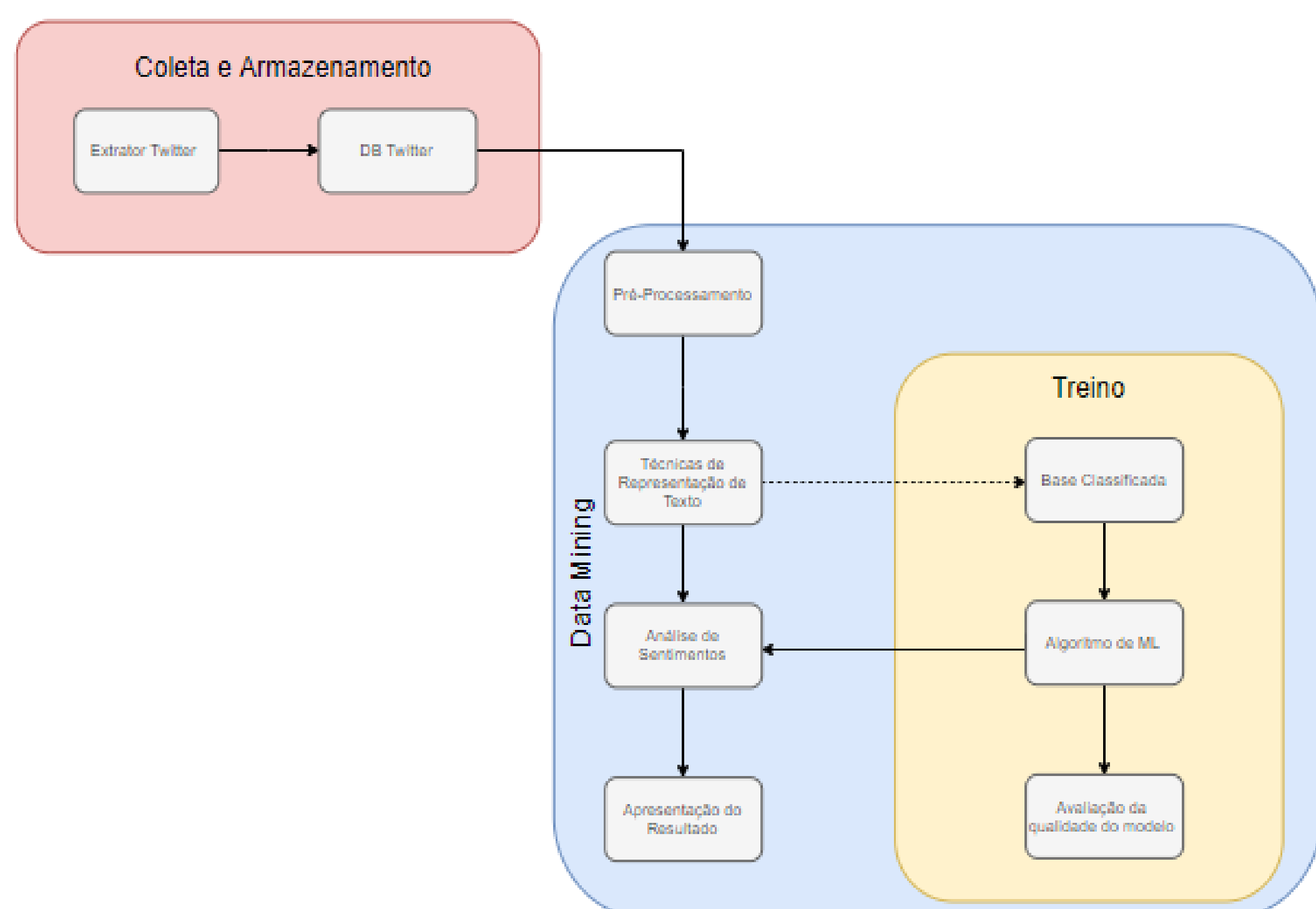


Objetivo

O objetivo deste trabalho é realizar a coleta, armazenamento e análise de sentimentos de dados de redes sociais de forma automatizada, para auxílio em pesquisas na área da ciência da conservação. A análise das interações humano-fauna é de grande importância para políticas de conservação de espécies e o uso de mensagens de redes sociais permite coletar eventos e opiniões em larga escala, gerando benefícios para esta área de pesquisa.

Arquitetura

A arquitetura completa do sistema é a mostrada abaixo.



