgula, de acordo com o padrão utilizado no Excel em Português.

Feito isso, o número máximo de casas decimais foi padronizado para no máximo dois valores, pois em alguns casos era possível observar mais de 10 casas decimais.

Finalizada a parte de formatação dos campos presentes nas bases, foi feita uma análise para descobrir valores que estivessem fora do padrão encontrado para aquele atributo, ou seja, se para o atributo chlorides a média encontrada é 5.8 e possui uma variação máxima de ± 3, e foi encontrado um valor muito abaixo ou acima dessa variação máxima, toda a linha que pertence a esse atributo seria removida, ou seja, aquela amostra de vinho seria removida da base por conter valores muito fora do padrão e que poderia prejudicar o aprendizado da IA.

Em seguida, realizou-se a estatística descritiva através da ferramenta de análise de dados, do Microsoft Excel, para exibir a análise estatística dos dados fornecidos de cada dos atributos. Ao realizar tal tarefa, foram considerados dois valores, o mínimo e o intervalo entre o valor máximo e o mínimo, sendo assim aplicou-se a fórmula: (valor - valor mínimo)/intervalo, onde valor refere-se aos valores de cada atributo.

Quando foi concluída essa análise, foi possível observar uma diminuição no número de registros para ambas as bases, de acordo com o descrito na tabela abaixo:

|  |  |  |  |
| --- | --- | --- | --- |
| Base analisada | n° original de registros | n° final de registros | Perdidos |
| Vinho branco | 4899 | 4780 | 117 |
| Vinho vermelho | 1600 | 1515 | 85 |

**DESENVOLVIMENTO DO ALGORITMO**

ARQUITETURA DA REDE

A arquitetura é composta de 12 entradas, sendo uma delas o bias, 12 elementos na primeira camada oculta, sendo um deles o bias e uma saída, referente à qualidade do vinho.



CRITÉRIO DE SELEÇÃO ADOTADO

Embora os valores de qualidade do vinho fossem variar de 0 a 10 onde 0 seria um vinho muito ruim e 10 seria um vinho muito bom, foi adotado um critério de seleção mais rigoroso, onde caso a qualidade seja menor do que 6, seria considerado um vinho simples e/ou barato, e caso seja um valor maior ou igual a 6, seria considerado um vinho mais seleto, ou seja, mais caro, fazendo com que o usuário do sistema (no caso o dono da produção de vinhos) possa tomar atitudes referentes a qualidade da safra que foi produzida, podendo vender a preços maiores ou menores.