

**UNIVERSIDADE PRESBITERIANA MACKENZIE**  
**FACULDADE DE COMPUTAÇÃO E INFORMÁTICA**

Emerson Moreira Baliza  
Erick Meyer Machado Terceiro  
Glaysen José Oliveira da Silva  
Thiago da Silva Fregulha

**Modelo de Classificação de Dados**

**SÃO PAULO**  
**2023**

Emerson Moreira Baliza – 22015851  
Erick Meyer Machado Terceiro – 22008225  
Glaysen José Oliveira da Silva - 22016449  
Thiago da Silva Fregulha – 22017534

### **Modelo de Classificação de Dados**

Trabalho final da disciplina de Projeto Aplicado II apresentado à Faculdade de Computação e Informática da Universidade Presbiteriana Mackenzie, como parte dos requisitos obrigatórios de conclusão de curso.

Docente: Prof. Anderson Adaime de Borba

**SÃO PAULO**  
**2023**

# **1 INTRODUÇÃO**

## **1.1 Contexto do estudo**

Foi realizado uma análise referente ao dataset. Os documentos referentes a análise foram armazenados na nuvem, em uma pasta do Github que pode ser acessada pelo link:

<https://github.com/tsfregulha/Projeto-Aplicado-Emotions/projects?query=is%3Aopen>

## **1.2 Descrição da origem**

O dataset foi retirado do Kaggle, sendo esta uma plataforma fundada em 2010, que possui um repositório de dados disponíveis publicamente para prática e experimentação de novos modelos.

É voltada também para competições de ciência de dados, onde indivíduos e equipes podem competir para produzir os melhores modelos preditivos para um determinado conjunto de dados.

## **1.3 Descrição do dataset**

O dataset selecionado busca ser capaz de identificação das emoções dos usuários, classificando entre positivas e negativas com base no reconhecimento de emoção de texto.

## **2 DESENVOLVIMENTO E VALIDAÇÃO DO ALGORITMO**

### **2.1 Modelo**

O modelo pretende seguir conforme esquema a seguir:

Abrir > Ler > Copiar > Colar > Classificação

### **2.2 Objetos**

Criar modelo capaz de ler arquivos e realizar a classificação e destinação do documento para o setor adequado.

### **2.3 Metas**

- a) Criar modelo em Python e Knime.
- b) Aplicar modelo em dados selecionados.
- c) Validar modelo com dados de teste.

### **2.4 Metodologia**

- a) Revisão Bibliográfica

Nesta etapa, será feito a revisão da literatura, a fim de identificar trabalhos similares, de modo que tenhamos referência de um caminho a ser seguido.

- b) Instalação e configuração do Knime, Python e Bibliotecas

Será instalado e configurado o Knime, Python no computador, bem como, das bibliotecas, de darão suporte para o avanço do projeto.

- c) Seleção de dataset

Nesta etapa, será feito a busca de um dataset compatível com o projeto.

- d) Planejamento de trabalho

Será feito o planejamento detalhado de cada etapa do projeto.

e) Teste de manipulação

Com o dataset definido, Knime e o Python instalado, será feito os testes.

f) Validação do modelo

Conforme o modelo for sendo construído, será feito etapas de validação, de modo que no final, tenhamos um modelo validado e funcional.

## 2.5 Resultados esperados

É desejado criar, testar e validar um modelo capaz de manipular arquivos, para dar suporte em uma empresa, de modo que o modelo acesse a fonte de dados, faça a leitura e classificação dado um série de regras à serem estabelecidas.

## 2.6 Cronograma

Descrição	Março/23	Abril/23	Maiio/23
1.Revisão Bibliográfica - Data X	x		
2.Instalação e configuração do Python e Bibliotecas		x	
3.Seleção de dataset	x		
4.Planejamento de trabalho	x	x	
5.Teste de manipulação		x	
6.Validação do modelo		x	x