



E380

CLASSIFICAÇÃO DE TEXTOS UTILIZANDO REGRESSÃO LOGÍSTICA

Lenina Velly Lotufo (Bolsista PIBIC/CNPq) e Profa. Dra. Nancy Lopes Garcia (Orientadora), Instituto de Matemática, Estatística e Computação Científica - IMECC, UNICAMP

É de grande interesse estudar a língua portuguesa e verificar as diferenças existentes entre o português europeu e brasileiro: são diferenças na pronúncia, diferenças lexicais, diferenças sintáticas e mesmo diferenças pragmáticas, ou seja, diferenças de adequação ao uso da língua. Este estudo visa classificar estas duas variedades do português utilizando as diferenças rítmicas, em particular a distância entre as sílabas tônicas das palavras. Os textos utilizados para o estudo constituíram-se em textos modernos do final do século XX. Foi necessário estudar maneiras de codificar estes textos e realizar uma análise exploratória de dados cuidadosa a fim de encontrar variáveis que discriminassem os textos. O objetivo final foi classificar os textos Português Europeu e Brasileiro utilizando a regressão logística.

Classificação - Regressão logística - Lingüística