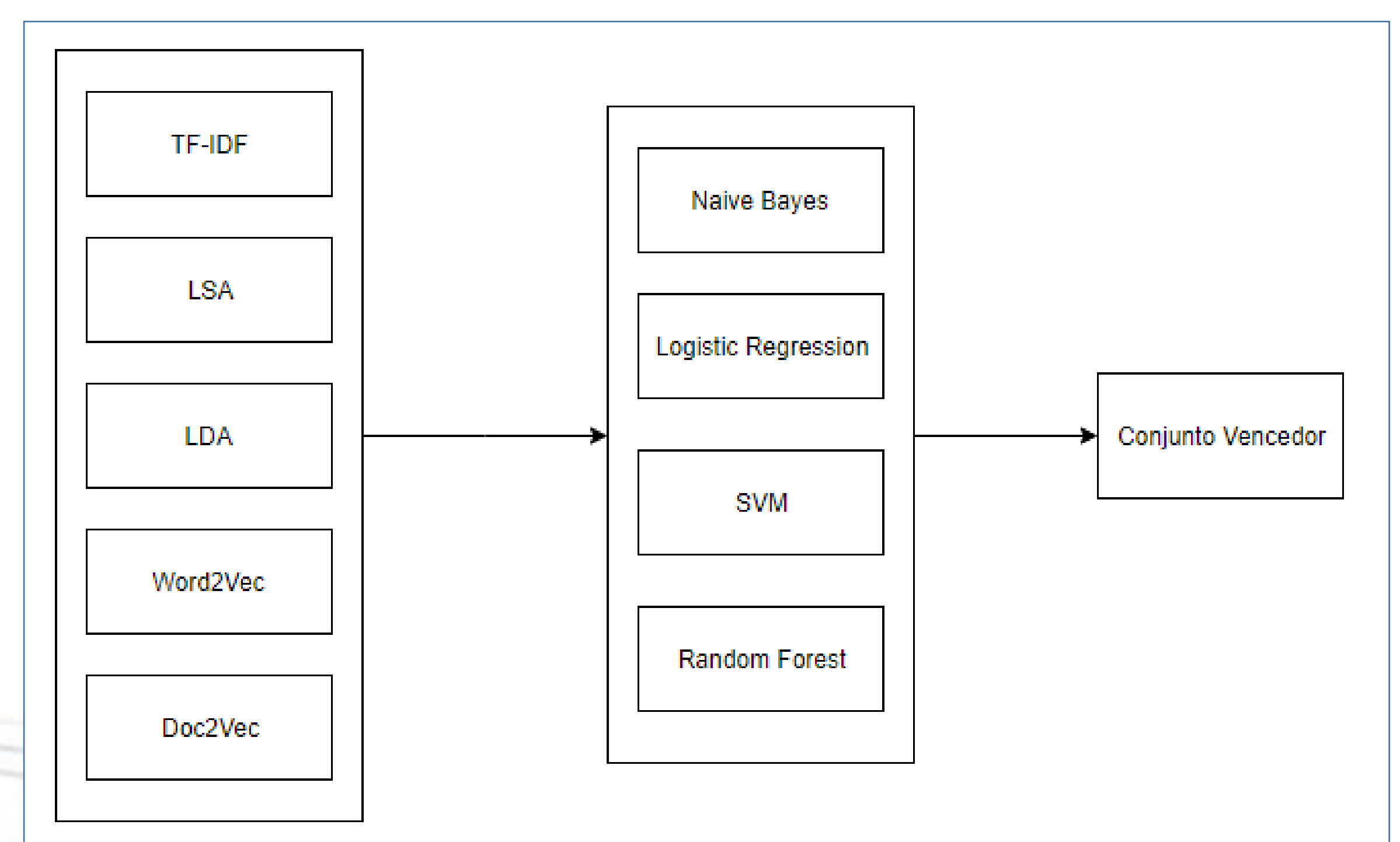
Inicialmente, extraem-se mensagens do Twitter através de palavras chaves relevantes informadas pelo usuário. As mensagens são armazenadas em um banco de dados em conjunto com outros metadados relevantes. Posteriormente é feito o pré-processamento das mensagens extraídas e utilizando técnicas de representação vetorial de texto (embeddings) e algoritmos de machine learning é feita a análise de sentimentos dos dados extraídos. Para o treino do algoritmo é utilizada uma base previamente classificada pelo usuário.

