

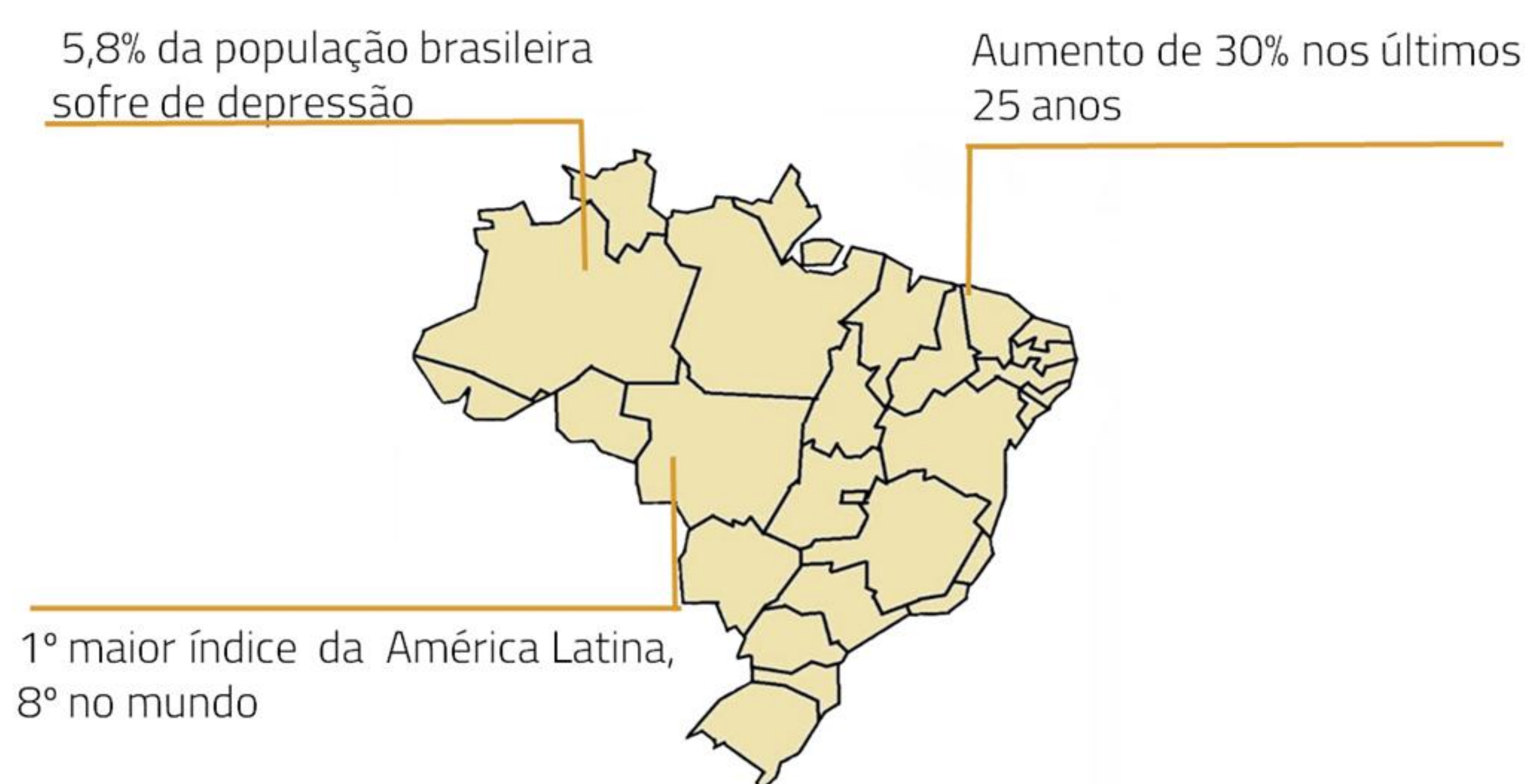
IDENTIFICAÇÃO DE DEPRESSÃO A PARTIR DA ANÁLISE DE TEXTOS



Adriana Maria Padilla, Felipe de Araújo Sarmiento, Fernanda Tanajura Piva, Vinicius Lago Fernandes
Professor Orientador: Tiago Sanches da Silva

