

# Previsão Da Religião Predominante Em Um País Com Base Em Sua Bandeira

1<sup>st</sup> Victor Bruno de Moura Souza  
Centro de Informática - CIn  
Universidade Federal de Pernambuco - UFPE  
Recife, Brasil  
vbms@cin.ufpe.br

2<sup>nd</sup> Marcelo Menezes Valois  
Centro de Informática - CIn  
Universidade Federal de Pernambuco - UFPE  
Recife, Brasil  
mmv4@cin.ufpe.br

3<sup>rd</sup> Joao Gabriel de Araújo Vasconcelos  
Centro de Informática - CIn  
Universidade Federal de Pernambuco - UFPE  
Recife, Brasil  
jgav@cin.ufpe.br

4<sup>th</sup> Edinaldo Barbosa da Costa Filho  
Centro de Informática - CIn  
Universidade Federal de Pernambuco - UFPE  
Recife, Brasil  
ebcf2@cin.ufpe.br

**Abstract**—Será discutido neste artigo a previsão da religião predominante em um país usando como base a sua bandeira, usando o classificador probabilístico de Naive Bayes.

**Index Terms**—Bandeira, religião, predominante, classificador, bayes

de tal metodo é justificada fundamentalmente visto que as instâncias providas no nosso objeto de estudo (bandeira) são condicionalmente independentes (o teorema de Bayes trata antes de tudo sobre probabilidade condicional).

## I. INTRODUÇÃO

A proposta do projeto em questão define o uso do Algoritmo Probabilístico de Classificação Ingênua de Bayes (Naive Bayes), com o propósito de avaliar o que as características presentes na bandeira de um país podem dizer quanto à religião predominante no mesmo. O modelo a ser desenvolvido faz uso de um Dataset (banco de dados) provido pelo repositório "<https://archive.ics.uci.edu/ml/datasets/Flags>", cujos dados foram coletados do documento "Collins Gem Guide to Flags"(Collins Publishers - 1986). Como esse Banco de Dados também possui informações complementares ao país (tais como nome, área, continente ao que pertence, população e linguagem), estas também exercerão interferência na previsão da religião dominante.

## II. OBJETIVO

O objetivo principal é desenvolver um modelo que irá tentar determinar a religião predominante em um país com base nas características de sua bandeira, tais como a presença de círculos, triângulos, listras, etc. Isso nos dirá, por exemplo, se há uma influência histórica, da religião, na criação das bandeiras ao longo do tempo.

## III. JUSTIFICATIVA

O modelo a ser desenvolvido será baseado no Algoritmo de Classificação Ingênua de Bayes (Naive Bayes), concomitante as bibliotecas matemáticas de aprendizado de máquinas, implementadas em Python (como Pandas, Scikit etc). A escolha

## IV. ALGORITMO NAIVE BAYES

O algoritmo "Naive Bayes" é um classificador probabilístico muito utilizado em machine learning. Baseado no "Teorema de Bayes", o modelo foi criado por um matemático inglês chamado Thomas Bayes (1701 – 1761). Atualmente, o algoritmo se tornou popular na área de Aprendizado de Máquina (Machine Learning) para categorizar textos baseado na frequência das palavras usadas, e assim pode ser usado para identificar se determinado e-mail é um SPAM ou sobre qual assunto se refere determinado texto, por exemplo. Por ser muito simples e rápido, possui um desempenho relativamente maior do que outros classificadores. Além disso, o Naive Bayes só precisa de um pequeno número de dados de teste para concluir classificações com uma boa precisão. A principal característica do algoritmo, e também o motivo de receber "naive" (ingênuo) no nome, é que ele desconsidera completamente a correlação entre as variáveis (features). Ou seja, se determinada fruta é considerada uma "Maçã" se ela for "Vermelha", "Redonda" e possui "aproximadamente 10cm de diâmetro", o algoritmo não vai levar em consideração a correlação entre esses fatores, tratando cada um de forma independente.

## V. SOBRE OS DADOS UTILIZADOS

### A. Base de dados

Foi utilizado um banco de dados que está disponível na plataforma UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Flags>).

### *B. Análise exploratória de dados*

Foi utilizado a biblioteca "pandas", criada para a linguagem Python que permite a manipulação e análise de dados. Em particular, oferece estruturas e operações para manipular tabelas numéricas e séries temporais.

## VI. APLICAÇÃO

### *A. Experimentos*

O primeiro experimento foi a bandeira de Nilfgaard, do império de "The Witcher", uma série de livros. O programa analisou todos os dados que colocamos e concluiu que a bandeira se refere a um território de religião desconhecida, que não está no banco de dados. No segundo experimento, usamos a bandeira de Wakanda, um país fictício presente nas histórias em quadrinhos da Marvel. Após analisar os dados da bandeira, o programa concluiu que a possível religião desse território é "Marxist".

### *B. Análise dos resultados*

Vimos que quando aplicado em bandeiras reais, o nosso algoritmo produziu um resultado satisfatório com uma boa porcentagem de acertos. Devido ao algoritmo de Bayes não precisar de um vasto banco de dados e seu rápido aprendizado, se torna uma boa ferramenta para ser utilizada nesse tipo de experimento.

## CONCLUSÕES E DISCUSSÕES

Concluímos que o Algoritmo de Classificação Ingênua de Bayes é satisfatório para o tipo de atividade proposta, uma vez que, experimentos realizados com objetos que não estavam no banco de dados, o modelo retornou resultados esperados.

## REFERENCES

- [1] Paul L. Meyer, "Probabilidade Aplicações à Estatística", 2ª Edição.
- [2] <https://www.vooo.pro/insights/6-passos-faceis-para-aprender-o-algoritmo-naive-bayes-com-o-codigo-em-python>
- [3] <https://github.com/Gvascons/Prediction—Naive-Bayes>
- [4] [https://pt.wikipedia.org/wiki/Teorema\\_de\\_Bayes](https://pt.wikipedia.org/wiki/Teorema_de_Bayes)