Integrantes: Beatriz de Oliveira Silva
Rodrigo da Matta Bastos

Professor Orientador: Ricardo Luis de Azevedo da Rocha
Co-orientador: Pedro Luiz Pizzigatti Corrêa

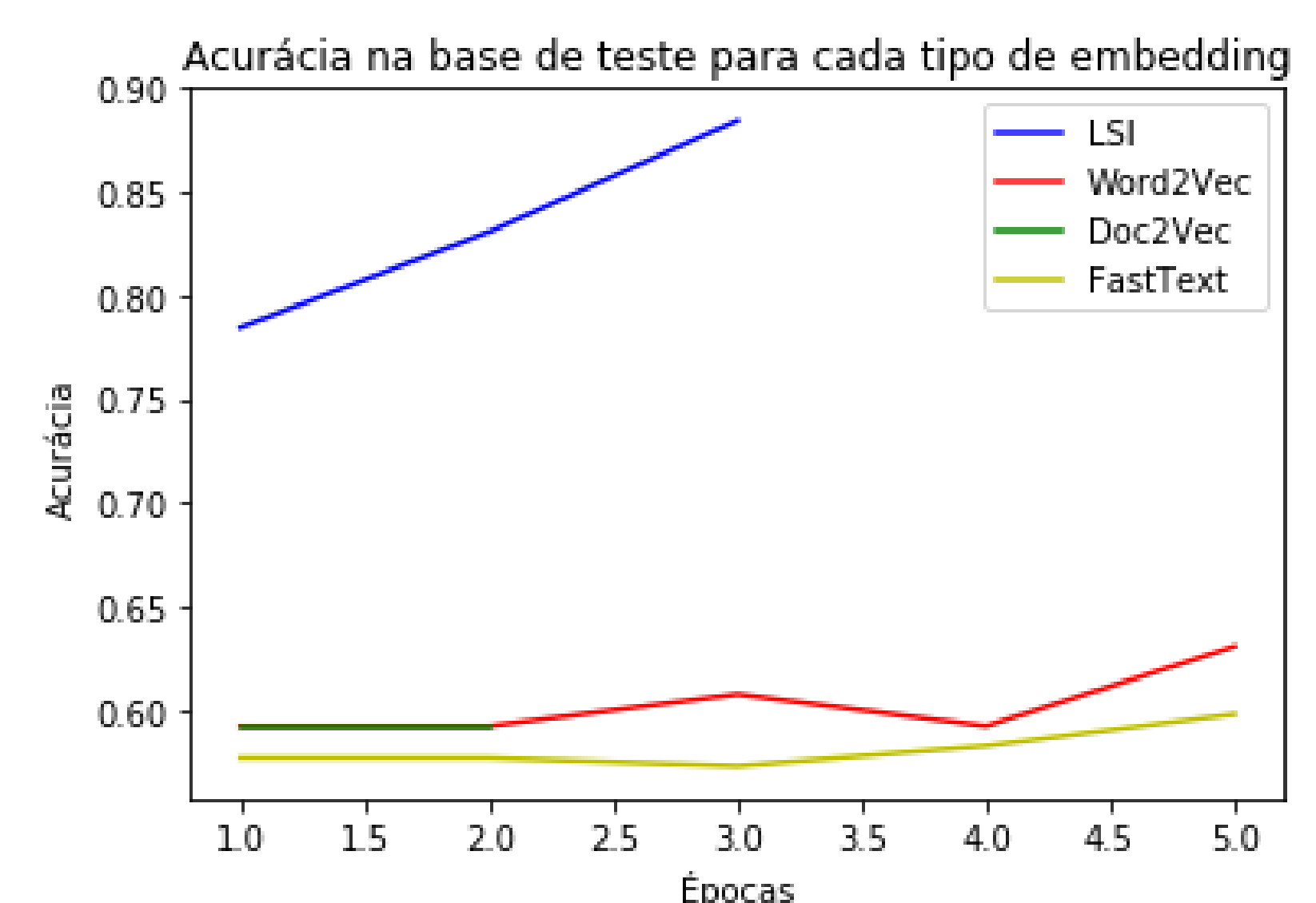
Seleção de Algoritmo

Para selecionar o tipo de embedding e o algoritmo de machine learning utilizado foi criada uma rotina de busca na etapa de treino, representada abaixo.



Modelos Neurais

O trabalho também envolveu o estudo de modelos de redes neurais para a análise de sentimentos. Os modelos implementados foram construídos utilizando o API Keras, Python e com o *backend* TensorFlow. Os modelos têm em comum uma camada de neurônios escondida com dimensão variável, uma ou mais camadas de entrada de acordo com o *embedding* que entra na rede neural e saída com a classificação. Além da quantidade de neurônios na rede ser variável, cada modelo tem um otimizador, função de ativação, função de custo (ou perda), tamanho de *batch* e quantidade de épocas diferentes. Os modelos com melhor desempenho foram os que receberam *embedding* LSA como entrada.



Conclusão

O trabalho aqui apresentado explora a possibilidade do uso de mensagens de redes sociais na área da ciência da conservação. Isto permite expandir o alcance dos estudos realizados e portanto serve como contribuição para esta área de pesquisa.