Introdução

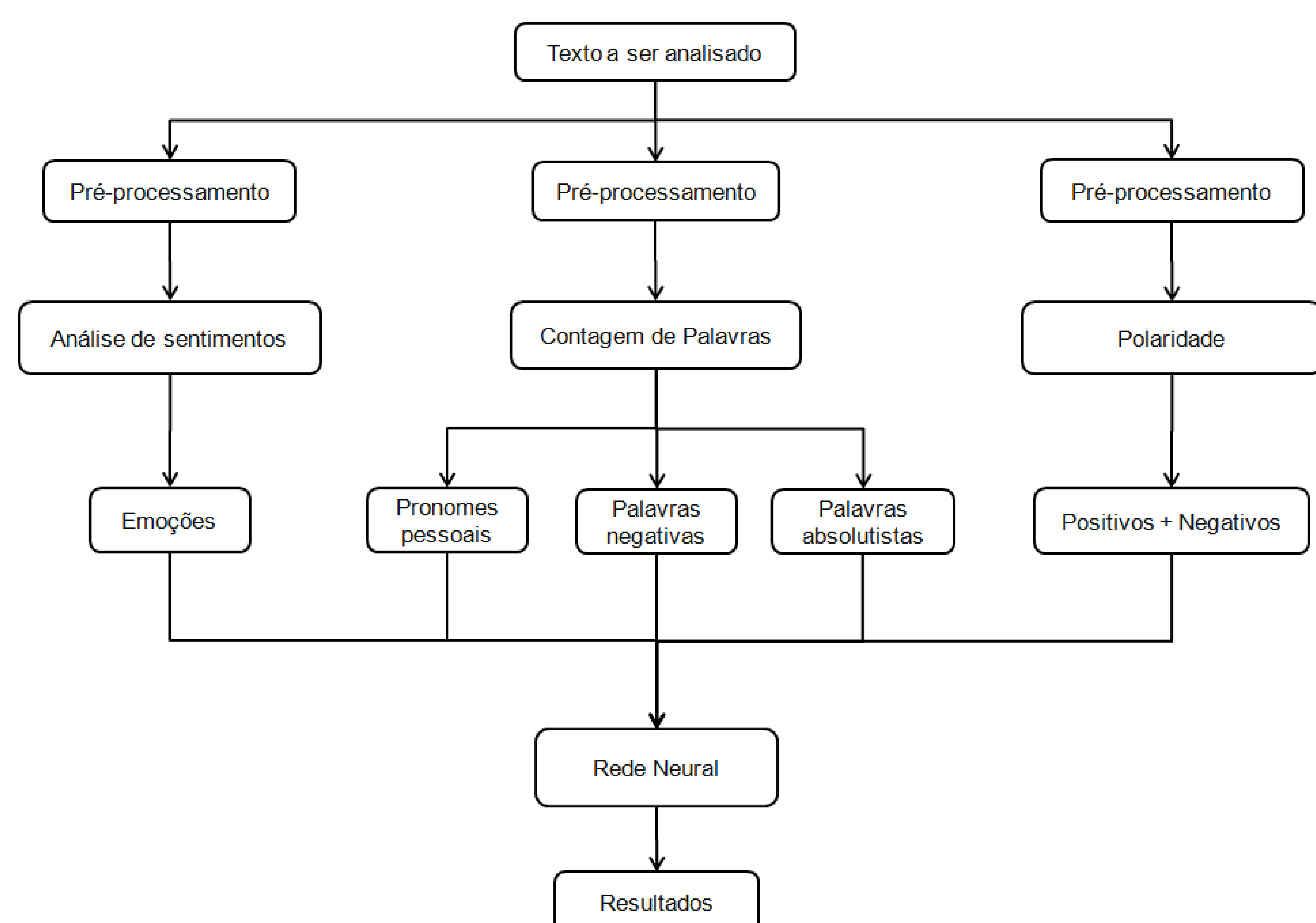
A depressão é uma doença psiquiátrica crônica e recorrente que produz alteração de humor caracterizada principalmente por tristeza profunda e forte sentimento de desesperança. É estimado que cerca de 322 milhões de pessoas sejam afetadas no mundo pela doença, segundo dados da Organização Mundial da Saúde (OMS), e que, se não tratada corretamente, pode causar grandes alterações na vida dessas pessoas ou até mesmo, em casos mais sérios, o suicídio.



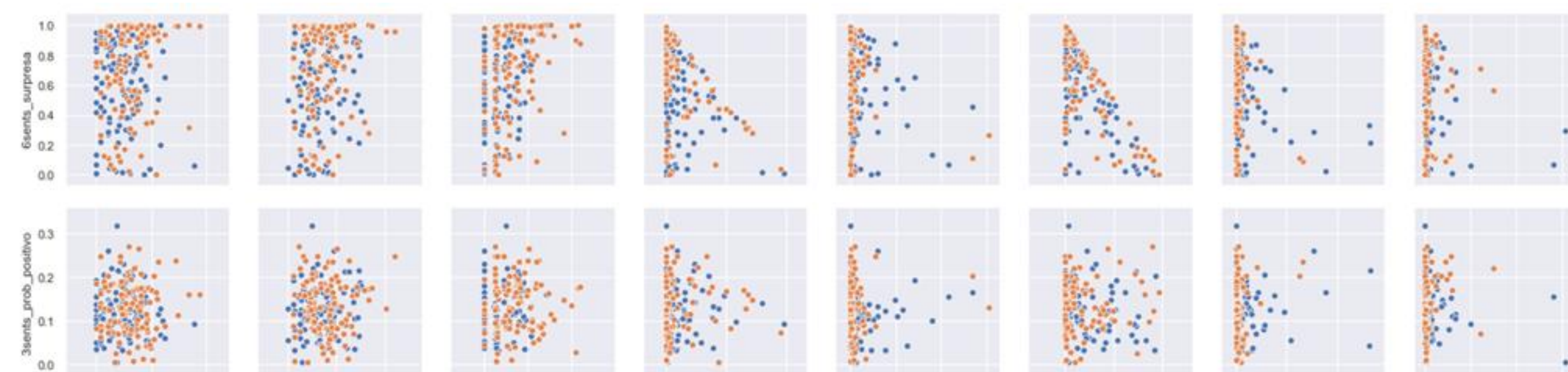
Os dados divulgados pela OMS também falam a respeito do Brasil, que possui o maior índice da América Latina de pessoas que sofrem com a depressão e está em 8º no mundo.

