Após fazer a classificação dos documentos

Nayves Bayes – importar a biblioteca dentro do Python

Metodologia (Abordagem metodológica)

- Fazer a metodologia em cima do pdf que ela enviou

- Caracterização da pesquisa

Fundamentação/Teoria

Primeiro modelo de classificação.

1. **Abordagem Metodológica**

No processo de desenvolvimento desta pesquisa, optou-se pelo campo da pesquisa quantitativa, tendo em vista que é uma aplicação onde seus objetivos principais envolvem a coleta de dados e sua análise e trabalha com medições objetivas e verificáveis, tais quais se assemelham aos principais objetivos de uma pesquisa quantitativa. Já com relação a natureza da pesquisa, foi optado por uma “natureza aplicada”, onde visa produzir uma aplicação que possa ser utilizada na vida real. Observando então às especificidades dos objetivos propostos à aplicação, podemos enquadrá-la então em uma pesquisa exploratória que, de acordo com Gil (2010), tem como objetivo conquistar maior familiaridade com o problema a fim de torná-lo mais explícito ou para fins de construção de hipóteses, dando-se pelo fato da busca dos algoritmos classificadores de texto para a realização do encaminhamento de documento automaticamente. Por fim, relacionado aos procedimentos técnicos seguidos nessa aplicação, pode ser assemelhado a uma pesquisa-ação junto a uma pesquisa documental, onde é preciso compreender e basear-se em estudos relacionados à classificação de documentos via redes neurais e elaborar o método de encaminhamento automático desses documentos previamente classificados.

1. **Ferramentas utilizadas**

As ferramentas utilizadas para as etapas de coleta, pré-processamento, construção do modelo de classificação e exibição são as descritas

a seguir:

Base de dados local – Base de dados construída para a recuperação de documentos e analisar dados. A partir dela podemos obter dados genéricos dos documentos para sua classificação inicial

Python 3 – Python é uma linguagem de programação de fácil escrita e

entendimento, e foi escolhida pelo fato de sua vasta gama de bibliotecas relacionadas aos temas principais.

NLTK – Biblioteca python que trata e trabalha com o Processamento de Linguagem Natural, voltado para a manipulação da linguagem humana no pré-processamento de dados.

Scikit Learn – Biblioteca Python com várias funcionalidades voltadas para aprendizado de máquina, como algoritmos classificadores.

Pandas - Biblioteca Python capaz de simplificar tarefas de manipulação de dados. Fornece ferramentas de análise de dados e estruturas de dados de alta performance, compreendendo manipulação, leitura e visualização de dados.

Banco de dados

Dados do Experimento

Para o treinamento e testes dos algoritmos, será desenvolvida uma base de dados própria que consiste em inúmeros tipos de documentos, tais documentos foram obtidos através do desenvolvimento e solicitações de militares a seus superiores imediatos, os quais foram salvos e armazenados para a geração do modelo de classificação. Todos os documentos contêm autorização prévia do uso para testes autorizada pelos autores e foram salvos no formato de *html*, para melhor utilização das bibliotecas de interpretação, a fim de obter um melhor resultado.

Após isso, será utilizado a biblioteca *NLTK* (*python*) para fazer o pré-processamento dos dados desses documentos, tais quais serão analisados e tratados, como exemplo a remoção de “stop words” e espações em brancos, trabalhando apenas com o necessário para classificação do documento.

Em seguida, com a base de dados em ordem e após o pré-processamento, será aplicado com o auxílio da biblioteca *Scikit Learn* os algoritmos de classificação (*redes neurais convolucionais - CNN*) e será analisado as porcentagens de assertividade de classificação para o armazenamento de seus documentos dentro dessas classes.

Por fim, com os documentos armazenados dentro de suas classes, será utilizado a biblioteca *Pandas* para o preenchimento automático de e-mails destinados ao responsável por aquela classificação de documento, podendo então seguir com o fluxo de vida útil do documento encaminhado.