Visto estes dados, foi escolhido como tema a detecção de sinais de depressão em textos utilizando técnicas de Programação Neurolinguística (NLP) com o auxílio da Inteligência Artificial. Como o objetivo foi focar no Brasil, o trabalho foi realizado para textos em português.

Metodologia



O Pré-processamento ilustrado no fluxograma anterior tem o objetivo principal de transformar os textos a serem analisados em números para que dessa forma seja possível manipular as informações para o âmbito computacional.

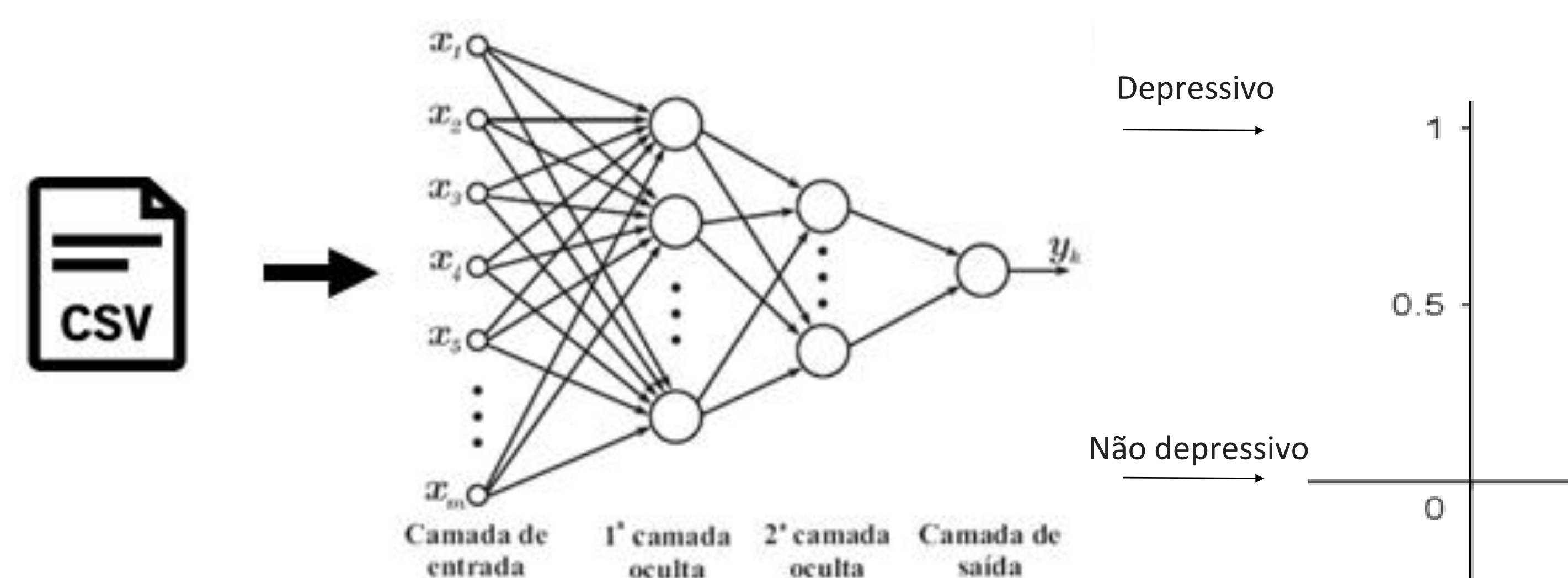


A imagem acima está ilustrando uma parte dos dados já pré-processados. A cor laranja é representada pelas pessoas que têm depressão e a cor azul pelas pessoas que não tem depressão.

O treinamento da rede neural, nesse trabalho, está inserido no contexto de aprendizagem de máquina supervisionada, em que as amostras fornecidas dos dados contêm a que classificação elas pertencem, ou seja, se possuem depressão ou não.

Essas amostras foram divididas em duas partes, 70% delas foi utilizado para o treinamento do modelo da rede e 30% foi utilizado para testar esse modelo treinado.

À seguir está representado a arquitetura da solução proposta:



Resultados

Atualmente o modelo está com uma acurácia de 75%. Uma das métricas de avaliação utilizadas chama-se Matriz de Confusão, que identifica quantos dados estão sendo previstos corretamente pela rede neural conforme ilustrado abaixo:

